Final Project Results Report

Deliverable ID:	D5-3
Dissemination Level:	PU
Project Acronym:	MALORCA
Grant:	698824
Call:	H2020-SESAR-2015-1
Topic:	Data Science in ATM
Consortium Coordinator:	DLR
Edition date:	30th April 2018
Edition:	2.00
Template Edition:	02.00.01

Authoring & Approval

Authors of the document

Name/Beneficiary	Position/Title	Date
Hartmut Helmke	Project Leader	2018-02-25
Mittul Singh	WP3 Leader since Feb 18	2018-02-24
Petr Motlicek	WP4 Leader	2018-02-25
Christian Kern	Dissemination Manager	2018-02-12
Aneta Cerna	WP1 Leader	2018-02-23
Dietrich Klakow	USAAR	2018-02-22

Reviewers internal to the project

Name/Beneficiary	Position/Title	Date
Heiko Ehr	WP5 Leader	2018-02-26
Petr Motlicek	WP4 Leader	2018-04-30 (of sect. 4-3)

Approved for submission to the SJU –y - Representatives of beneficiaries involved in the project

Name/Beneficiary	Position/Title	Date
Hartmut Helmke	Project Lead	2018-02-28
Hartmut Helmke	Project Lead	2018-04-30

Rejected –y - Representatives of beneficiaries involved in the project

Name/Beneficiary	Position/Title	Date

Document History

Edition	Date	Status	Author	Justification
0.01	2018-01-15	Draft	H. Helmke	First version
	2018-02-05	Draft	H. Helmke	Desc. of all deliverables
	2018-02-12	Draft	Chr. Kern	Contrib. to Master Plan
	2018-02-14	Draft	H. Helmke	Review of Chr. K. and updating information from self-evaluation excel sheet in comments of sections

	2018-02-19	Draft	H. Helmke	References added
	2018-02-19	Draft	H. Helmke	Section 2.2 written
	2018-02-23	Draft	A. Cerna	Executive summary prepared, document review
0.02	2018-02-24	Draft	H. Helmke	Explaining text for maturity assessment prev. changes of HHe in ch. 3 accepted
		Draft	H. Helmke	Reformulation of Executive summary
		Draft	H. Helmke	Scope section 2.2 integrated into context section (2.1), Abstract written
0.03	2018-02-24	Draft	M. Singh	Section 2.3 and 2.4
		Draft	H. Helmke	Scope section 2.2 integrated into context section (2.1), Abstract written
0.04	2018-02-05	Draft	H. Helmke	Acceptance of all previous changes; section 4.3 written
0.05	2018-02-26	Draft	H. Ehr	Review; section 1 and 4
0.06	2018-02-26	Draft	P. Motlicek	Review; section 4 updated
0.07	2018-02-26	Draft	H. Helmke	Review of Heiko's and Petr's improvements, changing dates of deliverables, first version of sect. 4.3
0.08	2018-02-27	Draft	Petr Motlicek	Update section 4.3, full review other sections
0.50	2018-02-27	Draft	H. Helmke	Acceptance of all changes accept Petr, deleting all hidden instructions, review of all chapter and review of 4.3; accepting all changes of Petr
	2018-02-28	Draft	H. Helmke	Correcting authors of D5-2 and D5-3 in reference list, adding interface aspect to sect. 4.3
	2018-02-28	Draft	J. Rataj	Review sect. 3.2 and improving sect. 4.3
1.00	2018-02-28	Final	H. Helmke	Accepting all changes and preparation for upload to SJU, Proof- reading of 4.3
1.01	2018-03-12	Draft	H. Helmke	References, integration of URL's to public deliverables on MALORCA's homepage (req. of SJU)
	2018-03-20	Draft	H. Helmke	Minor improvements
	2018-03-20	Draft	H. Helmke	Sect. 3.2, adding results info maturity assessment table and not only the references, (req. of SJU)
1.02	2018-03-21	Draft	H. Helmke	End-to-end learning added to 4.3
1.03	2018-03-21	Final	H. Helmke	Acceptance of all changes and prepared for SJU submission
1.05	2018-03-29	Draft	H. Helmke	Integration of SJU feedback with respect to Maturity Assessment
	2018-04-20	Draft	H. Helmke	Real-time aspect added to next steps

	2018-04-23	Draft	H. Helmke	Sect. 4.3 assignment of research questions to ER, ER-IR, Industrial research (wave-2)
1.06	2018-04-30	Draft	H. Helmke	Acceptance of all previous changes and final review, integration of feedback of Aneta Cerna with respect to next steps in excel sheet
1.07	2018-04-30	Final	H. Helmke	Acceptance of all previous changes, repair of link for D1-1 and deleting green colours in links (Feedback of A. Prister from 2018-04-30), integration of improvements of PM in document with name 4.3
2.00	2018-04-30	Final	H. Helmke	Accepting all changes a preparation for final upload to SJU portal

MALORCA

MACHINE LEARNING OF SPEECH RECOGNITION MODELS FOR CONTROLLER ASSISTANCE

This document is part of a project that has received funding from the SESAR Joint Undertaking under grant agreement No 698824 under European Union's Horizon 2020 research and innovation programme.

Abstract

The Active Listening Assistant (AcListant®) project has shown that Assistant Based Speech Recognition (ABSR) is a solution to reduce controller's workload and to increase ATM efficiency. For Dusseldorf airport all speech recognition models were manually created which is too expensive if the manual work is needed for many mid-size airports. The Horizon 2020 funded project MALORCA offers machine learning as a solution. Each Air Navigation Service Provider generates Mega Bytes or even Giga Bytes of radar data and voice recordings on a daily basis. More than 100 hours of radar data and utterance from controller pilot communication for Vienna and for Prague approach area were recorded. An initial basic ABSR system was set-up for both Prague and also for Vienna approach. MALORCA developed a very novel machine learning approach, which was not applied before in Automatic Speech Recognition domain. The algorithms can rely on two independent information sources: (1) acoustic scores are combined with (2) scores extracted from radar data. This approach reduces command recognition error rates of the baseline system from 7.9% below 0.6% for Prague and from 18.9% to 3.2% for Vienna. The performance of the trained ABSR system was successfully evaluated in proof-of-concept trials by ATC controllers in Vienna and Prague end of January 2018.

Table of Contents

List of Tables

List of Figures

1 Executive Summary

Although air traffic control is a very innovative business, paper flight strips are still in use. Written information on them is not available in digital form. Modern controller working positions (CWP), therefore, offer digital flight strips. However, manual input is required from the controllers. Others have the benefits and the controllers get additional workload.

The Active Listening Assistant (AcListant®) project has shown that Assistant Based Speech Recognition (ABSR) is a potential solution to reduce controllers' workload. A speech recognizer analyses the controller pilot communication and shows the recognitions in the radar label directly to the controller. As command recognition rates better than 95% are possible, the controller only needs to manually correct the output of speech recognizer in one of twenty cases. The controller gets additional free cognitive resources which could e.g. increase safety.

For Dusseldorf airport all speech recognition models were developed from manually transcribed data which is too expensive if the manual work is required for mid-size airports. The Horizon 2020 funded project MALORCA offers machine learning framework as a potential solution. Each Air Navigation Service Provider generates Mega Bytes or even Giga Bytes of radar data and voice recordings on a daily basis.

More than 100 hours of radar data and utterance from controller pilot communication for Vienna and for Prague approach area were recorded (WP2). 20% of these speech recordings were manually transcribed, i.e. an ATC expert listens to the controller-pilot communication and writes down word by word what was said and what are the ATC relevant elements: "good morning" e.g. is not relevant to the final comment. We need to correctly recognise the callsign, whether we have a DESCEND or a REDUCE command and the command value e.g. 8000 feet or 220 knots. All in all we transcribed four hours of voice recordings (excluding silence) for both Vienna and Prague.

An initial basic ABSR was set-up for both Prague and also for Vienna (WP3). We achieved a command recognition rate of approximately 80% respectively 60%. We added then 25% of the untranscribed data to improve the models through machine learning framework. The system performance has significantly increased. We have then plugged in another set of untranscribed data (to reach 50%, 75% and 100% of the total set) in order to emulate the learning effect on monthly basis. Command recognition rates have eventually increased to 92% respectively 83%.

In order to reach these very positive results, full-fledged Assistant Based Speech Recognition is required. A normal Speech Recognizer transforms the speech signal recorded by a microphone to a sequence of words. Assistant Based Speech Recognition developed by DLR, Idiap and Saarland University uses the output of an Arrival Manager to predict a set of controller commands which are possible in the current situation (i.e. called as situational context), where radar data is used as a second sensor. This approach can first significantly reduce the search space of the speech recognizer, correct the ASR hypotheses and can also be used for plausibility checking of the output of the speech recognizer. For Prague approach, the developed ABSR yields the command recognition error rates below 0.6% and for Vienna below 3.8%. Prague results are generally better than Vienna results (especially due to better audio quality, SNR 5 dB difference).

MALORCA developed a very novel approach, which was not applied before in Automatic Speech Recognition domain (WP4) since MALORCA's learning algorithms can rely on two independent information sources: (1) acoustic scores, which are then combined with (2) scores extracted from the situational context, provided by radar data available for each automatically transcribed utterance.

The performance of the trained ABSR system was evaluated on proof-of-concept trials by nine controllers in Vienna and Prague in end of January 2018 (WP5). These trials overstep the initial objective and allow the end-users, the controllers, to put their hands on the live-mode platform with basic HMI. The performed work does not cover only the objectives of MALORCA projects to develop a basic adaptable ABSR system and to improve it by unsupervised learning, but it goes beyond and provides the clear heritage of MALORCA project to SESAR2020 project PJ.16-04. Received feedback of end-users together with an Operational Concept Document [1] and System Requirement Specification from WP1 clearly specifies controllers' preferences in the domain of speech recognition. Results of MALORCA project were discussed during two Stakeholder workshops (Prague April 2017 and Vienna February 2018) with 58 resp. 38 stakeholders from industry, research and Air Navigation Service Providers from Europe and US.

In contrast to AcListant® project and other speech recognition exercises, MALORCA project has moved out from the laboratory environment. It approaches the daily situations by using recorded data directly from the operational rooms in Prague and Vienna.

As the MALORCA team comprises experts from ATM industry and research as well as from ATM and speech signal processing and machine learning, several challenges were discovered and put already on table in this early stage e.g. dealing with 8 kHz sampling rate instead 16 kHz, tackling noisy speech environment (i.e. low speech to noise ratio), quite frequently used deviations of ATC controllers from standard phraseology, etc.

Nevertheless MALORCA was a first demonstration of the potential of (semi) automatic adaption of controller assistance tools by learning from the large amount of audio and radar data recorded in daily manner in the ops rooms. Automatic adaptation of a generic Arrival Manager by machine learning algorithms could be a next step. Further research is required to verify or adapt the approach to other airports than Vienna and Prague. More complex approach areas like London exists. More than 200 airports with annual movement numbers between 50,000 and 200,000 are distributed all over the planet. Google exploits much larger datasets (they claim more than 200 thousand hours of transcriptions for improvement of its speech recognition tasks). MALORCA relies on less than 50 hours of in-domain data. Pilot readback and other sensors like mode-S or controller inputs into electronic flight strips need be integrated into the learning algorithms. Estimation of confidence scores related to reliability of the recognized commands is still a challenging task not properly addressed in MALORA.

MALORCA proved for Prague and Vienna approach area that unsupervised learning is able to notably improve command recognition rate and that automatic learning from radar data and voice recordings can reduce costs of data, speeds up development and reduce manual adaptation effort.

2 Project Overview

2.1 Operational/Technical Context

One of the main causes hampering the introduction of higher levels of automation in the Air Traffic Management (ATM) world is the intensive use of spoken language as the natural way of communication. Data link will be another media of communication with its known advantages and disadvantages, but it is assumed that data link will never fully replace voice communication. In 2015, the project AcListant® has achieved command error rates below 1.7 % based on Assistant Based Speech Recognition (ABSR). It was validated that not only significant controller workload reductions were possible, but also significant improvement for ATM efficiency (e.g. 50 € to 65 € less fuel consumption per A320 flight for validation trials for Dusseldorf Approach Area). One main issue to transfer ABSR from the laboratory to the ops-rooms is its costs of deployment. Currently each ABSR model must manually be adapted to the local environment due to e.g. different accents and deviations from standard phraseology.

MALORCA has proposed a general, cheap and effective solution to automate this re-learning, adaptation and customisation process to new environments, taking an advantage of the large amount of speech data usually available in the ATM world. Speech recognition is a first show-case to use machine learning framework to tailor general assistance systems to the local needs, which often is a significant cost factor preventing introduction of assistance systems.

Demonstration airports for MALORCA's approach are Vienna and Prague. The available data concentrates on the BALAD sector and the feeder position in Vienna. For Prague the Arrival Executive Controller (AEC) and the Director Executive Controller (PEC) are considered. No runway configuration changes are modelled, i.e. the data is limited to runway 24 (Prague) respectively runway 34 (Vienna inbounds) and runways 29 and 34 (Vienna outbounds). Speech recognition commands are modelled for arriving and departing traffic in the air only, i.e. no clearances by tower/ground- or delivery (issuing enroute-clearances) -controllers are considered.

All radar recordings are based on Asterix CAT 62 format, whereas the speech recordings are 8 kHz mono with low Signal-to-Noise (SNR) ration. [1]

2.2 Project Scope and Objectives

The objectives of the MALORCA project can be subdivided into three categories:

1. ASR and ABSR related objectives

[1] In proposal and in grant [46], [47] MALORCA project team still assumed that 16 kHz data (as in AcListant® project) is available and the SNR is better than 20 dB for all recordings. Although these assumptions needed a modification during the first six months of the project lifetime, MALORCA has achieved its objectives and developed more robust algorithms than promised in the grant.

2. Machine learning related objectives
3. Knowledge sharing related objectives, i.e. bringing together experts from multiple disciplines

Two research hypotheses are formulated

1. Unsupervised learning will improve command recognition rates and
2. Automatic Learning from radar data and voice recordings will reduce costs of manually annotating the data used for developing and updating the system, will speed up development and will reduce manual adaptation effort

2.2.1 ASR and ABSR related objectives

The ASR and ABSR (Assistant Based Speech Recognition) related objectives can be (according to the grant agreement [47]) further detailed into

- Provide tools to develop ABSR system for different deployment areas
- Proof of concept and roadmap for automatic adaptation of ASR in ATC
- Show cases are Prague and Vienna
- Improvement of command recognition and error rates by machine learning

Automatic or manual adaptation of ABSR to new deployment areas requires an initial basic system. MALORCA has developed an **acoustic model** and an **initial language model** for both Vienna and Prague approach area, a **context integration system** including the **text-to-concept** and **concept-to-command** modules.

Moreover, tools were further developed which process available radar data, evaluate performance with respect to different metrics (e.g. command prediction error rates, command recognition rates, command recognition error rates, response times, recognition time, linking these partially proprietary rates to standard information retrieval task – precision/recall). The basic ABSR system has been implemented in one year and was used to boost the automatic learning system, allowing exploiting additional data (i.e. untranscribed controller utterances).

The following learning curve depicted in Figure 1 (see D5-2 for more details [22]) shows that the basic ABSR system is able to offer initial command recognition rate of 79.8% (for Prague approach), which was later improved (by running an iterative framework by adding more and more untranscribed learning data) to 91.9%. For Vienna Approach Area we started with 60.0% before the learning process and ended by adding 18 hours of untranscribed data at 85.2%. The proof-of-concept trials were conducted in Prague (23/24 January 2018) and Vienna (30 January 2018). MALORCA's results are used in PJ.16-04. To conclude: All ABSR related objectives defined in the grant agreement are fully fulfilled.

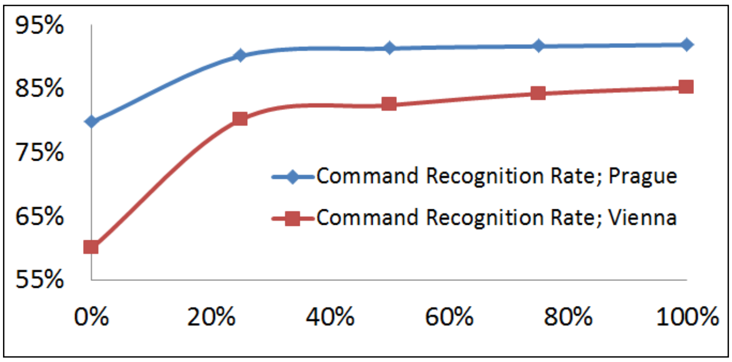

Figure 1: Learning Curve of Command Recognition depending on amount of provided untranscribed training data for Prague and Vienna

2.2.2 Machine learning related objectives

The (machine) learning related objectives can be further detailed into

- Develop a multi-modal, state-of-the-art, automatic learning system for ABSR
- Automatic Learning from radar data and voice recordings
  - To reduce costs of data
  - To speed up development
  - To reduce manual adaptation effort

Three models have been learned

- The acoustic model
- The language model and
- The command hypotheses prediction model

Automatic transcription of all speech data sets is available (approx. 22 hours of clean speech for Vienna and for Prague), i.e. the output of the speech recognizer without manual interaction. The automatic transcriptions are automatically filtered, if they are possible in the current situation (e.g. is recognized callsign existing, DESCEND only seldom for outbounds etc.). All automatic command transcriptions which pass the filter are used to develop the first command hypotheses prediction model. 4 hours of manual transcriptions are available for Prague and Vienna. The developed evaluation set is then used to compare automatic transcriptions with the manual transcriptions to calculate command recognition and command error rates. The plausibility checker (evaluating command hypotheses with respect to a situational context generated by command prediction model) is then used to filter out possible false recognitions (i.e. false positive commands). This results in a reduction of the command recognition error rate from 10% to 2% (for the cost of rejecting commands which were correctly recognised). More details can be found in common MALORCA paper submitted to SID 2017 in Belgrade [41] and in D4-4 [19].

A lot of work was done to extend the first approaches, usually working on top of 1-best hypothesis generated by basic ABSR system, to support N-best hypotheses output. The second part of the work is related to extracting robust confidence values, combining acoustic scores with scores extracted from command hypotheses predictor (relying on a situational context). We see this as a very novel approach, which was not applied before in ASR domain since our algorithms can rely on two independent information sources offering complementary results. This work was presented at ASRU conference in 2017 [42].

To conclude: MALORCA has spent a whole work package on developing an initial set of 2.5 hours for Vienna [6], [7] and four hours for Prague [8], [9]. Although the effort for one hour of transcription is not exactly measurable anymore at the end of the project it was more than 20 hours for one hour of high quality transcription including the time consuming word to command transcription.. Automatic annotation only requires an effort of two to five hours and some hours (< 10) of computer runtime. Even though the quality of the transcripts does not really reach the level of those transcribed manually, it is sufficient to use these automatic transcripts in a machine learning framework to increase recognition rates from 80% to 92% resp. 60% to 85% (for Prague and Vienna approaches respectively).

2.2.3 Knowledge sharing related objective

Many experts from Idiap, USAAR, DLR and the two ANSPs are involved. Telcos between all five partners took place every 2 weeks. A first Stakeholder workshop was organized with 58 participants from 26 different organizations in Prague in 2017. The follow-up workshop was organized in Vienna in February 2018.

The gap analysis with respect to command transcription was used by PJ.16-04 in the ASR activity. Ontology for command transcription was enabled for PJ.16-04 built on top of MALORCA's problem analysis. The requirements of MALORCA project were reused in PJ.16-04 Functional Requirement Document (FRD). A COOPANS workshop in Braunschweig in January 2017 was organized by Austro Control to demonstrate benefits of Assistant Based Speech Recognition to the COOPANS stakeholders.

2.2.4 Summary

The research hypothesis "Unsupervised Learning will improve command recognition rate" is validated and improvements are quantified for Prague and Vienna approach. The second hypothesis that "Automatic Learning from radar data and voice recordings will reduce costs of data, speed up development and reduce manual adaptation effort" is also validated for Prague and Vienna approach (see also sect. 3.2 with the maturity assessment).

Regardless of the unexpected data quality (relatively noisy and only 8 kHz recordings) and quantity issues (using only 45 hours of speech data) that we face in MALORCA, all objectives mentioned above are fully achieved.

Additionally, (to the grant) MALORCA developed algorithms to exploit 16 kHz recorded speech data with 8 kHz recorded acoustic data. Large out-of-domain speech data sets (mostly transcribed) are available and were integrated in the recognition models of MALORCA which was not planned when signing the grant agreement. However the data recorded in the ops rooms which we currently (and in the near future) get are 8 kHz noisy recordings. Algorithms (not foreseen in grant agreement) developed in MALORCA for these challenges enable the use of the large data sets already available at ANSPs side without investing into new voice recording techniques.

Although all objectives of MALORCA are achieved, we do not conclude that research on Assistant Based Speech Recognition and Machine Learning should stop now. The opposite is the case. Research was able to deliver and influence already in the middle of the project industrial research. Suggestions for next steps are further elaborated in section 4.3 "Plan for next R&D phase (Next steps)".

2.3 Work Performed

As part of the MALORCA project, first and foremost, we concentrated on developing the engineering requirements (Task 1.1, 1.2 and 1.3).

As Task 1.1, we defined the use cases in an Operation Concept Document (D1.1) in connection to speech recognition application in ATM environment with the main focus on automatic learning of speech recognition parameters. This task also identified the current operating environment from the view of the end users.

As Task 1.2, we defined the system requirements (D1.2). Functional requirements defined in this process specified what the MALORCA system is assumed to accomplish. As part of this initial preparation, we constructed an architecture design document (D1.3) specifying and defining the system architecture and the needed tools for data preparation, data transcription and proof-of-concept. These three tasks setup the MALORCA system as given in Figure 2.

The Figure 2 below gives the final schematic overview of the complete MALORCA ABSR system. The four text boxes at the bottom: DATA, TEXT, COMMAND and USER are namely the major conceptual modules tackled in the MALORCA system. In this section, we first describe these major concepts of the MALORCA system and then classify the work done as part of these components.

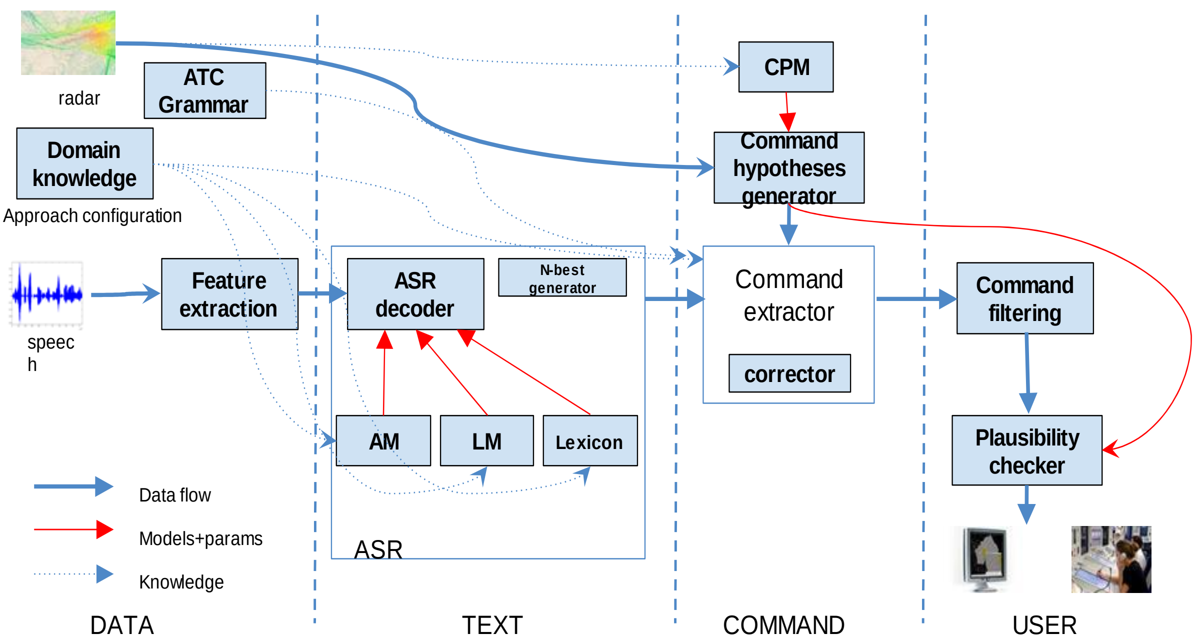

Figure 2: Schematic overview of the MALORCA system. This figure shows the four important concepts of the MALORCA system.

Work performed to establish the DATA conceptual module of the system consists of the data collection process. Data, which is used as the input to the MALORCA system, includes the formation of the Grammar as specified by ICAO grammar rules (and later phraseology adaptations); domain knowledge containing specific information about each airport; speech data collected by ANSPs and features extracted on speech audio. We describe the various efforts made in the DATA conceptual module in Section 2.3.1.

TEXT conceptual module of the system consists of the research and development work performed to convert the features extracted from acoustic data to TEXT. This conceptual module describes the Assistant Based Speech Recognition system consisting of the various models like acoustic model, language model and the pronunciation dictionary in the form of Lexicon. These sub-modules are combined in an automatic speech recognizer decoder and then used by an acoustic hypotheses generator to output the raw text. Further details on the work packages completed in this module are described in Section 2.3.2.

COMMAND conceptual module is used to convert the raw text or speech transcriptions obtained from the previous conceptual module to air traffic commands. Important blocks used in the COMMAND module are the Command Predictor Model (CPM), Command Hypotheses Generator and Command Extractor and corrector engines. The Command Hypotheses Generator with the help of the CPM trained using the context files generates a set of commands which are then used in the Command Extractor Module to transform the acoustic (text) hypotheses into command hypotheses, and to potentially correct them based on a situational concept. The details of the research and development efforts made in this module are described in the section 2.3.2.2.

USER conceptual module makes a two-fold contribution. It consists of the work packages completed to produce the Command Filtering and Plausibility Checker blocks; and the work done to test and analyse the MALORCA system with the help of air traffic controllers. We describe this conceptual module later in Section 2.3.4.

2.3.1 DATA

The collection of the audio and subsequent transcription of portion of these audio forms are part of this conceptual module, mainly covered in tasks WP2.1 to WP2.6 and delivered as part of D2.1 to D2.4. This module also includes some initial feature extraction carried out as part of acoustic modelling necessary for the speech recognition system. So a part of this feature extraction process forms work package tasks to train acoustic models, specifically in WP3.1.

Major effort in WP2 was spent in collecting the audio data from the Prague and Vienna approaches (WP2.1 and WP2.4) and then transcribing partial portions of this collected set (WP2.4 and WP2.5). For transcription, ANSPs used a tool provided by USAAR to transcribe the audios. Both USAAR and IDIAP supported training for this tool. DLR then prepared data for automatic learning for Vienna and Prague approach for facilitating building speech recognition system as part of WP3.

Final deliverables as part of the DATA conceptual module, made available transcribed data and complete non-transcribed data collected as part of WP2.

2.3.2 TEXT

Speech recognition systems are generally based on two main separate components, namely, 1) the **acoustic model** to learn and model the phonetic and other acoustic aspects of speech, and 2) the **language model** to capture the structure of a language by learning the possible/allowed sequences of words and their distribution in a given context. The lexicon, a pronunciation dictionary, is composed and is strongly aligned with the acoustic model. This task is performed as part of the WP3.1. Also, for the N-best acoustic hypotheses generator (i.e. word recognition lattices), we utilize a similar scheme as the AcListant® project.

2.3.2.1 Acoustic Model

The acoustic models generated as part of this module for the MALORCA speech recognition is generated in two iterations. First, a simplistic version of the acoustic model is created using conventional techniques like HMM/GMM as part of WP 3.1. This simplistic version is built using the audio data collected in the DATA conceptual module.

In the next iteration of task 3.1, deep learning techniques are used to develop a sophisticated and more advanced version of the (speaker-dependent) acoustic model. These two iterations plus some initial feature extraction form the complete task 3.1. The deliverables produced out this work package are namely: D3.2 and D3.4.

Further enhancements are carried out leveraging out-of-domain data. As shown in [40], we can harness this out-of-domain audio data to build an initial version of the acoustic model and then adapt with data collected in MALORCA to obtain performance boosts in speech recognition. This learning process for acoustic modelling allows training the model independently of an expert. In a deployed version of the system, these algorithms remain important to updating the acoustic model as more and more data is collected by deployed system. These learning algorithms are delivered across three deliverables D4.1, D4.3 and D4.5.

2.3.2.2 Language Modelling

Air traffic control tasks are generally performed based on a standardized phraseology which allows an easy, clear and safe communication between the controllers and the pilots. These standard command specification protocols tell the controllers and the pilots how to issue/exchange flight related information. An initial version of the language model is developed as an optimal context-free grammar modelling this standard phraseology. This context-free grammar serves as baseline, which is learned when experts manually design rules to be used as a language model based on available transcriptions. However, the manual transcriptions form only a portion of the speech data collected and do not cover many possible events that are never transcribed. Thus, in MALORCA, using this grammar-based language model results in an inflexible system and hence, yields relatively low word recognition performance. Development of this grammar-based model is carried out as part of WP3.2 and delivered as D3.3.

To further improve the language modelling, we get motivated by more statistical approaches as part of WP4.2. As a first iteration, we have developed n-gram statistical language models which can deal deviations which cannot be easily modelled in context-free grammar. Still such models offer only estimates, based on what has been seen in the transcriptions. In our experiments, we have found out that these relatively simple statistical language models are able to significantly outperform (decreases the command error rate by 50% relatively) on the test data as reported in deliverable D4.3 and D4.5.

These statistical language models are further enhanced in their capabilities to encounter unseen events with help of a data augmentation technique developed as part of the automatic adaptation task for language modelling in WP4.2. Using a two-fold approach, we first build a concept-based class language model, which helps capturing events in the concept space. Intuitively, it is assumed that even if a specific command is not seen in a certain form in the training data, the corresponding concepts can occur more often. And hence, this class-based language model helps alleviate the problem of unseen events in the concept space. Second, as there is a lack of training data, we sample large amounts of data from these class-based models to create artificial data and to better model events not present in the training data. As we use this artificial data for training, we further improve the performance of the system in comparison to a simple statistical language modelling approach [19]. Above language modelling approaches are implemented as part of WP4.2 and WP4.5 and delivered as part of the D4.3, D4.4 and D4.5.

2.3.3 COMMAND

As part of MALORCA, the COMMAND conceptual module requires the implementation of external context integration (task 3.3) and uses this situational context to predict plausible commands at a certain time (task 3.4). Also, we have designed a conversion of raw text to commands as part of the Context Extractor (task 4.3) and hence, allow comparing these commands with the commands available after plausibility checker. Additionally, we have developed a command prediction model (task 4.4) to support command hypotheses generation as done in task 3.4.

The command generator engine is implemented in two iterations. First, to aid generation of initial dataset for automatic model improvement in task 2.3 and task 2.6 and then the second iteration is used to generate efficient command hypotheses that can be used in the ABSR as described in task 3.3. These two iterations are delivered as part of D3.1 and D3.6. Alongside, we also develop Context corrector part of the COMMAND conceptual module and deliver it as part of the D3.5.

In designing the next generation Context Extractor (task 4.3), we describe in detail how we can further reduce the dependency on experts within MALORCA project. The present Context Extractor engine is maintained and updated by experts as a rule-based system. Hence, continuous adaptation and management of this resource requires expert help, which in a deployed system leads to additional costs. By designing a statistical system to learn the context extraction, we can remove the dependency on an expert. Also learning under statistical framework is more expressible than under a rule-based framework. The design of context extractor engine is delivered as part of D4.2.

Furthermore, we learn to predict hypotheses generation given the ATC radar situational context in a context prediction model. This model is implemented and delivered as part of D4.3, D4.4 and D4.5.

2.3.4 USER

This conceptual module of the system involves two parts, where we first test the system on the recorded test data and then test the system with controllers from Prague and Vienna at a demonstration trial. These constitute the work package 5.

To enable testing on the recorded data, further data is recorded in task 5.1, employed to evaluate the MALORCA system. MALORCA system reveals approximately 15 – 20% absolute improvements in recognition rates as shown in Figure 1 as part of the task 5.3.

To perform the proof-of-concept trials with controllers, a proof-of-concept plan is created in task 5.4 and delivered as D5.1. The trials are organized at the two participating ANSPs with the help of the controllers from respective approach areas. Controllers are asked to actively use a prototype of the system and then provide a feedback through a questionnaire and personal interview. Full reports of these experiments are presented in the Gap Analysis Report in D5.2 which also contains an analysis of the further improvements to the system.

In addition to the tasks performed above, we also respond to the ethical concerns by publishing the ethical requirements in D7.1 to D7.7. Reports on the implementation of these requirements are the annual ethical reports delivered in D6.3 and D6.4.

The above-described work was detailed in the Project Management Plan (PMP) D6.1 and the results of the two stakeholder workshops are published as D6.2.

2.4 Key Project Results

MALORCA has focused on realizing the objectives set out in the Section 2.2. To achieve these objectives, we have presented important project related results. This section summarizes these key results.

With AcListant®, the key project result is the proof of concept (Speech Recognition System for an ATM environment) that can be built with the help of experts that lead to superior operation quality for ATMs. In contrast, MALORCA has shown that with the help of Machine Learning framework such a system can be transformed into an automated system, which needs minimal expert support. Figure 3 also summarizes this key result.

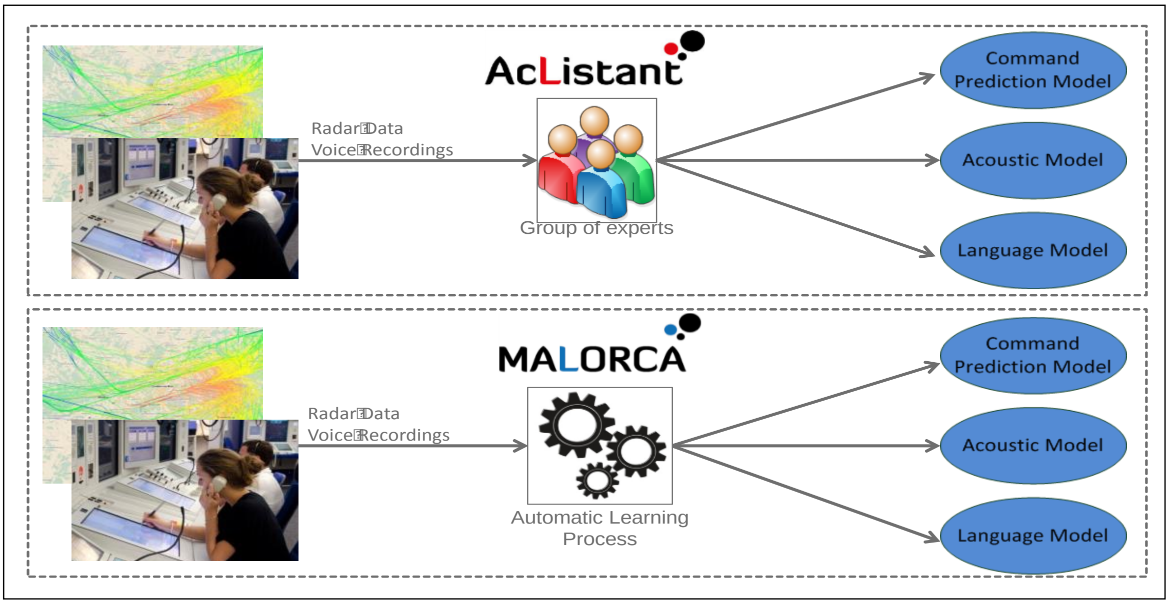

Figure 3: Diagram comparing the key difference between AcListant and MALORCA project systems.

MALORCA developed a very novel approach, which was not applied before in Automatic Speech Recognition domain since MALORCA's algorithms rely on two independent information sources: (1) acoustic information is enhanced with (2) the information extracted from command hypotheses prediction (employing situational context). For each automatically transcribed utterance, the output of a Command Hypothesis Predictor is available, resulting in a limited set of possible commands in each situation. Moreover, in MALORCA, we have shown that even with small amounts of transcription data collected we can achieve close-to 92% of command recognition rate (while command recognition error rate is below 0.6%, refer Figure 1). The machine learning algorithms developed to achieve this performance are the backbone of this project. In terms of human effort, these machine learning algorithms have roughly brought down the transcription effort (to transcribe one hour of ATC speech) from tens of hours to 2 hours. This result can be considered as a key-result to possibly port the ABSR system in ATM operation rooms.

MALORCA has focused on implementing a robust and high-quality solution for ABSR in ATM environment while minimizing cost and resource requirements during deployment time. Another output of this process is a definition of framework allowing to model and measure deviations in phraseology from the controller's side. Such a framework might be an important thing in understanding and measuring a controller's workload (i.e. allowing to better performing work management and optimization in their daily tasks). Future versions of such system will benefit from the MALORCA's system design. SESAR 2020 PJ.16-04 has already benefited by extending the framework to an ontology for controller/pilot command transcription.

2.5 Technical Deliverables

Reference	Title	Delivery Date[2]	Dissemination Level[3]
Description			
D1.1	Operational Concept Document	20/02/2017	Public
This document focuses on possible implementation of speech recognition systems in air traffic control environment with special emphasis on Vienna and Prague needs. The Operation Concept Document including the use cases was the input for the requirements document D1-2.			
D1.2	System Requirement Document	23/09/2016 [4]	Public

[2] Delivery data of latest edition

[3] Public or Confidential

[4] SRS was a living document during the life time of the MALORCA. The final version 3.00 was created 23/02/2018 and sent to SJU. The previous version 2.00, approved by SJU, was created 23/09/2016.

Reference	Title	Delivery Date[2]	Dissemination Level[3]
Description			
This document collects and describes the general technical requirements of a controller support tool, based on speech recognition in ATM environment. Only a subset of the requirements was addressed by MALORCA project.			
D1.3	Architecture Design Document	21/11/2016	Confidential
The Architecture Design Document describes the technical aspects of MALORCA project related to possibilities and efforts to integrating blocks of automatic speech recognition and machine learning into existing ATM systems. Besides describing architecture of ABSR, machine learning and command hypotheses predictor, D1.3 also describes data formats which can be used to exchange information among all modules, so-called data elements.			
D2-1	Transcribed Data Set for Vienna Approach Area	14/06/2017	Confidential
D2.1 describes the work of preparation, recording of voice data, transcribing of the data and processing by annotator tool for the Vienna Approach Data-Set. In grant agreement the data set itself was agreed between MALORCA project and SJU. In addition to that, the project team also created a short report describing the data.			
D2-2	Complete non-transcribed data set for Vienna Approach Area	21/06/2017	Confidential
D2-2 describes the contents of the complete Vienna data set of radar and speech data, containing approx. four hours of fully transcribed data and 18 hours of untranscribed (no voice to text transformation) data. In grant agreement the data set itself was agreed between MALORCA project and SJU. In addition to that, the project team also created a short report describing the data.			
D2-3	Deliverable Report for Transcribed Data Set for Prague Approach Area	13/01/2017	Confidential
D2-3 report brings description of work and introduction of recorded and delivered data of Prague Approach Area processed by annotator tool. In grant agreement the data set itself was agreed between MALORCA project and SJU. In addition to that, the project team also created a short report describing the data.			
D2-4	Complete Non-Transcribed Data Set for Prague Approach Area	14/03/2017	Confidential
D2-4 brings description of work and introduction of delivered data processed by Annotator tool and the set of predicted commands for Prague Approach Area. In grant agreement the data set itself was agreed between MALORCA project and SJU. In addition to that, the project team also created a short report describing the data.			
D3-1	Command Hypotheses Predictor, first iteration	30/11/2016	Confidential
D3-1 describes the first prototype of the command hypotheses predictor and its application to Vienna and Prague approach area. Additionally reduction rate and error rate of the predicted commands of			

<table>
<tr><th>Reference</th><th>Title</th><th>Delivery Date[2]</th><th>Dissemination Level[3]</th></tr>
<tr><th colspan="4">Description</th></tr>
<tr><td colspan="4">the current prototype status are evaluated.
In grant agreement the software itself was agreed between MALORCA project and SJU. In addition to that, the project team also created a short report describing the software and its performance.</td></tr>
<tr><td>D3-2</td><td>General acoustic model, first iteration</td><td>27/02/2017</td><td>Confidential</td></tr>
<tr><td colspan="4">D3.2 presents first results of acoustic modelling in ATC automatic speech recognition: Two types of English acoustic models, namely HMM/GMM and HMM/DNN were developed. Hence, these two models are the basic components that initialized the automatic leaning of the acoustic model adaptation of WP4. The deliverable briefly presents the implementation together with recognition performance.
In grant agreement the software itself was agreed between MALORCA project and SJU. In addition to that, the project team also created a short report describing the software and its performance.</td></tr>
<tr><td>D3-3</td><td>General, Context-free Grammar Based Language Model, first iteration</td><td>27/02/2017</td><td>Confidential</td></tr>
<tr><td colspan="4">D3-3 describes the first prototype of the basic language model (primarily its grammar-based version) and its application to Vienna approach data. Word Error Rates of speech recognition, as well as error rates for concept and command extraction are evaluated. An approach based on statistical language model is combined with the context-free grammar to examine a promising alternative to the pure grammar-based approach.
In grant agreement only the grammar-based approach and here the software itself was agreed between MALORCA project and SJU. In addition to that, the project team also created a short report describing the software and also developed the statistical language model.</td></tr>
<tr><td>D3-4</td><td>Neural Network based acoustic model, second iteration</td><td>06/06/2017</td><td>Confidential</td></tr>
<tr><td colspan="4">D3.4 presents the second deliverable devoted to acoustic modelling and describes further development of a hybrid (deep learning based) acoustic models towards MALORCA environment, and acts as an important step towards WP4 to use D3-4 models for unsupervised and semi-supervised adaptation. The deliverable also presents current recognition performance.
In grant agreement the software itself was agreed between MALORCA project and SJU. In addition to that, the project team also created a short report describing the software and its performance.</td></tr>
<tr><td>D3-5</td><td>Context Integration into ASR system</td><td>13/06/2017</td><td>Confidential</td></tr>
<tr><td colspan="4">D3.5 presents the work on context integration into the ABSR system, i.e. using the actual airspace conditions which are processed and transformed by the command hypotheses predictor into a set of meaningful and possible commands that can be given by the controllers to the pilots. The goal of context integration is to exploit this information during speech recognition and thereby improving command recognition.
In grant agreement the software itself was agreed between MALORCA project and SJU. In addition to that, the project team also created a short report describing the software and its performance.</td></tr>
<tr><td>D3-6</td><td>Improved Context provided by Command Hypotheses Prediction, 2nd iteration</td><td>28/03/2017</td><td>Confidential</td></tr>
</table>

Reference	Title	Delivery Date[2]	Dissemination Level[3]
	Description		
D3.6 describes the concept of the command hypotheses predictor which was used for improvements by automatic learning. This current command hypotheses predictor reduces the command recognition error rate from 15.4 to 1.0% for Prague and from 19.7 to 4.2% for Vienna Approach Area. The average context size for Prague is below 650 commands and for Vienna below 1600 commands. In grant agreement the software itself was agreed between MALORCA project and SJU. In addition to that, the project team also created a short report describing the software and its performance.			
D4-1	Application of model learning algorithms (initial version for acoustic model)	06/06/2017	Confidential
While ATC uses standard phraseology and a limited vocabulary, D4.1 provides initial solution to adapt the speech recognition systems to local acoustic conditions and vocabularies at target airport to reach higher performance. Due to continuous operation of ATC systems, a large and increasing amount of untranscribed speech data is available, allowing for semi-supervised learning methods to build and adapt ASR models. Different methods are explored on data selection for adapting baseline acoustic and language models by exploiting the continuously increasing amount of untranscribed data. Initial results on relative improvements in word error rates, concept error rates and command error-rates are presented, when adapting ASR models to different ATC conditions in a semi-supervised manner.			
D4-2	Concept Extractor Definition	08/11/2017	Confidential
D4.2 presents the definition of the post-recognition steps and approaches needed to transform the plain text ASR hypotheses into an abstracted ATC representation that can be presented to following systems or to the controller. That is, after recognizing the sequence of spoken words issued by the controller to a pilot, we need to extract the ATC information carried by this sequence and then transform it to the command level using concept extractor. We also present results on automatic adaptation of the ABSR engine as it was intended in the MALORCA project.			
D4-3	Acoustic, language and hypotheses rule model adaptation for Vienna approach	22/12/2017	Confidential
D4-3 briefly presents the performance of the generic MALORCA ABSR engine when adapted (in a semi-supervised way) to the Vienna approach area. Detailed description of algorithms and use of data is given in the deliverable D4-4. In grant agreement the finalization of a demonstrator was agreed between MALORCA project and SJU. In addition to that, the project team also created a short report describing the performance of the demonstrator.			
D4-4	Application of model learning algorithms (final version)	22/12/2017	Confidential
D4-4 describes the concept - how the necessary ABSR models (acoustic model, language model, controller command prediction model) can be automatically created from recorded transcribed/untranscribed controller utterances and the corresponding radar data. The trained models are validated against transcribed controller commands for Vienna and Prague approach. D4-4 describes recent findings, proposing to adapt all three components of the ABSR system. D4-4 does not present the current best performance of the ABSR system. Instead, we analyze each of the three modules independently w.r.t. capability to learn from un-transcribed data. The combination of all			

Reference	Title	Delivery Date[2]	Dissemination Level[3]
Description			
three modules trained using semi-supervised training is rather given in the following deliverables of final work-package 5.			
D4-5	Acoustic, language and hypotheses rule model adaptation for Prague approach	22/12/2017	Confidential
D4-5 briefly presents the performance of the generic MALORCA ABSR engine when adapted (in a semi-supervised way) to the Prague approach area. Detailed description of algorithms and use of data is given in the deliverable D4-4. In grant agreement the finalization of a demonstrator was agreed between MALORCA project and SJU. In addition to that, the project team also created a short report describing the performance of the demonstrator.			
D5-1	Proof-of-Concept Plan	04/08/2017	Confidential
D5-1 details the proof-of-concept plan split into two technical (T1, T2) and two operational (O1, O2) activities. T1 is a workshop with technical experts to evaluate the ABSR prototype implementation against the technical requirements document D1-2. T2 is an offline evaluation and quantifies the improvements of the ABSR system with respect to the amount of available training data. O1 involves controllers who concentrate only on the different outputs of a baseline ABSR system and on an ABSR system trained with all the available MALORCA training data. O2 puts the trained ABSR system in a simulation environment with a replay of historic radar data and controller voice recordings from real Prague and Vienna. ABSR is used here to support the controllers in maintaining radar labels.			
D-5-2	Gap Analysis Report	28/02/2018	Public
MALORCA's Proof-of-Concept is split into two technical (T1, T2) and two operational (O1, O2) activities. T1 is a workshop with technical experts to evaluate the ABSR prototype implementation. T2 is an offline evaluation to quantify the improvements of the Assistant Based Speech Recognizer (ABSR system) with respect to the amount of available training data. O1 involves controllers who concentrate only on the different outputs of a baseline ABSR system and a trained system. O2 puts the trained ABSR system in a simulation environment with a replay of historic radar data and controller voice recordings from real Prague and Vienna. ABSR is used here to support the controllers in maintaining radar labels. D5-2 first presents the outcomes of the four activities, and the controller's feedback in debriefing sessions and finally analyses the results with respect to gaps and further challenges.			
D5-3	Final Project Results Report	30/04/2018	Public
This document, see abstract at page 5.			
D6-1	Project Management Plan	12/12/2016	Confidential
D6-1 presents the Project Management Plan (PMP) of the MALORCA project, which was detailed during project kick-off meeting in April 2016 in Braunschweig at DLR and updated several times.			
D6-2	Report on Stakeholder Workshop Results	26/02/2018	Public

Reference	Title	Delivery Date[2]	Dissemination Level[3]
Description			
The first stakeholder workshop was conducted in April 2017 with 58 participants. The second stakeholder workshop took place in Vienna in February 2018. The workshops consisted of presentation to the stakeholders and more important different and parallel working groups with a limited number of participants. The outputs of the working groups are summarized in D6-2.			
D6-3	Ethical Report 1	28/04/2017	Confidential
MALORCA project proposes a general, cheap and effective solution to automate re-learning, adaptation and customisation process of speech recognition models to new environments. For this purpose speech and radar data of controllers are recorded. At least recording speech data touches ethical issues. The implementation of the ethical issues with respect to the security of data privacy of the MALORCA project was described in seven ethical deliverables D7-1 to D7-7. D6-3 summarizes the usage of D7-1 to D7-7 during the first year of the MALORCA project.			
D6-4	Ethical Report 2	17/02/2018	Confidential
MALORCA project proposes a general, cheap and effective solution to automate re-learning, adaptation and customisation process of speech recognition models to new environments. For this purpose speech and radar data of controllers are recorded. At least recording speech data touches ethical issues. The implementation of the ethical issues with respect to the security of data privacy of the MALORCA project was described in seven ethical deliverables D7-1 to D7-7. D6-4 summarizes the usage of D7-1 to D7-7 during the second year of the MALORCA project.			
D6-5	Period Technical Report 1, Part B	17/01/2017	Confidential
The first periodic report summarizes the outputs of the first eight months of the MALORCA project and gives an outlook of the planned work for the remaining months of the project.			
D6-6	Period Technical Report 2, Part B	27/07/2017	Confidential
The second periodic report summarizes the outputs of the MALORCA project from October 2017 to April 2017 and gives an outlook of the planned work for the remaining months of the project.			
D6-7	Period Technical Report 3, Part B	16/11/2017	Confidential
The third periodic report summarizes the outputs of the MALORCA project from April 2017 to October 2017 and gives an outlook of the planned work for the remaining months of the project.			
D6-8	Period Technical Report 4, Part B	May 2018	Confidential
The last periodic report summarizes the financial part of the last six months of the project. It will be created after the final project review meeting.			
D7-1 to D7-7	Implementation of Ethical Requirements	06/09/2016	Confidential

Reference	Title	Delivery Date[2]	Dissemination Level[3]
Description			
MALORCA project proposes a general, cheap and effective solution to automate re-learning, adaptation and customisation process of speech recognition models to new environments. For this purpose speech and radar data of controllers are recorded. At least recording speech data touches ethical issues. The implementation of the ethical issues with respect to the security of data privacy of the MALORCA project was described in seven ethical deliverables D7-1 to D7-7. The implementation of each of this seven ethical is described in these seven deliverables.			

Table 1: Project Deliverables

3 Links to SESAR Programme

3.1 Contribution to the ATM Master Plan

Code	Name	Project contribution	Maturity at project start	Maturity at project end
OI/EN code not provided, no enablers associated in the ATM Master Plan yet. Suggested enablers are industry partners and ATM-system providers with an expertise in CWP (Controller Working Position) development, automatic speech recognition and/or voice communication systems	*Associated project is SESAR2020 PJ16-04, Controller Workstation Productivity*	*Today Speech Recognition Technology has reached a level of reliability that is sufficient for implementation into an ATM-system. Several economic (e.g. increased ATCO-productivity, landing capacity, competitive advantage) and safety benefits (e.g. reduced cognitive load, improved decision making, constant ATCO performance over longer time periods) are targeted and are proven to be achievable.*	*TRL-1*	*TRL-2*
		The project furthermore provides a cheap automatic learning system to adapt the Assistance-Based Speech Recognizer and a roadmap for automatic learning to be integrated in the ATC domain. When achieved, the integration of such supporting system at different places is possible with less tuning and training effort and therefore less cost.		*TRL-1*
		The show case of adapting speech recognition models to different approach areas for applying Machine Learning techniques in ATM can be transferred for adaptation of Controller Assistance tools to different location in general (e.g. Tailoring a generic Arrival Manager to a specific airport). This will reduce software development costs – "learning instead of programming".		*Maturity assessment pending*

Table 2: Project Maturity

3.2 Maturity Assessment

The main results of the first MALORCA stakeholder workshop [25] was that MALORCA considers at least two each other influencing, but different roadmaps.

1. A(B)SR for ATM applications (ABSR roadmap) and
2. Machine learning for ATM application (ML roadmap).

The A(B)SR roadmap started 2013 with unproven ideas by AcListant® project of DLR and Saarland University [48], [49]. At the end of the project basic principles were observed, document in a patent and published at different conferences (TRL1) [50], [52], [53], [54], [55], [56]. The successor project AcListant®-Strips started the formulation of the technology concept (TRL2) and performed an experimental proof of concept (TRL3) which was however restricted to Dusseldorf approach area.

Nevertheless the possible improvements with respect to mouse inputs into the radar label (three times less clicking time for controllers, up to two arrivals more per hour, flight time reduction of 77 seconds per aircraft) were quantified and published [57], [58]. These results influenced the definition of SESAR 2020 solution PJ.16.04-ASR, i.e. industrialization of ASR activities in ATM has started. However, PJ.16-04-ASR hardly considers Assistant Based Speech Recognition and also neglects machine learning results.

MALORCA started in the ML roadmap with unproven ideas that radar data as an additional sensor will improve unsupervised learning of speech recognition models (TRL 0).

Two research hypotheses

1. Unsupervised learning will improve command recognition rates and
2. Automatic Learning from radar data and voice recordings will reduce costs of manually annotating the data used for developing and updating the system, will speed up development and will reduce manual adaptation effort

are also formulated and could be tested for Vienna and Prague. Nothing to the contrary was observed. Even more, command recognition rates increases from 80% to 92% for Prague respectively from 60% to 85% for Vienna by applying unsupervised learning. Manual transcription (voice to commands) including radar data processing and first command prediction in work package 2 needed more than 1,500 of working hours. The result was less than 2 hours for Vienna and less than 4 hours for Prague.[5] At the end of the project we got 18 hours of automatically transcribed data for both Vienna and for Prague. Adding more hours would now require just the radar recording, voice recording and splitting task.

We, however, showed it only for Vienna and Prague relying on a basic speech recognizer with acceptable recognitions rates. Further research is needed to verify or adapt the approach to other airports (see mayor gap analysis in [22]) The developed prototype uses data from the ops room, but the data needed manual improvement steps in advance.

To conclude, more than basic principles are observed and reported (TRL 1), first technology concepts have been formulated (TRL 2). The experimental proof of concept (TRL 3) with (much) more data, in a more complex environment (e.g. London, Paris or Frankfurt), considering different configuration (e.g. weather, runway configuration changes) are necessary. On the other hand the portability to many different mid-sized airports needs to addressed also.[6]

[5] Command transcription was improved in other work packages by DLR afterwards.

[6] The ICAO runway safety list from 2017 [59] contains about 200 airports which have between 50,000 and 250,000 movements per year. They cover all parts of the planet. In total these 200 airports have 23 million movements per year. The London area with five different airports just covers over one million movement per year.

Table 3: ER Fund / AO Research Maturity Assessment

ID	Criteria	Satisfaction	Rationale - Link to deliverables - Comments
TRL-1.1	Has the ATM problem/challenge/need(s) that innovation would contribute to solve been identified? Where does the problem lie?	Achieved	Adaptation costs are too expensive, if explicit manual adaptation would be needed for each new airport / approach area, see sect. 2.3 and 2.4 and D5.2 (Gap Analysis Report), [22].
TRL-1.2	Has the ATM problem/challenge/need(s) been quantified?	Achieved	Adaptation costs are too expensive, if explicit manual adaptation would be needed for each new airport / approach area, see sect. 2.3 and 2.4 and D5.2 (Gap Analysis Report), [22]. The benefits of ABSR were already quantified in AcListant®-Strips project. It is assumed that they are in the same order for Vienna and Prague as reported for Dusseldorf in AcListant®-Strips project.
TRL-1.3	Are potential weaknesses and constraints identified related to the exploratory topic/solution under research? - The problem/challenge/need under research may be bound by certain constraints, such as time, geographical location, environment, cost of solutions or others.	Achieved	MALORCA has chosen two European mid-size airports for demonstrations. It was not demonstratrated for more complex approach areas and also not for tower or enroute, see sect. 4.1 and also D5.2 (Gap Analysis Report), [22], for details
TRL-1.4	Has the concept/technology under research defined, described, analysed and reported?	Achieved	An iterative concept, assuming that previous versions of the ABSR system can be used to automatically transcribe new data, and to further use this data for re-training. Learning of command prediction model does not need any transcribed data at all, but an initial speech recognizer with a command recognition rate of at least 50%, see sect. 4.2 and also D5.2 (Gap

ID	Criteria	Satisfaction	Rationale - Link to deliverables - Comments
			Analysis Report), [22], for details
TRL-1.5	Do fundamental research results show contribution to the Programme strategic objectives e.g. performance ambitions identified at the ATM MP Level?	Achieved	Yes, a reduction of controller workload, enabling an increase of controller productivity is contribution to strategic objectives, see section 3.1
TRL-1.6	Do the obtained results from the fundamental research activities suggest innovative solutions/concepts/ capabilities? - What are these new capabilities? - Can they be technically implemented?	Achieved	Radar data will improve unsupervised learning of speech recognition models and decrease costs see sect. 4.1, MALORCA developed a very novel machine learning approach, which was not applied before in Automatic Speech Recognition domain. If relies on two independent information sources: (1) acoustic scores are combined with (2) scores extracted from radar data, see sect. 2.3.3.
TRL-1.7	Are physical laws and assumptions used in the innovative concept/technology defined?	Not Applicable	
TRL-1.8	Have the potential strengths and benefits identified? Have the potential limitations and disbenefits identified? - Qualitative assessment on potential benefits/limitations. This will help orientate future validation activities. It may be that quantitative information already exists, in which case it should be used if possible.	Achieved	Command recognition rates of 92% for Prague and 85% for Vienna are possible with ABSR; unsupervised learning can increase recognition rate from 80% to 92% resp. 60% to 85%, see sect. 2.4 and chapter 4 and Gap Analysis Report D5-2, [22]

ID	Criteria	Satisfaction	Rationale - Link to deliverables - Comments
TRL-1.9	Have Initial scientific observations been reported in technical reports (or journals/conference papers)?	Achieved	yes, presented at ASRU, Interspeech and SID in Belgrade, see reference list
TRL-1.10	Have the research hypothesis been formulated and documented?	Achieved	see subsect. 2.2.3 and sect. 3.2 and Gap Analysis Report D5-2, [22]: The research hypotheses are (1) Unsupervised learning will improve command recognition rates and (2) Automatic Learning from radar data and voice recordings will reduce costs of manually annotating the data used for developing and updating the system, will speed up development and will reduce manual adaptation effort.
TRL-1.11	Is there further scientific research possible and necessary in the future?	Achieved	yes, see sect. 4.3
TRL-1.12	Are stakeholders interested about the technology (customer, funding source, etc.)?	Achieved	two stakeholder workshops were organized, see D6-2, [25] and results of MALORCA are already used in PJ.16-04-CWP for the Automatic Speech Recognition activity, e.g. the command ontology

4 Conclusion and Lessons Learned

4.1 Conclusions

MALORCA started with unproven ideas that radar data as an additional sensor will improve unsupervised learning of speech recognition models, resulting in the two main hypotheses that (1) unsupervised learning will improve both command recognition rates and command recognition error rates and (2) automatic learning from radar data and voice recordings will reduce costs of manually annotating the data, it will speed up development and reduce manual adaptation effort. Both hypotheses were successfully evaluated in laboratory tests and with technical experts as well as with controllers from Austro Control and ANS CR.

Although technical partners of MALORCA assumed to work with 16 kHz voice recordings, we have proven that even 8 kHz voice recordings are not a show stopper in reaching the MALORCA objectives. However, noisy signal quality or bad initial command transcription as well as different interpretations of the semantics of a given controller command significantly decreases the ABSR system performance. This is clearly demonstrated by the gap in command recognitions error rate between Prague (~0.6% command recognition error rate) and Vienna (~3.2% command recognition error rate). Ontology for command transcription agreed by the main European ATM players, i.e. ANSPs and industry, is necessary, if MALORCA approach aims to be extended to other different approach areas. SESAR 16.04-ASR ontology is here a first step.

MALORCA's scope was focused on Vienna and Prague approach area, which are two European mid-sized airports. The approach needs to be extended or adapted to other approach areas including more complex European hub airports or metroplex airport areas. Tower, ground and enroute controller communication as well as pilot readback need to be included.

A way must be found to directly use the controller-pilot communication from the ops room and not from the laboratory simulation tests. This includes technical as well as ethical aspects.

Moreover, in MALORCA, we have shown that even with small amounts of transcription data we can achieve close-to 92% of command recognition accuracy. In terms of human effort, developed machine learning algorithms have significantly brought down the transcription effort. Nevertheless manual effort for pre-processing the radar data is still needed which should be significantly reduced if learning directly in the ops-room or from thousands of hours is intended. This result together with the easy adaptable basic ABSR system for approach control will be the key to developing and deploying ABSR to different approach areas.

4.2 Technical Lessons Learned

MALORCA results clearly suggest that full understanding and agreeing on ontology among ATC community is necessary. Ontology is largely related to deviations in controller-pilot communication, which makes the command recognition of real ATC speech more challenging than expected at the beginning of the project. Despite the preceding AcListant® project, which was built around laboratory environment, MALORCA deals with real data (i.e. voice recorded from real communications). As an immediate difference, we see that real communications comprise much large amount of deviations from standard phraseology than simulated data. As a consequence, statistical language modelling yields better performance than the model based on context-free grammar.

According to MALORCA proposal and grant agreement [46], [47], we have assumed that the ATC speech will be of high quality (i.e. we will be able to process 16 kHz audio with high signal-to-noise ratio). This has been the case in 16 kHz data in AcListant® project. Apparently, a direct access to raw speech recordings is not straightforward, and MALORCA had to deal with speech recordings obtained from archives (i.e. the data was passed through a telephone channel and downsampled to 8 kHz). We also assumed that pilot communication will be automatically eliminated from these archives and that the recordings contain only single utterance from a controller starting at push-to-talk pressing and ending at push-to-talk button release. As this was not the case, an additional effort was paid to pre-process the data. Although measurements of signal-to-noise ratio indicate reasonable values (~17dB and 22dB for Vienna and Prague recordings), the level of noise especially in Vienna recordings is clearly audible. Despite these challenges, MALORCA project outcomes demonstrate high command recognition accuracies.

During the development phase, we exploited an annotation tool to produce a minimum-size development set augmented with manual transcriptions. The tool for manual annotations (or corrections) of both text and command transcripts has been inherited from previous AcListant® project, and used by Vienna and Prague partners for transcriptions. The resulting transcripts were used to build the first-iteration (basic) ABSR engine. Voice recordings with manual transcriptions were further augmented by "out-of-domain" data to build an initial (basic) ABSR system. Achieved results on both Prague and Vienna data clearly suggest that combining in-domain and out-of-domain data is useful and yields better performance than the system built using out-of-domain data only. Main objective of MALORCA is to improve command recognition rates while reducing the costs on manually annotating the development data. Among other solutions, MALORCA has focused on exploiting untranscribed data to improve the basic ABSR system. More specifically, MALORCA has presented an iterative concept, assuming that previous version of the ABSR system can be used to automatically transcribe new data, and to further use this data for re-training. This concept has been evaluated on both Prague and Vienna approaches, and noticeable improvements were achieved in terms of both word recognition rates and command recognition rates. MALORCA has started with only two hours of transcribed data from Vienna and from Prague. Compared to big IT companies, processing audio (in some extreme case using more than 200,000 hours of voice recordings), our data size is very small. Luckily USAAR and Idiap were able to provide 150 hours of out-of-domain data. Transcribed in-domain data from other projects was also used, but only on word level, because each project currently has its own semantics for command transcription. MALORCA improved the ontology from AcListant® project which works well, but it reaches its limits when further extensions are needed. This input from MALORCA was already successfully used by SESAR 2020 16-04 ASR solution.

4.3 Plan for next R&D phase (Next steps)

For the next R&D phase, three aspects need to be considered for assistant based speech recognition (ABSR). The first one deals with the improvements of the machine learning algorithms of generic automatic speech recognition employed in ABSR. The second aspect deals with the reduction of needed expert knowledge and maintenance costs to deploy ABSR to a variety of airports. The third aspect deals with implications on deployment and maintenance costs for air-traffic controller assistance systems generally. As the following description is mainly based on the outcome of the second stakeholder MALORCA workshop (February 2018), it mixes up some of the three aspects mentioned above. According to the first aspect, the progress from the AcListant® to MALORCA project has clearly indicated that there is a significant difference between a laboratory environment (i.e. controller-pilot communication) and a real ATC operations room. Some commands which appear in a simulated environment are only used very seldom in real life situations and vice versa (see results of T1/T2 in D5-2). Therefore, learning from data generated in a simulated environment is not sufficient for developing a speech recognition platform for real controller-pilot communication. Nevertheless the following open research questions can further be addressed in a clean and safe simulation environment:

- Which command recognition rate[7] is sufficient for a specific ASR application? (Short name used in Table 4 at page 37: "Needed Rates")
- Which command recognition error rate[8] is still tolerable? (Short name used in Table 4 at page 37: "Needed Rates")
- Is higher recognition performance absolutely required in busy traffic situations, or is lower performance still satisfactory?[9] (Short name used in Table 4 at page 37: "Needed Rates")
- Is a lower command recognition rate preferred by the controller, if it comes with a reduction of the command recognition error rate (Therefore, providing less false positive commands)? ? (Short name used in Table 4 at page 37: "Needed Rates")
- Should the controller be able to decide for an optimal operating point? (Short name used in Table 4 at page 37: "Needed Rates")
- Can bad recognition performance result in safety issues, or result in a loss of the mental picture, or is even the contrary the case, i.e. the human may trust too much in the automation if recognition performance is high in general? (Short name used in Table 4 at page 37: "Safety Issues")
- Estimation of confidence scores related to reliability of the recognized commands is still a challenging task, not fully addressed in MALORCA. Confidence scores can help to measure the quality of the ABSR (Short name used in Table 4 at page 37: "Confidence scores for ASR output") .
- How can we guarantee that daily learning improves the recognition performance and does not decrease an overall performance? This includes the question, if we can replace the current version running in the ops room by an updated version. Which tests are needed

[7] ASR command output which is neither rejected nor wrong recognized and, therefore, correctly shown to the end-user.

[8] ASR command output which is neither correctly recognized nor rejected and, therefore, wrongly shown to the end-user.

[9] In busy situations the controller has no time for corrections or in busy situations the controller is happy for any help he/she can get even a command recognition rate of 90% already reduces the clicking time by a factor of 10 compared to inputting every command manually.

before to verify it? Or can we expect that each new version is better than the previous version? This includes also the question when an update is required, because the performance of previous version has decreased, i.e. situation parameters have changed since last update (Short name used in Table 4 at page 37: "Daily Learning and Update").

- New algorithms trained under machine learning frameworks should be accurate, efficient and introspective, i.e. they should ask for additional information if they are not able to provide a confident result, i.e. a confidence measures are needed. Accuracy and efficiency was addressed by MALORCA. Introspective algorithms were not considered. (Short name used in Table 4 at page 37: "Introspective Learning Algorithm")
- Training end-to-end models (on big learning data sizes), i.e. direct recognition of the ATC commands (e.g. DLH123 DESCEND FL 70) from the radar data and voice utterance input avoiding the temporary steps transforming the voice utterance to the sequence of phonemes or words and then to final ATC commands (with the radar data). (Short name used in Table 4 at page 37: "End-to-end-Learning")
- Real-time aspect: Speech recognition has to start immediately when a controller starts to talk and/or allowing a partial output (e.g. callsign) should be sent to a controller as soon as possible. (Short name used in Table 4 at page 37: "Real-time A(B)SR")
- Generation of artificial data: Training of neural networks (especially deep neural networks) requires large amount of development data. Can we (and how) generate artificial data? The problem of artificial data is that they rely on a model. Image recognition is a successful example. The pixels of training images were just randomly disturbed to generate more training data. (Short name used in Table 4 at page 37: "Artificial Data")

The working group "ABSR in the ops room at Easter 2020" of MALORCA's second controller workshop (see D6-2 [10]) and the Gap Analysis Report (D5-2, chapter "Mayor Gaps") created clear hints for the following steps which are necessary to use real life data: Integration of pilot readbacks for developing further applications (e.g. safety net functions like readback/hearback errors to detect runway incursions and so on).

- Integration of pilot readbacks as additional input modality in addition to radar data to further reduce command recognition error rates of controller speech. The same, i.e. to increase confidentiality of the ABSR system, applies for other modalities (sensors) like mode-S, or controller inputs into electronic flight strips. This has also a chance to improve the

[10] Working Group "ABSR in the ops room at Easter 2020" of 2nd Stakeholder Workshop (see D6-2, page 25) stated that controller support "is very critical and needs to be started as soon as possible. It is important that controllers from the ops room can use the MALORCA prototype such that say can directly convince themselves of the benefits. Standardized user experience feedback needs to be elicited in order to further improve the system. When the actual enrolment in the ops room happens, the benefit has to be clear from the very beginning."

classification of automatically generated transcripts into good and bad examples. (Short name used in Table 4 at page 37: "Other Input Modalities as second sensor")

- Verifying or adapting MALORCA's approach to other airports than Vienna and Prague. More complex approach areas like London exist. (Short name used in Table 4 at page 37: "MALORCA for Hubs")
- On the other hand more than 200 airports with annual movement numbers between 50,000 and 200,000 are distributed all over the planet. How to create learning algorithm which will be able to build models from available training data of some mid-sized airports and can extrapolate to the other mid-sized airports? Or are only a few hours for each mid-sized airport sufficient to improve the models for other mid-sized airports? (Short name used in Table 4 at page 37: "MALORCA for 100 mid-size airports")
- How can we exchange command transcription including the ontology developed in SESAR 2020 PJ.16-04 ASR solution? And how can we exchange different A(B)SR implementations? (Short name used in Table 4 at page 37: "Standardization (Exchanging of transcriptions (Standardization)")
- How can we extend MALOCA's approach to other controller positions (Tower, Clearance Delivery, Enroute etc.)? (Short name used in Table 4 at page 37: "MALORCA for Tower")
- How can we automatically learn statistical language models? MALORCA has demonstrated that statistical language models outperform grammar based approaches if real world phraseology deviations have to be modelled. Nevertheless, sampling from context-free grammar can lead to significant improvements if combined with statistical language models. (Short name used in Table 4 at page 37: "Learning of Statistical language modeling and phraseologyy")
- MALORCA mostly concentrates on the voice recordings. Manual effort for pre-processing the radar data is still required. This needs to be significantly reduced if learning directly in the ops-room or from thousands of hours is intended. (Short name used in Table 4 at page 37: "Automatic Radar Data processing resp. Other second sensors")
- How can we update the recognition models directly in the operations room? Radar data and voice recordings are generated at the premises of the ANSPs. In MALORCA, these recordings were processed to further adapt and improve the recognition models in the labs of DLR, Idiap and USAAR. Can the learning algorithms be improved at the same time as the new data is released (i.e. real-time processing of data during training)? In this case we have less data privacy problems, as recordings never leave the ops room. (Short name used in Table 4 at page 37: "Learning directly in ops rooms for data privacy")

The working Group "ABSR in the ops room at Easter 2020" of 2nd Stakeholder Workshop also identified the challenge that ABSR needs to access the interfaces in the ops room. On the one hand radar data and voice recordings are needed as inputs; on the other hand the (processed) output of ABSR needs to be shown to the controllers in the radar label, or on a different HMI. Due to the situation of limited competition on supplier side in the ops room for primary systems, the working group has identified "an alternative via the providers of back-up/secondary systems. These are expected to be more open, because they could use this as an opportunity to strengthen their position. Regarding the radar data using ADS-B might be a suitable alternative. In case getting access to the primary screen is a problem, an additional screen (e.g. on a tablet) is possible. Regarding the software maturity level required for the MALORCA software certification, it seems to be feasible and specific industry partners have been identified" (see D6-2, page 25).Another interesting feedback of the Stakeholder Workshop in Vienna was related to other non-technical issues.

- The support of the controllers is needed, which is already achieved for controllers who were participating in AcListant® and MALORCA trials.
- The support of high-level and medium management is needed.
- Potential safety issues have to be evaluated, if ABSR leaves the laboratories and enters the ops room.
- How can we benefit in an early stage from speech recognition in the ops room and do not need to wait another seven years (AcListant® ideas were published in 2011 [48], [49])?
- Recording of thousands of hours of voice samples results also in ethical issues with respect to data privacy. The explicit agreement of controllers is possible, but how to get the commitment of the pilots (in case of using readback data) (see also row of short name "Learning directly in ops rooms for data privacy" in Table 4 at page 37)?
- As MALORCA project has started as an exploratory research (ER) project using new technologies for a completely new application, many questions remain open as shown above. MALORCA offers preliminary results, demonstrating potentials of (semi-) automatic adaption of controller assistance tools by learning from a large amount of daily recorded data in the ops rooms. In principle the project has shown that the use of machine learning algorithms for adapting the ABSR system to a specific environment is a promising approach in two aspects. First the learning algorithms result in an improvement of the command recognition error rate. Second MALORCA has shown that substantial reductions in needed expert knowledge and deployment cost are possible. Concerning the first aspect, AcListant® has clearly demonstrated that command recognition rates up to 95% with command recognition error rates below 2% are technically feasible. By using the developed learning mechanism, MALORCA has even reduced the command recognition error rate below 0.6%. AcListant® has validated that there are already efficiency and workload benefits by substituting manual inputs by speech recognition. We can significantly reduce controllers' workload by employing new systems extracting information from the voice samples. The released controller's resources result in increased airport inbound throughput by two landings per hour (i.e. shown for Düsseldorf airport). We can reduce fuel burn by 50 to 65 litres for an A320. This obviously leads to the reduction of CO_2 emissions. Concerning the second aspect, MALORCA has shown that machine learning can significantly reduce the necessary investments for deployment. Further it also seems possible to reduce maintenance costs for such systems.
- The next step into the future of reduced deployment and maintenance costs for controller assistance tools are the automatic adaptation of e.g. a generic Arrival Manager using a machine learning framework based on the results of MALORCA (see row at short name "Machine Learning for xMAN improvement and maintenance" " in Table 4 at page 37) . For these issues additional and new questions have to be answered, e g.: Do we need new algorithm or can we build mainly on MALORCA's results?

The following Table 4 summarizes the next steps again.

- The column "Short Name" relates to the bullets in the text above (the darker the gray of the cell the higher the interests of the MALORCA partners).
- Column "Maturity Level" describes where to address the subject: ER means that this is (still) fundamental research, i.e. it should be addressed by the Horizon 2020 means of an exploratory research project (comparable to MALORCA in the beginning). ER-IR (Industrial research within exploratory research) means that fundamental research results are already available and the next step is possible (this is comparable to MALORCA at the end of the project). TRL-2 means that a higher maturity level (e.g. thanks to MALORCA or AcListant®) has already been achieved. The subject is ready for an industrial research and could be e.g. addressed in wave two of SESAR 2020 program.
- Column "Needed Data" specifies if data from the labs is sufficient or whether usage of operational data is more useful.
- Research Direction specifies whether the subject is more related to Assistant Based Speech Recognition (A)(B)SR), more related to Machine Learning or whether it relates more to improvement of an Controller Assistant Tools (e.g. improvement/adaptation of an Arrival Manager).
- Column "ATCOs needed for trials?" specifies whether trials with controllers are needed or not. This does not mean, whether the knowledge/feedback of the controllers is needed for specification.
- The last column shows the subjects to which this subject is related to, i.e. these subjects could be combined.

Table 4: Summary of research actions initiated by MALORCA results

Short Name	Maturity Level	Needed Data	Research Direction	ATCOs needed for trials?	Can be combined with
Needed Rates	TRL-2	Lab	A(B)SR	yes	Safety Issues
Safety Issues	TRL-2	Lab	A(B)SR	yes	Needed Rates
Confidence scores for ASR output	ER-IR	Lab	A(B)SR	yes	Introspective Learning Algorithm
Introspective Learning Algorithm	ER	Lab	ML	no	Confidence scores for ASR output
Daily Learning and Update	ER	Lab/Real	ML	no	Learning directly in ops rooms also for data privacy
Other Input Modalities as second	ER-IR	Real	ML, xMAN	no	

Short Name	Maturity Level	Needed Data	Research Direction	ATCOs needed for trials?	Can be combined with
sensor					
End-to-end-Learning	ER	Real	ML	no	
MALORCA for Hubs	ER-IR, TRL-2	Real	ML	no	A(B)SR in the ops room
Applying MALORCA to many of the 100 mid-size airports	ER-IR, TRL-2	Real	ML	no	
Generalization of MALORCA for application to 100 mid-size airports	ER	Real	ML, xMAN	no	
Standardization (Exchanging of transcriptio)	TRL-2	Real	A(B)SR, ML	no	
MALORCA for Towers	ER-IR, TRL-2	Real	A(B)SR, ML	yes	A(B)SR in the ops room
Learning of Statistical language modeling and phraseology	ER	Real	ML	no	
Automatic Radar Data processing resp. other second sensors	ER-IR, TRL-2	Real	ML	no	
Learning directly in ops rooms also for data privacy	ER-IR, TRL-2	Real	ML	no	Daily Learning and Update
Machine Learning for xMAN improvement and maintenance	ER	Real	xMAN, ML	no	
Real-time A(B)SR	TRL-2,ER-IR	Lab	A(B)SR	yes	
Artificial Data	ER	Lab	ML, xMAN	no	
A(B)SR in the ops room	TRL-2	real	A(B)SR, ML	yes	MALORCA for Hubs, MALORCA for Tower

For the ABSR technology achieved so far, ANSPs have benefits with respect to workload reduction and increased user acceptability already now. Due to an increased throughput and reduced fuel consumption airlines and airports on the other hand would have financial benefits from the investments of the ANSPs. Hence, start-up financing for ANSPs to introduce ABSR technology could speed up the process.

Table 4 also shows that MALORCA (demonstrating that successful machine learning application in ATC environment is possible) has opened the door for many other ASR and machine learning related research questions. Three main next steps with respect to achieving higher TRL can are obvious:

- Application of MALORCA for a hub airport,
- Application of MALORCA in a tower environment and
- bringing ABSR (as it is now) very early to the ops room. Only in the ops room we can earn money.

The following research questions respectively supporting tasks (if already results are available)

- Needed Rates,
- Safety Issues,
- Confidence scores for ABSR output,
- Introspective Learning Algorithm,
- Daily Learning and Update,
- Other Input Modalities as second sensor,
- End-to-end-learning,
- Exchanging of transcriptions (Standardization),
- Learning of Statistical language modeling and phraseology,
- Automatic Radar Data processing resp. other second sensors,
- Learning directly in ops rooms also for data privacy
- Real-time A(B)SR
- Artificial Data
- Machine Learning for xMAN improvement and maintenance,
- Applying MALORCA to many of the 100 mid-size airports,
- Generalization of MALORCA for application to 100 mid-size airports

can be integrated into “Application of MALORCA for a hub airport”, “Application of MALORCA in a tower environment” resp. “Bringing ABSR to the ops room”

5 References

5.1 Project Deliverables

In project deliverables D1-1, D1-2, D5-2, D5-3 and D6-3 are **not** confidential. These deliverables are also assessable via MALORCA's homepage www.malorca-project.de. The direct links are also available in the following reference list.

Work Package 1

[1] Cerna, Aneta; Monhart, David: MALORCA project: D1-1: Operational Concept Document, version 3.00; 20. February 2017, see also MALORCA's homepage or type the following URL-link into your browser
http://www.malorca-project.de/wp/wp-content/uploads/D1-1-OCD.pdf

[2] Cerna, Aneta; Helmke, Hartmut; Finke, Michael; Nesvadba, Matej; Motlicek, Petr; Szaszak, György; Oualil, Youssef; Windisch, Christian; Kern, Christian: MALORCA project: D1-2: System Requirement Specification Document, version 2.00; 23. September 2016 ,

[3] Cerna, Aneta; Helmke, Hartmut; Finke, Michael; Nesvadba, Matej; Motlicek, Petr; Szaszak, György; Oualil, Youssef; Windisch, Christian; Kern, Christian: MALORCA project: D1-2: System Requirement Specification Document, version 3.00; 23. February 2018, project internal update of previous version D1-2, see also MALORCA's homepage or type the following URL-link into your browser
http://www.malorca-project.de/wp/wp-content/uploads/D1-2-SRS.pdf

[4] Cerna, Aneta; Helmke, Hartmut: MALORCA project: D1-2: System Requirement Specification, Annex 1, version 1.00; 23. February 2018, project internal document not officially submitted to SJU and not listed in grant agreement

[5] Helmke, Hartmut; Ehr, Heiko; Motlicek, Petr; Oualil, Youssef: MALORCA project: D1-3: Architecture Design Document, version 1.01; 21. November 2016

Work Package 2

[6] Windisch, Christian; Orlik, Peter: MALORCA project: D2-1: Transcribed Data Set or Vienna Approach Area, version 2.00; 14. June 2017

[7] Windisch, Christian; Orlik, Peter: MALORCA project: D2-2: Complete Non-Transcribed Data Set for Vienna Approach Area, version 2.00; 21. June 2017

[8] Cerna, Aneta: MALORCA project: D2-3: Deliverable Report for Transcribed Data Set for Prague Approach Area, version 1.00; 13. January 2017

[9] Cerna, Aneta: MALORCA project: D2-4: Complete Non-Transcribed Data Set for Prague Approach Area, version 1.00; 14. March 2017

Work Package 3

[10]Helmke, Hartmut; Siol, Gerald; Ehr, Heiko: MALORCA project D3-1: Command Hypotheses Predictor, first iteration, version 1.00, 30. November 2016.

[11]Motlicek, Petr; Szaszak, György; Helmke, Hartmut: MALORCA project: D3-2: General acoustic model, first iteration, version 1.00. 28 February, 2017.

[12]Szaszak, Gyorgy; Helmke, Hartmut: MALORCA project D3.3: General, Context-free Grammar Based Language Model, first iteration, version 1.00, 27. February 2017.

[13]Motlicek, Petr; Srinivasamurthy, Ajay; Helmke, Hartmut: MALORCA project D3.4: Neural Network based acoustic model, second iteration, version 1.00, 06. June 2017.

[14]Szaszak, Gyorgy; Oualil, Youssef; Helmke, Hartmut: MALORCA project D3.5: Context Integration into ASR system, version 1.00, 13. June 2017.

[15]Helmke, Hartmut; Siol, Gerald: MALORCA project D3-6: Improved Context provided by Command Hypotheses Prediction, 2nd iteration, version 1.00, 28. March 2017.

Work Package 4

[16]Motlicek, Petr; Srinivasamurthy, Ajay: MALORCA project D4-1: Application of model learning algorithms (initial version for acoustic model), version 1.00, 06. June 2017.

[17]Oualil, Youssef: MALORCA project D4-2: Concept Extractor Definition, version 1.00, 08. November 2017.

[18]Helmke, Hartmut; Kleinert, Matthias; Motlicek, Petr; Srinivasamurthy, Ajay: MALORCA project D4-3: Acoustic, language and hypotheses rule model adaptation for Vienna approach, version 1.00, 22. December 2017.

[19]Kleinert, Matthias; Motlicek, Petr; Singh, Mittul: MALORCA project D4-4: Application of model learning algorithms (final version), version 1.00, 22. December 2017.

[20]Helmke, Hartmut; Kleinert, Matthias; Motlicek, Petr; Srinivasamurthy, Ajay: MALORCA project D4-5: Acoustic, language and hypotheses rule model adaptation for Prague approach, version 1.00, 22. December 2017.

Work Package 5

[21]Friedrich, Maik; Helmke, Hartmut; Oualil, Youssef; Srinivasamurthy, Ajay; Cerna, Aneta; Kern, Christian; Motlicek, Petr: MALORCA project D5-1: Proof-of-Concept Plan, version 1.00, 04. August 2017.

[22]Helmke, Hartmut; Siol, Gerald; Kleinert, Matthias; Ehr, Heiko; Cerna, Aneta; Kern, Christian; Klakow, Dietrich; Motlicek, Petr: MALORCA project D5-2: Gap Analysis Report, version 1.00, 28. February 2018, , see also MALORCA's homepage or type the following URL-link into your browser
http://www.malorca-project.de/wp/wp-content/uploads/D5-2-GapAnalysisReport.pdf.

[23]Helmke, Hartmut; Singh, Mittul; Motlicek, Petr; Kern, Christian; Cerna Aneta; Klakow, Dietrich: MALORCA project D5-3: Final Project Results Report, version 1.00, 28. February 2018, see also MALORCA's homepage or type the following URL-link into your browser http://www.malorca-project.de/wp/wp-content/uploads/D5-3-FinalProjectResultsReport.pdf.

Work Package 6

[24]Helmke, Hartmut: MALORCA project: D6-1: Project Management Plan, version 4.00; 12 December 2016

[25]Helmke, Hartmut; Srinivasamurthy, Ajay; Cerna, Aneta; Windisch, Christian; Kern, Christian: MALORCA project: D6-2: Stakeholder Workshop Report, version 1.00; 26 February 2018, see also MALORCA's homepage or type the following URL-link into your browser http://www.malorca-project.de/wp/wp-content/uploads/D6-2-StakeholderWorkshopReport.pdf

[26]Helmke, Hartmut; Biella, Marcus: MALORCA project: D6-3: Ethical Report 1, version 1.00; 28 April 2017.

[27]Helmke, Hartmut; Biella, Marcus: MALORCA project: D6-4: Ethical Report 2, version 1.00; 17 February 2018.

[28]Helmke, Hartmut; Cerna, A.; Ehr, H.; Motlicek, P.; Oualil, Y.; Windisch, C.: MALORCA project: D6-5: Periodic Technical Report, Part B, Reporting Period 01/04/2016 to 30/09/2016, version 1.03; 17 January 2017

[29]Helmke, Hartmut; Cerna, A.; Ehr, H.; Motlicek, P.; Oualil, Y.; Windisch, C.: MALORCA project: D6-6: Periodic Technical Report, Part B, Reporting Period 01/10/2016 to 31/03/2017, version 1.02; 27 July 2017

[30]Helmke, Hartmut; Cerna, A.; Ehr, H.; Motlicek, P.; Oualil, Y.; Windisch, C.: MALORCA project: D6-7: Periodic Technical Report, Part B, Reporting Period 01/04/2017 to 30/09/2017, version 1.00; 16 November 2017

Work Package 7

[31]Helmke, Hartmut: MALORCA project: D7-1: Implementation of Ethical Requirements, version 1.01; 10 August 2016

[32]Helmke, Hartmut: MALORCA project: D7-2: Implementation of Ethical Requirements, version 1.03; 06 September 2016

[33]Helmke, Hartmut: MALORCA project: D7-3: Implementation of Ethical Requirements, version 1.01; 08 August 2016

[34]Helmke, Hartmut: MALORCA project: D7-4: Implementation of Ethical Requirements, version 1.01; 08 August 2016

[35]Helmke, Hartmut: MALORCA project: D7-5: Implementation of Ethical Requirements, version 1.01; 08 August 2016

[36]Helmke, Hartmut: MALORCA project: D7-6: Implementation of Ethical Requirements, version 1.01; 08 August 2016

[37]Helmke, Hartmut: MALORCA project: D7-7: Implementation of Ethical Requirements, version 1.01; 08 August 2016

See also the two ethical reports [26] and [27].

5.2 Project Publications

[38]Helmke, H.; Motlicek, P.; Himawan, Ivan; Ohneiser, O.; Windisch, C.; Kern, C.; Cerna, A.; Nesvadba, M.; Oualil, Y.; Klakow, D.; Szaszak. G., "MALORCA: Learning of Controllers' Behaviour from Recorded Radar Data and Speech Utterances", Poster presentation 6th SESAR Innovation Days, Delft, Nov. 2016

[39]Helmke, H.; Oualil, Y.; Schulder, M. "Quantifying the Benefits of Speech Recognition for an Air Traffic Management Application", 28th ESSV Congress, Saarbrücken, March. 2017

[40]Srinivasamurthy, A.; Motlicek, P.; Himawan, I.; Szaszák, G.; Oualil, Y.; Helmke. H.: "Semi-supervised Learning with Semantic Knowledge Extraction for Improved Speech Recognition in Air Traffic Control", in Proc. of Interspeech, 2017, Stockholm, Sweden, Aug. 2017

[41]M., Kleinert; H., Helmke; G., Siol; H., Ehr; M., Finke; Y., Oualil; A., Srinivasamurthy: "Machine Learning of Controller Command Prediction Models from Recorded Radar Data and Controller Speech Utterances," 7th SESAR Innovation Days, Belgrade, 2017.

[42]Y., Oualil; D., Klakow; G., Szaszák; A., Srinivasamurthy; H., Helmke; P., Motlicek: "A Context-Aware Speech recognition and Understanding System for Air Traffic Control Domain", in: Proceedings of the IEEE Automatic Speech Recognition and Understanding Workshop, Okinawa, Japan, 2017

[43]Project Website: www.malorca-project.de

5.3 Other

[44]Project Execution Guidelines for SESAR 2020 Exploratory Research, Edition 01.00.00, 08/02/2016

[45]European ATM Master Plan

[46]Helmke, Hartmut: Grant Preparation: MALORCA – Part A, 2016-03-18-698824_Annex 1 - Description of the action (part A).pdf, 18 March 2016

[47]Helmke, Hartmut: Grant Preparation: MALORCA – Part B, 2016-03-18-698824_Annex 1 - Description of Action (part B)_v-0-16.pdf, version 0.16, 18 March 2016

[48]T. Shore, "Knowledge-based word lattice re-scoring in a dynamic context," Master Thesis, Saarland University (USAAR), 2011.

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

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

[51]Ajay, Srinivasamurthy; Petr, Motlicek; Youssef, Oualil: "Semi-supervised Learning with Multimodal Semantic Confidence Measures for Speech Recognition in Air Traffic Control", under preparation to be submitted to Interspeech 2018.

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

[53]H., Gürlük; H., Helmke; M., Wies; H., Ehr; M., Kleinert; T., Mühlhausen; K., Muth; and O., Ohneiser: “Assistant based speech recognition - another pair of eyes for the Arrival Manager,” in IEEE/AIAA 34th Digital Avionics Systems Conference (DASC), Prague, Czech Republic, 2015.

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

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

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

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

[58]H., Helmke; O., Ohneiser; J., Buxbaum and C., Kern: “Increasing ATM Efficiency with Assistant-Based Speech Recognition”, 12th USA/Europe Air Traffic Management Research and Development Seminar (ATM2017). Seattle, Washington, 2017.

[59]ICAO: “ICAO Safety, Airport Movements 2010”, https://www.icao.int/safety/RunwaySafety/Lists/Airport%20Movements/AllItems.aspx

Appendix A

A.1 Glossary of terms)

Term	Definition	Source of the definition
Automatic Speech Recognition	An Automatic Speech Recognition (ASR) system gets an audio signal as input and transforms it into a sequence of words, i.e. “speech-to-text” following the recognition process. The sequence of words is transcribed into a sequence of ATC concepts (“text-to-concepts”) using ontology. The word sequence “lufthansa two alpha altitude four thousand feet on qnh one zero one four reduce one eight zero knots or less turn left heading two six zero” is transcribed into “DLH2A ALTITUDE 4000 ft, DLH2A INFORMATION QNH 1014, DLH2A REDUCE 180 OR_LESS, DLH2A HEADING 260 LEFT”. The resulting concepts can be used for further applications such as visualization on an HMI.	MALORCA project
Command Recognition Rate	The number of controller commands which are correctly recognized by ASR and are not rejected before divided by number of total given commands; in other words: the percentage of given commands correctly shown on the controllers’ HMI.	See definition in [50]
Command (Recognition) Error Rate	The number of controller commands which are wrongly recognized by ASR and which are not rejected divided by number of total given commands; in other words: the percentage of given commands wrongly shown on the controllers’ HMI.	See definition in [50]
Command Hypotheses Predictor	Components needed for Assistant Based Speech Recognition which predict a set of possible commands.	MALORCA [10], [15]
Command Prediction Error Rate	The number of controller commands which are not predicted by the Command Hypotheses Predictor divided by number of total given commands.	MALORCA [10], [15]

Table 5: Glossary

A.2 Acronyms and Terminology

Term	Definition
ABSR	Assistant Based Speech Recognition System
ADS-B	Automatic Dependent Surveillance Broadcast
ANSP	Air Navigation Service Provider
ASR	Automatic Speech Recognition
ATCO	Air Traffic Controller
ATM	Air Traffic Management
HMI	Human Machine Interface
SESAR	Single European Sky ATM Research Programme
SJU	SESAR Joint Undertaking (Agency of the European Commission)
SNR	Signal-to-Noise Ratio

Table 6: Acronyms and technology