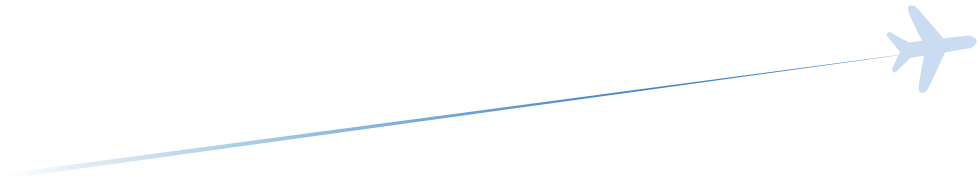

# Four Hours of Silence Reduced Annotated Controller and Pilot Utterances from Iceland


| | |
|---|---|
| **Deliverable ID:** | **D3.3** |
| **Dissemination Level:** | **PU** |
| **Project Acronym:** | **HAAWAII** |
| **Grant:** | **884287** |
| **Call:** | **H2020-SESAR-2020-2** |
| **Topic:** | **SESAR-ER4-18-2019** |
| **Consortium Coordinator:** | **DLR** |
| **Edition date:** | **25 May 2021** |
| **Edition:** | **01-00-00** |
| **Template Edition:** | **02.00.02** |


Founding Members

EUROPEAN UNION EUROCONTROL

SESAR JOINT UNDERTAKING

## Authoring & Approval

### Authors of the document

| Name/Beneficiary | Position/Title | Date |
|---|---|---|
| **Hartmut HELMKE (DLR)** | Project Lead | 24.05.2021 |
| **Shruthi Shetty (DLR)** | Lead of Command Extraction in WP3 | 17.05.2021 |

### Reviewers internal to the project

| Name/Beneficiary | Position/Title | Date |
|---|---|---|
| Hartmut HELMKE (DLR) | Project Lead | 24.05.2021 |

### Approved for submission to the SJU By - Representatives of beneficiaries involved in the project

| Name/Beneficiary | Position/Title | Date |
|---|---|---|
| Hartmut HELMKE | Project Lead | 24.05.2021 |
| Pavel SMRZ (BUT) | WP4 Leader | email feedback 18/05/21 |
| Petr MOTLICEK | WP3 Leader | by commenting in document 18/05/21 |
| Christian WINDISCH (ACG) | D6.4 Leader | 24.05.2021 in document and via email |
| Teodor S. SIMIGANOSCHI (Isavia) | WP1 Leader | By silent approval |
| Julia HARFMANN (NATS) | WP5 Leader | By silent approval |
| Martina ERCEG (CCL) | Contact Point CCL | By silent approval |

### Rejected By - Representatives of beneficiaries involved in the project

| Name/Beneficiary | Position/Title | Date |
|---|---|---|

## Document History

| Edition | Date | Status | Author | Justification |
|---|---|---|---|---|
| 0.01 | 05.05.2021 | Draft | H. Helmke | First Version and integration of block of Amrutha |
| | 10.05.2021 | Draft | H. Helmke | Updating chapters with automatically created transcriptions |
| | 11.05.2021 | Draft | H. Helmke | Creating Next Steps chapter |

| | | | | |
|---|---|---|---|---|
| 0.02 | 13.05.2021 | Draft | H. Helmke | Adding section 6.4 |
| 0.50 | 16.05.2021 | Draft | H. Helmke | Preparation for intermediate review |
| 0.50 | 17.05.2021 | Draft | S. Shetty | Review of section 2 and 3 |
| | 18.05.2021 | Draft | H. Helmke | Integration of feedback from Pavel |
| 0.50 | 18.05.2021 | Draft | P. Motlicek | Review all the sections |
| | 18.05.2020 | Draft | S. Shetty | Corrected text in Section 3 |
| | 20.05.2021 | Draft | S. Shetty | Some corrections with respect to table names and others |
| 0.51 | 24.05.2021 | Draft | H. Helmke | Integration of feedback of Petr from 18.05 into another document, TS deleted from author list, deleted Isavia from copyright list |
| 1.00 | 25.05.2021 | Final | H. Helmke | Accepting al previous changes and preparation for submission to SJU |

## Copyright Statement

© 2021 DLR–
All rights reserved. Licensed to the SJU under conditions.

# HAAWAII

## HIGHLY ADVANCED AIR TRAFFIC CONTROLLER WORKSTATIONS WITH ARTIFICIAL INTELLIGENCE INTEGRATION

This General document is part of a project that has received funding from the SESAR Joint Undertaking under grant agreement No 884287 under European Union's Horizon 2020 research and innovation programme.


## Abstract

Advanced automation support developed in Wave 1 of SESAR IR includes using of automatic speech recognition (ASR) to reduce the amount of manual data inputs by air-traffic controllers. Evaluation of controllers' feedback has been subdued due to the limited recognition performance of the commercial of the shell ASR engines that were used, even in laboratory conditions. Past exploratory research funded project MALORCA, however, has shown (on restricted use-cases) that satisfactory performance can be reached with novel data-driven machine learning approaches. The project builds on very large collection of data, organized with a minimum expert effort to develop a new set of models for complex environments of Icelandic en-route and London TMA. The deliverable is public.

The deliverable D2.3 has described the collecting of surveillance data and voice recordings from Isavia ANS enroute airspace, including the transcription process. As already Alan Turing pointed out, speech recognition does not include speech understanding. Therefore, this deliverable concentrates on the semantic interpretation, i.e. the annotation, of the transcribed voice recordings by using the information from corresponding voice recordings. Currently 7.5 hours of manually transcribed pilot and ATCo utterances are available from Isavia airspace. 90 minutes of them are manually annotated, the remaining 6 hours are already automatically annotated. A recognition rate on command level of 92.6% for ATCo and of 89.9% for pilot utterances is currently – after one third of project runtime -- achieved, if the used transcriptions are manually transcribed. From automatic transcriptions, based on current implementation of speech-to-text transformation a performance of 67.3% and 70.5% for pilots, respectively is achieved, if surveillance data for callsign extraction is used. If this surveillance data is not available, 58.7% for ATCos and 56.2% for pilots are achieved. The challenge of the next months will be to manually check the automatic annotations and to improve the extraction rate especially from automatic transcriptions by improving both extraction performance and speech to text transformation, i.e. reducing the word error rate (WER). Recognition performance of above 90% from the output of the speech-to-text process is the objective to enable readback error detection support for ATCos.

# Table of Contents



# List of Tables

# 1 Executive Summary

As already Alan Turing pointed out, speech recognition does not include speech understanding. Even with a perfect speech to text transformation, i.e. a perfect automatic speech recognition system, we do not know the semantics of the following two air traffic controller (ATCo) utterances “good morning speed bird two zero zero zero alfa reduce one eight zero knots until DME four miles contact tower on frequency one one eight decimal seven zero zero” and “one eighty to DME four tower one eighteen seven speed bird two thousand alfa” we do not even have a clue that both utterances mean the same thing on a semantic level.

Therefore, this deliverable concentrates on the semantic interpretation, i.e. the annotation, of the transcribed voice recordings by using the information from corresponding voice recordings. Currently 7.5 hours of manually transcribed pilot and ATCo utterances are available from Isavia airspace. 90 minutes of them are manually annotated, the remaining 6 hours are already automatically annotated. A recognition rate on command level of 92.6% for ATCo and of 89.9% for pilot utterances is currently – after one third of project runtime -- achieved, if the used transcriptions are manually transcribed. From automatic transcriptions, based on current implementation of speech-to-text transformation a performance of 67.3% and 70.5% for ATCos and pilots, respectively is achieved, if surveillance data for callsign extraction is used. If this surveillance data is not available, 58.7% for ATCos and 56.2% for pilots are achieved. Table 1 provides a good summary of current command extraction performance and also about the available data.

| Type | Total | User-Rec | Relevance | Rec-Rate | Rec-Rate-ASR |
|---|---|---|---|---|---|
| AFFIRM | 63 | 54 | 2.2% | 85.7% | 50.7% |
| ALTITUDE | 318 | 276 | 11.1% | 86.8% | 75.8% |
| CALL_YOU_BACK | 22 | 21 | 0.8% | 95.5% | 66.7% |
| CLIMB | 164 | 158 | 5.7% | 96.3% | 80.2% |
| CONTACT | 159 | 145 | 5.5% | 91.2% | 70.9% |
| CONTACT_FREQUENCY | 242 | 229 | 8.4% | 94.6% | 80.1% |
| CORRECTION | 4 | 4 | 0.1% | 100.0% | 60.0% |
| CPDLC | 101 | 99 | 3.5% | 98.0% | 93.1% |
| DESCEND | 63 | 54 | 2.2% | 85.7% | 61.5% |
| DIRECT_TO | 210 | 165 | 7.3% | 78.6% | 58.0% |
| FAREWELL | 166 | 159 | 5.8% | 95.8% | 81.8% |
| GREETING | 344 | 337 | 12.0% | 98.0% | 92.6% |
| HEADING none | 11 | 4 | 0.4% | 36.4% | 36.4% |
| INFORMATION ACTIVE_RWY | 10 | 7 | 0.3% | 70.0% | 58.3% |
| INFORMATION QNH | 41 | 40 | 1.4% | 97.6% | 56.8% |
| INFORMATION TRAFFIC | 10 | 7 | 0.3% | 70.0% | 36.4% |
| INIT_RESPONSE | 145 | 141 | 5.1% | 97.2% | 84.6% |
| MAINTAIN ALTITUDE | 38 | 28 | 1.3% | 73.7% | 63.9% |
| NEGATIVE | 7 | 7 | 0.2% | 100.0% | 71.4% |
| NO_CONCEPT | 176 | 147 | 6.1% | 83.5% | 59.3% |
| REPORT_MISCELLANEOUS | 27 | 22 | 0.9% | 81.5% | 64.0% |
| SQUAWK | 42 | 39 | 1.5% | 92.9% | 86.1% |
| STATION | 454 | 441 | 15.8% | 97.1% | 82.2% |
| Sum of relevant types | 2817 | 2584 | 98.1% | 91.7% | 75.5% |

**Table 1: Summary of current command extraction performance**

All command types which were observed at least 10 times or are considered as important for the read back error detection application are shown. Seldom occurring, but relevant are e.g. NEGATIVE and CORRECTION. Currently the data base consists of only 90 minutes of annotated data.

We see how often each type occurred (column “Total”), how often the command is correctly extracted and the relevance of this type (“Total” of this type divided by the number of all command types). Column “Rec-Rate” contains “User-Rec” divided by column “Total”, i.e. the command recognition rate for this type. Column “Rec-Rate-ASR” contains the command recognition rate, when the command extraction [4] is applied on the automatically transcribed data and not applied on the manually transcribed and manually verified transcriptions. In yellow we mark types which are significantly below the average and in green which are significantly better.

The improvable performance for the types “ALTITUDE” and “DESCEND” results from the fact that these commands are often used with a conditional clearance, which needs to be modelled better. Improvable performance for “DIRECT_TO” type results also from conditional clearance, but also from the fact, that currently only waypoints and not lat/long coordinates as “six one north one two west” are not modelled.

The challenge of the next months will be to check all automatic annotations for manual created transcriptions and to improve the extraction rate especially from automatic transcriptions by improving both extraction performance and speech to text transformation, i.e. reducing word error rate (WER). Recognition performance of above 90% from the output of the speech-to-text process supported by command prediction is the objective to enable readback error detection support for ATCos.

# 2 Introduction

## 2.1 Purpose of the document

The HAAWAII project builds on a very large collection of data, organized with a minimum expert effort to develop a new set of models for complex environments of Icelandic en-route and London TMA.

This deliverable describes annotation and the automatic annotation process for Isavia's enroute airspace data recorded during July to September 2020.

## 2.2 Scope

The HAAWAII project aims to research and develop a reliable, error resilient and adaptable solution to automatically transcribe voice commands issued by both air-traffic controllers and pilots. To develop new models, large audio data with corresponding transcription is required. The ANSPs of Icelandic en-route and London TMA (Terminal Manoeuvring Area) collect the audio data of the controllers and the pilots.

In order to process the large amount of voice data for building machine learning models, it is important to first transcribe and annotate commands in the most efficient way possible.

The **transcription** task involves the speech-to-text transformation, writing down word-by-word, what the ATCo has said. Examples are: "lufthansa two bravo alfa descend flight level eight zero and reduce speed two two zero knots" and "bonjour air_france two seven three [unk] confirm vien* correction contact vienna radar on one two nine decimal five". The **annotation** task extracts the semantic concepts from the transcriptions (text-to-concepts transformation), e.g., "DLH2BA DESCEND 80 FL, DLH2BA REDUCE 220 kt" and "AFR273 CORRECTION, AFR273 CONTACT VIENNA_RADAR, AFR273 CONTACT_FREQUENCY 129.500".

The transcriptions and annotations are needed on the one hand for creating first models for the speech recognizers, which are later improved by automatic transcriptions and annotations. On the other hand, the manual transcriptions and annotations are needed for evaluating the recognition performance of the developed speech recognitions engines. Therefore, high quality transcriptions and annotations are needed. The performance of the speech recognizer is limited by the quality of the available training data.

This deliverable focusses on the annotations, while the transcriptions are already described in D2.3 [1].

## 2.3 Intended readership

This document is mainly intended for:

- **HAAWAII** consortium members in order to have a common and shared view of the command extraction and annotation process
- **SESAR JOINT UNDERTAKING (SJU)** as Horizon 2020 Programme coordinator.

The information HAAWAII consortium members is characterized as follows:

- Chapter 3 is interesting for Isavia technicians providing the surveillance data. It focuses (see footnote 4 at page 27) on a problem with aircraft with the callsign “arctic eagle” (ICAO designator FEI).

- Chapter 4 concentrate on the current performance of command extraction from manually automatically transcribed utterances. It describes the used metrics (e.g. Command Recognition Rate, Callsign Recognition Rate) in detail. It shows performance on command level and on callsign level. The chapter is interesting for Idiap, BUT, Isavia and NATS.

- Chapter 5 addresses the performance, when command extraction is performed with using automatically transcribed data. It is interesting for Idiap and BUT to understand current HAAWAII challenges of speech understanding.

- Table 19 and Table 20 contain a summary of performance on command level and considering only callsigns by comparing performance from manually transcriptions and automatic transcriptions and also by comparing the performance when surveillance data information is used and when not. These tables might be interesting for all partners and also for the bodies from SJU.

- Section 6.1 gives hints for BUT and Idiap, how to avoid systematic errors in speech-to-text transformation for the ATM domain.

- The tables in section 6.2 show for Isavia which command types from the ontology [5] are currently considered and which command types were observed how often in the transcribed and already manually annotated voice recordings. Table 27 shows the command types, which were observed at least 10 times. These types will be considered for readback error detection. If different or more types are needed, Isavia has the chance to concentrate also on these new types during data preparation during the next months.

- Section 6.4 is also interesting for Isavia and maybe for NATS. It shows for which of the example use cases in D1-1 and D6-2 the status of current implementation and which improvements are planned during the next months. Use cases with priority 3 and above will not be considered. They were currently not observed in the provided data.

- Last but not least the whole document is important for DLR staff, which is responsible for improving the command extraction (annotation) performance. It shows the current status with quantifiable numbers. It shows that many challenges are remaining, but is also shows that even more is already achieved.

## 2.4 Background

The background mainly consists of the recent projects on Assistant Based Speech Recognition

- Helmholtz Validation Fund supported AcListant®, AcListant®-Strips (Active Listening Assistant)
- SESAR2020 Exploratory Research MALORCA (Machine Learning of Speech Recognition Models for Controller Assistance)
- SESAR2020 Industrial Research PJ.16-04-02 ASR (CWP HMI)

- Publication for the DASC 2020 by DLR, which publishes an initial version of the command extraction.

## 2.5 Structure of the document

The structure of this document is based on the Horizon 2020 template for project deliverables. It is organized as follows:

- **Chapter 1: Executive Summary**. Provides a summary of the key information and elements contained in the Technical Validation Report document.
- **Chapter 2: Introduction** (this chapter). Introduces the document.
- **Chapter 3:** Provides a description of the recorded voice data and the corresponding surveillance data, especially addressing the performance of current implementation of command prediction and the limitations due to other used data formats and missing data items compared to e.g. MALORCA project.
- **Chapter 4:** Provides an overview of the current performance based on manually checked voice transcriptions.
- **Chapter 5:** Provides an overview of the current performance based on automatically transcribed data
- **Chapter 6:** Describes the challenges of the next months with respect to improvements of requirements, speech-to-text transformation and command and callsigns extraction.

## 2.6 Glossary of terms

The HAAWAII project has more than 20 different deliverables. Therefore, the HAAWAII project decides to have one separate document containing the glossary of terms, so that maintenance of the terms is eased and errors or misunderstandings only need to be changed in one place.

For simplifying the task of the readers, the contents of the master document are shown in the following table.

| Term | Definition | Source of the definition |
|---|---|---|
| **AcListant®** | Venture Capital funded project Active Listening Assistant being conducted by DLR and Saarland University from 2013 to 2015. | PJ.16-04 |
| **Annotation** | This task extracts the semantic concepts from the Transcription (i.e. text-to-concepts transformation), e.g., “DLH2BA DESCEND 80 FL, DLH2BA REDUCE 220 kt” and “AFR273 CORRECTION, AFR273 CONTACT VIENNA_RADAR, AFR273 CONTACT_FREQUENCY 129.500”. | D3.1 |
| **Assistant Based Speech Recognition (ABSR)** | Special Instance of Automatic Speech Recognition which needs an assistant system to provide context in order to improve recognition rate and/or reduce error rate | See definition in [1] |

| 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 an 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. | PJ.16-04 |
| **Callsign (Recognition) Error Rate** | The number of callsign, which are wrongly recognized by ABSR and which are not rejected divided by the number of total given callsigns; in other words: the percentage of given callsigns wrongly shown on the controllers’ HMI. “oscar kilo one” must be mapped to “OACK1” if this is the only “O..K1” in the air. Otherwise it is counted as an error. | in D1.2 |
| **Callsign Recognition Rate** | The number of callsigns, which are correctly recognized by ABSR and are not rejected before divided by the number of total given callsigns; in other words: the percentage of given callsigns correctly shown on the controllers’ HMI. “oscar kilo one” must be mapped to “OACK1” if this is the only “O..K1” in the air. | in D1.2 |
| **Callsign Rejection Rate** | The number of callsigns, which are said by the ATCo, but mapped to NO_CALLSIGN divided by the number of total given callsigns; in other words: the percentage of given callsigns not shown at all on the controllers’ HMI. | in D1.2 |
| **Chunk** | | D3.1 |
| **Clearance transmission identifier** | The Clearance transmission identifier is part of the readback information and represents the Transmission unique identifier from the Transmission information. This will be used to trace and check a specific transmission from the multiple transmissions. See example in Table 2 Example of transmission information and identifiers | in D1.2 |
| **CoCoLoToCoCo** | Controller Command Logging Tool for Context Comparison that provides a user-friendly interface to carry out transcriptions and various annotations for air traffic control voice commands. | D3.1 |

| Term | Definition | Source of the definition |
|---|---|---|
| **Command Prediction Error Rate** | The number of controller commands which are given but not predicted (by the Command Hypotheses Predictor) divided by number of total given commands; in other words: the percentage of errors of the Command Hypotheses Predictor. | See definition in [1] |
| **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 [1] |
| **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 [1] |
| **Communication group** | Communication group is part of transmission information and it is a generated value or index that is used to identify and group multiple ATCO/Pilot transmissions that represent a single communication/dialogue.<br>The single communication/dialogue is for example when pilot asks for higher flight level and the ATCO provides clearance for that flight level.<br>See example of multiple transmissions grouped into communication groups in Table 2 Example of transmission information and identifiers. | in D1.2 |
| **Concept of Operations [ConOps]:** | Concept of Operations [ConOps]: The ConOps is jointly elaborated by all ATM stakeholders, from the civil and military airspace users and service providers, to airports and the manufacturing industry to gain common understanding of the ATM system. It describes the operational targets, to move ATM towards trajectory-based operations whereby aircraft can fly their preferred trajectories, considering the matching between constraints and optimization. The ConOps allows all ATM stakeholders, from the civil and military airspace users and service providers, to airports and the manufacturing industry to gain common understanding of the ATM system. In this context, the ConOps is the operational answer to reach the ATM Performance improvements targeted by the ATM MP. Furthermore, the ConOps is an important reference for global interoperability and harmonization, as it has been | See definition in [2] |

| Term | Definition | Source of the definition |
|---|---|---|
| | adapted for Europe from the ICAO Global Air Traffic Management Operational Concept. | |
| **Controlling Working Position Identifier** | The controlling working position identifier is part of the Transmission information and represents a name or index to identify the position that generated that specific transmission. See example in Table 2 Example of transmission information and identifiers. | in D1.2 |
| **Exploratory Research** | The exploratory research investigates relevant scientific subjects (during the ATM Excellent Science & Outreach phase) and conducts feasibility studies looking for potential application areas in ATM (during the ATM application-oriented research phase). | See definition in [2] |
| **Horizon 2020** | The EU Framework Programme for Research and Innovation. | SESAR 1, WP14, SESAR 2020 |
| **MALORCA** | Machine Learning of Speech Recognition Models for Controller Assistance, Horizon 2020 funded project from 2016 to 2018 | |
| **PMP deliverable** | Output produced by the projects that is submitted to the SJU via the SESAR 2020 collaborative platform and that is subject to quality assessment by the SJU. However, these deliverables do not appear in the grant agreement as contractual deliverables. The production of PMP deliverables is done in support of subsequent contractual deliverables and is described in the PMP. | See definition in [2] |
| **Project Management Plan** | Formal, approved document, provided by each SESAR 2020 Solution Project, used to manage its execution. It defines how the project is executed, monitored, controlled, and closed. | See definition in [2] |
| **Read-back error detection rate** | The number of correctly detected read-back errors (with or without correction) divided by the total number of read-back errors (with or without correction). | |
| **Read-back error false alarm rate** | The number of detected read-back errors, which are not a read-back error, divided by the total number of read-back errors (with or without correction). | |

| Term | Definition | Source of the definition |
|---|---|---|
| **SESAR 2020** | The SESAR 2020 (Single European Sky ATM Research) Research and Innovation (R&I) Programme will demonstrate the viability of the technological and operational solutions already developed within the SESAR R&I Programme (2008-2016) in larger and more operationally-integrated environments.<br>At the same time, SESAR 2020 will prioritise research and innovation in a number of areas, namely integrated aircraft operations, high capacity airport operations, advanced airspace management and services, optimised network service performance and a shared ATM infrastructure of operations systems and services.<br>SESAR 2020 will retain its founding members, the European Union and Eurocontrol. | SESAR 1, WP14,<br>SESAR 2020, PJ.17-03 |
| **Transcription** | This task involves the speech-to-text transformation, writing down word-by-word, what the ATCo has said. Examples are: “lufthansa two bravo alfa descend flight level eight zero and reduce speed two two zero knots” and “bonjour air_france two seven three [unk] confirm vien* correction contact vienna radar on one two nine decimal five”. | D3.1 |
| **Transmission Direction** | This is either “ATCo” when the ATCo (ground) speaks to the pilot or “Pilot”, if the pilot (air) speaks to the ATCo. | D1.2 |
| **Transmission unique identifier** | Transmission unique identifier is part of transmission information and represents a generated unique value or index that is used to distinguish one single transmission from either ATCO or Pilot. | D1.2 |
| **TRL 2 (V1)** | Technology concept and/or application formulated: Applied research. Theory and scientific principles are focused on very specific application area(s) to perform the analysis to define the concept. Characteristics of the application are described. Analytical tools are developed for simulation or analysis of the application. | See definition in [2] |
| **TRL 3** | **Analytical and experimental critical function and/or characteristic proof-of concept:** Proof of concept validation. Active Research and Development (R&D) is initiated with analytical and laboratory studies including verification of technical feasibility using early prototype implementations that are exercised with representative data. | See definition in [2] |

| Term | Definition | Source of the definition |
|---|---|---|
| **TRL 4 (V2)** | Component/subsystem validation in laboratory environment: Standalone prototyping implementation and test with integration of technology elements and conducting experiments with full-scale problems or data sets. | See definition in [2] |
| **True Positives (tp)** | The total number of correctly predicted commands, i.e., the number of commands which were predicted which were actually given. | |
| **False Positives (fp)** | The total number of falsely predicted commands, i.e., the number of commands which were predicted but actually NOT given. | |
| **False Negatives (fn)** | The total number of commands which were falsely not predicted, i.e., the number of commands which were NOT predicted but were actually given. | |
| **True Negatives (fn)** | The total number of commands which were correctly not predicted, i.e., the number of commands which were NOT predicted and actually NOT given. | |
| **Recall** | Recall represents the percentage of actually given commands which were predicted.<br><br>tp / (tp + fn) | |
| **Precision** | Precision represents the percentage of true predictions out of all the commands which were predicted.<br><br>tp / (tp + fp) | |
| **Accuracy** | Accuracy represents the prediction rate. It also takes into account the number of commands which were correctly NOT predicted.<br><br>(tp + tn) / (tp + fp + fn +tn) | |
| **Segment** | A part of the audio recording without any specific property | D3.1 |
| **Utterance** | Segment of an audio file, which consists of a complete message by only one speaker to the other dialogue participants . In case of ATC it contains complete message of ATCO to one pilot or complete answer of pilot to ATCO. Utterance can contain one or more sentences e.g. “Good morning. Speed bird one three seven descend flight level | D3.1 |

| Term | Definition | Source of the definition |
|---|---|---|
| | eighty". Utterance segments can be automatically or manually created. | |
| **SpokenData** | A generic web based tool which allows to transcribe the speech recordings, while transcribers are supported by several functions to minimise their effort. | D3.1 |

**Reference used in Glossary of terms**

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

[2] SESAR 2020 Execution guidance of ER4 projects : https://ec.europa.eu/research/participants/data/ref/h2020/other/guides_for_applicants/jtis/h2020-guide-project-handbook-er4-sesar-ju_en.pdf

**Table 2 Example of transmission information and identifiers.**

| Transmission unique identifier | ATCO/Pilot Transmission | Clearance transmission Identifier | Controlling Working Position Identifier | Communication Group |
|---|---|---|---|---|
| 1 | ATCO: XYZ descend flight level three one zero | 1 | CWP1 | 1 |
| 2 | Pilot: XYZ descending level three one zero | 1 | CWP1 | 1 |
| 3 | ATCO: ASD here Reykjavik control 1, 2,3,4,5 audio check. | NULL | CWP1 | 2 |
| 4 | Pilot: I hear you 5 by 5. | NULL | CWP1 | 2 |
| 5 | ATCO: ABC descend flight level three one zero | 2 | CWP1 | 3 |
| 6 | Pilot: ABC level one three zero | 2 | CWP1 | 3 |
| 7 | Pilot: ABC correction descending flight level three one zero | 2 | CWP1 | 3 |
| 8 | ATCO: XYZ descend flight level one zero zero | 3 | CWP1 | 4 |
| 9 | Pilot: XYZ descending level one zero zero | 3 | CWP1 | 4 |
| 10 | Pilot: And how is the weather in Keflavik? | NULL | CWP1 | 4 |
| 11 | ATCO: Its always still wind and sunny. | NULL | CWP1 | 4 |

## 2.7 Acronyms and terminology

The HAAWAII project has more than 20 different deliverables. Therefore, the HAAWAII project decides to have one separate document containing these acronyms, so that maintenance of the acronyms is eased and errors or misunderstandings only need to be changed in one place.
For simplifying the task of the readers, the contents of the master document are shown in the following table.

| Term | Definition |
|---|---|
| **ABSR** | Assistant Based Speech Recognition |
| **ACC** | Area Control Centre |
| **ACG** | Austro Control Österreichische Gesellschaft für Zivilluftfahrt (Austrian ANSP) |
| **ADS-B** | Automatic dependent surveillance–broadcast |
| **AEC** | Approach executive controller |
| **AFIS** | Aerodrome Flight Information Service |
| **AG** | Attention Guidance |
| **AI** | Artificial Intelligence |

| Term | Definition |
|---|---|
| **ANRIC** | Aeronautical Radio Incorporated |
| **ANSP** | Air Navigation Service Provider |
| **ANS-CR** | Air Navigation Services of the Czech Republic |
| **APC** | Approach planning controller |
| **APP** | Approach |
| **ARR** | Arrival |
| **ARTAS** | ATM suRveillance Tracker And Server |
| **ASR** | Automatic Speech Recognition |
| **ASTERIX** | All Purpose Structured Eurocontrol Surveillance Information Exchange |
| **ASW** | Air situation window |
| **ATC** | Air Traffic Control |
| **ATCo** | Air Traffic Controller; also ATCO used, but ATCo preferred in HAAWAII project |
| **ATM** | Air Traffic Management |
| **Avg** | Average |
| **BUT** | Brno University of Technology |
| **CBA** | Cost Benefit Analysis |
| **CER** | Command or Context (Prediction) Error Rate, also used as CtxER |
| **Cmd** | Command (files containing annotations) |
| **CmDER** | Command Error Rate |
| **CmDRR** | Command Recognition Rate |
| **CoCoLoToCoCo** | Controller Command Logging Tool for Context Comparison |
| **Cor** | Correct (files containing transcriptions) |
| **COTS** | Commercial of the shell |
| **CPP** | Context Portion Predicted |
| **CONOPS** | Concept of operations |
| **CPDLC** | Controller Pilot Data Link Communications |
| **CTA** | Control area |
| **CTR** | Controlled traffic region |
| **CtxER** | See CER |
| **CV** | Clearance verification |

| Term | Definition |
|---|---|
| **CWP** | Controller Working Position |
| **DASC** | Digital Avionics Systems Conference |
| **DEC** | Departure executive controller |
| **DEP** | Departure |
| **DFS** | Deutsche Flugsicherung GmbH (German ANSP) |
| **DLR** | German Aerospace Center, Deutsches Zentrum für Luft- und Raumfahrt e.V. |
| **DNN** | Deep neural network |
| **DPO** | Data Protection Officer |
| **DVI** | Direct Voice Input |
| **DVO** | Direct Voice Output |
| **EATMA** | European Air Traffic Management Architecture, An architectural Model of European ATM for each SESAR Concept Story board step containing information relating to Operational activities. |
| **EDR** | Event Detection Rate |
| **EML** | European Media Laboratory |
| **ENAIRE** | Spanish ANSP |
| **ER** | En-Route |
| **Err** | Error (files containing errors) |
| **EU** | European Union |
| **EXE** | Exercise |
| **FAA** | Federal Aviation Administration |
| **FANS** | Future Air Navigation System |
| **FDPS** | Flight Data Processing System |
| **FL** | Flight level |
| **FIR** | Flight Information Region |
| **ft** | Feet |
| **GDPR** | General Data Protection Regulation |
| **GUI** | Graphical User Interface |
| **HF** | Human factors |
| **HMI** | Human Machine Interface |
| **HUP** | Human Performance |
| **IB** | Information Bottleneck |

| Term | Definition |
|---|---|
| **ICAO** | International Civil Aviation Organization |
| **ICE** | Intelligent Communications Environment |
| **ID** | Identifier |
| **Idiap** | Idiap Research Institute |
| **IEC** | Information executive controller |
| **ILS** | Instrument landing system |
| **IFR** | Instrument Flight Rules |
| **ISA** | Instantaneous self assessment |
| **JSON** | JavaScript Object Notation |
| **khz** | Kilo hertz |
| **KPA** | Key Performance Area |
| **kt** | Knots |
| **KWA** | Keyword Spotting Algorithm, special implementation of callsign recognition |
| **LAC** | London Area Control |
| **LTCC** | London Terminal Control Centre |
| **LTMA** | London Terminal Manouvering Area |
| **MALORCA** | Horizon 2020 funded project MACHINE LEARNING OF SPEECH RECOGNITION MODELS FOR CONTROLLER ASSISTANCE |
| **MWM** | Mental Workload Model |
| **N/A** | Not applicable |
| **NASA TLX** | NASA Task load index |
| **NATS** | United Kingdom ANSP |
| **NAT OTS** | NORTH ATLANTIC ORGANIZED TRACK SYSTEM |
| **Nm** | Nautical miles |
| **No.** | Number |
| **NOK** | Not Ok |
| **NPR** | Noise Preferential Route |
| **OA** | Open Access |
| **Obj** | Objective |
| **OSED** | Operational services and environment description |
| **OTS** | ORGANIZED TRACK SYSTEM |

| Term | Definition |
|---|---|
| **PC** | Prestwick Centre |
| **PEC** | Director executive controller |
| **PERF** | Performance |
| **PJ** | Project |
| **POK** | Partly Ok |
| **PST** | Performance Stability |
| **PSS** | Paperless Strip System |
| **PTT** | Push to talk |
| **R/T** | Radio Telephony |
| **RabbitMQ** | is an open-source message-broker software (sometimes called message-oriented middleware) |
| **REF** | Reference |
| **REQ** | Requirement |
| **ReTi** | Reaction Time |
| **RMA** | Radar Manoeuvring Areas |
| **RNAV** | Area navigation |
| **RTP** | Real Time Protocol |
| **RWY** | Runway |
| **(S)VFR** | (Special) Visual Flight Rules |
| **S2T** | Speech-To-Text |
| **SA** | Situation Awareness |
| **SAD** | Speech Activity Detection |
| **SAF / SAFE** | Safety |
| **SAR** | Safety assessment report |
| **SASHA** | Situation Awareness for SHAPE (Solutions for Human Automation Partnerships in European ATM) |
| **SC APP** | Approach Senior Controller |
| **Scn** | Scenario |
| **SDK** | Software Development Kit |
| **SDDS** | Surveillance Data Distribution |
| **SESAR** | Single European Sky ATM Research |
| **SID** | Standard instrument departure |

| Term | Definition |
|---|---|
| **SJU** | SESAR Joint Undertaking |
| **SME** | Subject Matter Experts |
| **SOL** | Solution |
| **STAR** | Standard terminal arrival route |
| **STCA** | Short Term Conflict Alerting |
| **T2C** | Text-to-Concept |
| **T2S** | Text-to-Speech |
| **TC** | Terminal Control |
| **TMA** | Terminal Manoeuvring Area |
| **TRL** | Technology Readiness Level |
| **TS** | Technical Specification |
| **TSWR** | Tower |
| **TTC** | Text-to-Concept |
| **TTS** | Text-to-Speech |
| **TVALP** | Technical Validation Plan |
| **TVALR** | Technical Validation Report |
| **V2T** | Voice to Text |
| **V&V** | Validation & Verification |
| **VAD** | Voice activity detection |
| **VCS** | Voice communication system |
| **VFR** | Visual flight rules |
| **VieAPP** | Vienna Approach |
| **VRR** | Voice Recognition and Response |
| **VTT** | Voice to Text |
| **WDR** | Word Detection Rate, approx.. 100% - WER |
| **WER** | Word Error Rate |
| **WL** | Workload |
| **w.r.t.** | with respect to |
| **XML** | eXtenable Markup Language |

# 3 Description of Used Data

The following Table 3 lists the directories, for which transcribed voice data for Isavia enroute and oceanic airspace is available.[1]

| Day and Work Station | # Wav | # Cor | # Cmd | Wav[s] | Cor[s] | Cmd[s] |
|---|---|---|---|---|---|---|
| 2020-03-12__Sector1_001 | 93 | 93 | 93 | 253 | 253 | 253 |
| 2020-07-21__Sector1_001 | 46 | 46 | 46 | 177 | 177 | 177 |
| 2020-07-24__Sector1_001 | 827 | 827 | 32 | 3471 | 3471 | 183 |
| 2020-07-24__Sector1_002 | 523 | 523 | 29 | 2557 | 2557 | 161 |
| 2020-07-31__Sector1_001 | 943 | 943 | 47 | 3642 | 3642 | 210 |
| 2020-07-31__Sector1_002 | 364 | 364 | 25 | 1354 | 1354 | 111 |
| 2020-08-01__Sector1_001 | 672 | 672 | 12 | 2549 | 2549 | 48 |
| 2020-08-02__Sector1_001 | 392 | 392 | 392 | 1177 | 1177 | 1177 |
| 2020-08-04__Sector1_002 | 333 | 333 | 3 | 1265 | 1265 | 11 |
| 2020-08-05__Sector1_001 | 597 | 597 | 32 | 2133 | 2133 | 166 |
| 2020-08-07__Sector1_001 | 554 | 554 | 20 | 2279 | 2279 | 120 |
| 2020-09-02__Sector1_001 | 927 | 927 | 30 | 4379 | 4379 | 180 |
| 2020-09-02__Sector1_002 | 460 | 460 | 23 | 2073 | 2073 | 128 |
| | 6731 | 6731 | 784 | 7:35:09 | 7:35:09 | 0:48:45 |

**Table 3: Transcribed Voice Recordings from Isavia enroute airspace from 2020**

In total, 6731 wave files with voice recordings are available currently. All of them are transcribed. The duration of these 6731 wave files is 7 hours and 35 minutes. Currently only 784 files of them are manually annotated. The duration of these annotations is approximately 49 minutes (already silence reduced). Three directories (grey colour in Table 3) are already fully manual annotated. The other files are all automatically annotated.

The following Table 4 shows the number of aircraft which are in the air, or in other words the aircraft for which surveillance data is available. The last three columns show the number of callsigns which are predicted, i.e. callsigns for which it is assumed that the ATCo may give a command during the next minute or callsigns corresponding to aircraft where the pilot may initiate a call to the ATCo.

[1] The statistic was generated 2021-05-05.

| | | Number of Aircraft | | | Callsigns in Context | | |
|---|---|---|---|---|---|---|---|
| **Day and Work Station** | **# Files** | **# min** | **# max** | **#aver** | **# min** | **# max** | **#aver** |
| 2020-03-12__Sector1_001 | 93 | 23 | 28 | 25 | 16 | 21 | 18 |
| 2020-07-21__Sector1_001 | 46 | 7 | 30 | 21 | 3 | 23 | 14 |
| 2020-07-24__Sector1_001 | 827 | 4 | 39 | 21 | 2 | 29 | 16 |
| 2020-07-24__Sector1_002 | 523 | 8 | 42 | 29 | 5 | 28 | 21 |
| 2020-07-31__Sector1_001 | 943 | 6 | 32 | 25 | 5 | 28 | 20 |
| 2020-07-31__Sector1_002 | 364 | 1 | 29 | 19 | 1 | 23 | 15 |
| 2020-08-01__Sector1_001 | 672 | 14 | 37 | 23 | 8 | 30 | 18 |
| 2020-08-02__Sector1_001 | 392 | 5 | 47 | 27 | 2 | 37 | 22 |
| 2020-08-04__Sector1_002 | 333 | 5 | 50 | 24 | 2 | 42 | 19 |
| 2020-08-05__Sector1_001 | 597 | 5 | 47 | 29 | 2 | 34 | 20 |
| 2020-08-07__Sector1_001 | 554 | 4 | 47 | 23 | 3 | 38 | 17 |
| 2020-09-02__Sector1_001 | 927 | 4 | 47 | 27 | 1 | 38 | 19 |
| 2020-09-02__Sector1_002 | 460 | 7 | 46 | 26 | 3 | 35 | 18 |
| Sum / Average | 6731 | 7.2 | 40.1 | 24.5 | 4.1 | 31.2 | 18.2 |

**Table 4: Number of aircraft with surveillance data and number of callsigns predicted**

Looking over all days on average we have at least (# min) 7.2 aircraft with surveillance data. On 31st of July on workstation 002 we have the absolute minimum of just one aircraft with surveillance data. The maximum number of aircraft is 40.1 (absolute maximum is 50 on 4th of August) and the average number is 24.5. On average we predict 18.2 aircraft /callsigns to which the ATCo may talk within the next minute. The minimum number is 4.1 and the maximum is 31.2, whereas the absolute maximum is 42.

This is the current status of callsign prediction, which is expected to improve. We hope that a reduction of 50% in the average number and in the maximal number of predicted callsigns is possible. One of the challenges is that no flight plan information is available compared to MALORCA project, i.e. it is not clear from the actual surveillance data itself that an aircraft is an inbound or an outbound or an overflight. The majority of flights is of course just an overflight not landing or starting from an Icelandic airport. The decision whether an aircraft is an arrival can only be taken from future surveillance data, i.e., whether it is landing on an Icelandic runway or not.

The following Table 5 will, however, show a general problem with HAAWAII callsign prediction for Isavia airspace. The table shows the number of given commands, which are derived for most of the directories automatically, i.e. no manual checking was performed. We expect that these numbers corresponding to callsign errors will be slightly decrease, when automatic extraction is improved resp. when manual checking and correction would be done.

| Day and Work Station | Sum Cmds | NO_CALLSIGN | NO_CONCEPT | # Errors | Area Errors | No Radar | % Err |
|---|---|---|---|---|---|---|---|
| 2020-03-12__Sector1_001 | 92 | 39 | 10 | 0 | 0 | 0 | 0.0% |
| 2020-07-21__Sector1_001 | 77 | 10 | 4 | 9 | 0 | 9 | 11.6% |
| 2020-07-24__Sector1_001 | 1172 | 198 | 70 | 93 | 0 | 93 | 7.90% |
| 2020-07-24__Sector1_002 | 843 | 79 | 35 | 104 | 21 | 83 | 12.3% |
| 2020-07-31__Sector1_001 | 1436 | 161 | 92 | 124 | 0 | 124 | 8.60% |
| 2020-07-31__Sector1_002 | 577 | 70 | 21 | 35 | 0 | 35 | 6.0% |
| 2020-08-01__Sector1_001 | 1127 | 102 | 55 | 29 | 0 | 29 | 2.50% |
| 2020-08-02__Sector1_001 | 583 | 90 | 30 | 18 | 0 | 18 | 3.0% |
| 2020-08-04__Sector1_002 | 515 | 56 | 50 | 83 | 15 | 68 | 16.1% |
| 2020-08-05__Sector1_001 | 920 | 109 | 58 | 122 | 8 | 114 | 13.2% |
| 2020-08-07__Sector1_001 | 865 | 112 | 44 | 126 | 26 | 100 | 14.5% |
| 2020-09-02__Sector1_001 | 1447 | 162 | 89 | 172 | 0 | 172 | 11.8% |
| 2020-09-02__Sector1_002 | 740 | 87 | 37 | 70 | 0 | 70 | 9.4% |
| Sum / Average | 10394 | 1275 | 595 | 985 | 70 | 915 | 8.99% |
| Only for the manual annotated directories | 752 | 139 | 44 | 27 | 0 | 27 | 4.87% |

**Table 5: Accuracy of command prediction**

"NO_CALLSIGN" means that for this file/utterance no callsign could be extracted, which in most cases means that no callsign was said by the ATCo or pilot. "NO_CONCEPT" means that no command of the ontology [5] was extracted for this utterance, which is e.g. the case for the utterances "euro trans three" or for "if requesting to deviate north of valdi should be no problem".

"#Errors" counts the number of commands with a callsign different from NO_CALLSIGN for which this callsign is currently not predicted. It is the sum of the following two columns.

"Area Errors" count the number of errors, for which surveillance data is available, but the aircraft is currently outside the expected lat/long rectangle.

"No Radar" counts the number of commands, for which the extracted callsign is not found in the surveillance data. For utterances currently automatically annotated the reason could be that the extracted callsign is still wrong.[2],[3],[4]

"% Err" is the error percentage, i.e. column "# Errors" divided by column "Sum Cmds".

[2] The manual annotated directories however show that also missing surveillance data could be reason. In the directory 2020-07-21__Sector1_001 the callsign BCS3998 is responsible for all 9 errors. The BCS3998 is visible in the surveillance data from 05-44-44 to 06-53-18, but all the utterances occur roughly 6min earlier.

[3] The 18 errors in folder 2020-08-02__Sector1_001 result from 6 different callsigns

- GRL901 has 4 commands at 07-57-23-51, but the first surveillance data is from 08:34:48, and then continuously starting in FL 380.
- FLI47 not found in surveillance data, only FLI470 and FLI471 have surveillance data, but much later, FLI47 not in radar data, checking transcription with wave file, HHe cannot even understand that FLI47 said in that utterance.
- MYA77 has radar data from day before until 18:42:25 and new surveillance data start at 09:09:39, utterance from 2020-08-02__08-58-45-19_P, first surveillance data from 09:09:42, starting in FL 41 and then climbing, next conversation at 2020-08-02__09-12-04-25_P. The surveillance data from the day before seems to be a bug on DLR side, which needs to be solved.
- FEI720 not found in surveillance data at all, transcription seems to be correct, no similar aircraft ending at 'zero' found in predicted callsigns at that time
- FEI721 gets commands from 12-58-27-53 to 13-10-45-57_P (3 utterances, but no surveillance data available at all for that aircraft during the whole day.
- ICG31 has hours before and after radar data, but not between 13:38:21 and 14:54:37, Therefore, at 13-39-02-28 the callsign is not in context, the aircraft is before called with ICG31B (bravo means emergency, however not clear, why 76 minutes of data are missing, last altitude 65, first altitude 64.75, positions also not strange.

[4] A special problem are aircraft with the designator FEI (arctic eagle). As not flight plan information is available the callsign is in I062/380#2 Ident, i.e. Characters 1-8 (coded on 6 bits each) defining a target identification when flight plan is available or the registration marking when no flight plan is available. For "FEI33" only "TFORC" is in the data. For artic eagle we found for that day only FEI760, FEI741 and FEI740.

# 4 Current Performance of Automatic Annotation on gold transcriptions

The following Table 6 shows the performance of command extraction [4] for the ATCo utterances which are already manually annotated. As seen before, three directories are already completely annotated. For the other directories, only the interesting cases are already annotated. In most cases these cases are the challenging cases for which the extraction code most be updated. Therefore, the performance is slightly lower compared to the fully manually annotated folders.

| Day and Work Station | # gold | RcR | ErR | RjR | # Words | Unknown Cl | Unknown 2 |
|---|---|---|---|---|---|---|---|
| 2020-03-12__Sector1_001 | 81 | 95.1% | 1.2% | 6.2% | 444 | 128 | 4 |
| 2020-07-21__Sector1_001 | 42 | 95.2% | 2.4% | 2.4% | 193 | 19 | 1 |
| 2020-07-24__Sector1_001 | 45 | 82.2% | 15.6% | 4.4% | 256 | 51 | 6 |
| 2020-07-24__Sector1_002 | 25 | 92.0% | 16.0% | 0.0% | 172 | 29 | 2 |
| 2020-07-31__Sector1_001 | 45 | 97.8% | 0.0% | 2.2% | 233 | 25 | 8 |
| 2020-07-31__Sector1_002 | 23 | 91.3% | 8.7% | 0.0% | 122 | 15 | 2 |
| 2020-08-01__Sector1_001 | 21 | 100% | 0.0% | 0.0% | 104 | 0 | 0 |
| 2020-08-02__Sector1_001 | 355 | 96.1% | 4.2% | 0.6% | 1696 | 219 | 10 |
| 2020-08-04__Sector1_002 | No manual annotation for ATCos done at all | | | | | | |
| 2020-08-05__Sector1_001 | 46 | 89.1% | 10.9% | 6.5% | 375 | 89 | 4 |
| 2020-08-07__Sector1_001 | 30 | 83.3% | 16.7% | 6.7% | 194 | 53 | 2 |
| 2020-09-02__Sector1_001 | 23 | 91.3% | 8.7% | 0.0% | 134 | 17 | 2 |
| 2020-09-02__Sector1_002 | 16 | 93.8% | 6.3% | 6.3% | 103 | 20 | 0 |
| Sum / Average | 752 | 93.9% | 5.7% | 2.3% | 4026 | 665 | 41 |

**Table 6: Accuracy of automatic command extraction for ATCo utterances**

Command recognition rates are computed by comparing instructions from **manual** human annotation (gold annotation) to the results of the **automatic** semantic extraction (command extraction). For a given speech utterance, each instruction is treated as one big word. Then, the Levenshtein distance between the gold annotation and the results of command extraction is calculated, resulting in the number of substitutions (subs), insertions (ins) and deletions (del). Table 7 gives an overview about the different metrics and illustrates an example how they are calculated. In the table #gold defines the total number of commands in the gold annotation. #match defines the number of matches, which is #gold – subs – del. The table also shows, why the sum of RcR, ErR and RjR can be bigger than 100%. This is the case when more commands are recognized than really said.

| Metric | | Calculation |
|---|---|---|
| Command Recognition Rate (**RcR**) | | **RcR** = #matches / #gold |
| Command Recognition Error Rate (**ErR**) | | **ErR** = (subs + ins) / #gold |
| Command Rejection Rate (**RjR**) | | **RjR** = del / #gold |
| **Example** | | |
| **Gold Annotation** | | **Command Extraction** |
| AFR123 INIT_RESPONSE<br>AFR123 TURN LEFT<br>AUA1AB SPEED 140 kt<br>DLH123_NO_CONCEPT | | AFR123 DIRECT_TO OKG none<br>AFR123 INIT_RESPONSE<br>AFR123 TURN_RIGHT<br>AUA1AB NO_CONCEPT<br>DLH123 NO_CONCEPT |
| ***Result:*** | | |
| *RcR = 2/4 = 50% (green)* | *ErR = 2 / 4 = 50% (purple)* | *RjR =1/4 = 25% (yellow)* |

**Table 7: Metric Definition for Command Recognition Performance**

If the result of the command extraction contains either NO_CONCEPT or NO_CALLSIGN, these substitutions and insertions are always calculated as deletions, i.e., these extractions contribute to the rejection rate and not to the error rate (as shown in the example in Table 7).

"#Words" counts the number of words in the corresponding manually transcribed utterance. "Unknown Cl" counts the number of words, which are not used for extracting the commands, i.e. they are neither classified as callsign, nor type, nor value, nor unit, nor qualifier nor as condition. "Unknown 2" counts the number of consecutive word pairs which are classified as unknown. Three and more consecutive unknown classification are not counted here. Table 8 provides an example.

| one | one nine | one | one nine decimal zero | vista | jet seven zero eight | have | a | good | day |
|---|---|---|---|---|---|---|---|---|---|
| vare | vare | vare | unkn unkn | unkn unkn | csgn csgn | csgn csgn | csgn | unkn unkn | type type |
| | | | ########## | | | | | ########## | |

**Table 8: Example for Number of Consecutive Unknowns**

We have 17 word (#Words). 6 words are classified in the second row with "unkn" (Unknown Cl) and we have 1 sequence of two consecutive unknown classification in the word sequence "have a" (column Unknown 2).

For calculation of the callsign recognition rates CaR, CaE and CaRj, see definitions in Table 9 we just compare the callsigns from the gold annotation and from the automatic extraction. For each utterance we consider the callsign only once, except when different callsigns are annotated or extracted. For the example in Table 7 this results in the three annotated and extracted callsigns AFR123, AUA1AB and DLH123.

| Metric | Calculation | |
|---|---|---|
| Callsign Recognition Rate (**CaR**) | Same as RcR but only for callsigns without instructions, which is number of all utterances minus the wrong callsign recognitions divided by all utterances (UttCnt -WrongCsgn) / UttCnt | |
| Callsign Recognition Error Rate (**CaE**) | Same as ErR, but only for callsigns without instructions (InventedCsgn + NoCsgnMissed + BreakBreak) / UttCnt, | |
| Callsign Rejection Rate (**CaRj**) | Same as RjR, but only for callsigns without instructions NoExtraction / UttCnt. | |
| *If the command extraction results in different callsigns, the calculation is done for each callsign. See example below, which also illustrate that the sum of RcR, ErR and RjR can exceed 100%.* | | |
| **Example** | | |
| **Gold Annotation** | **Command Extraction** | |
| AFR123 INIT_RESPONSE<br>AFR123 TURN LEFT<br>AUA1AB SPEED 140 kt<br>DLH123_NO_CONCEPT | AFR123 DIRECT_TO OKG none<br>AFR123 INIT_RESPONSE<br>AFR123 TURN RIGHT<br>AUA1AB NO_CONCEPT<br>AFRTYY NO_CONCEPT | |
| ***Result:*** | | |
| *CaR = 2/3 = 67% (green)* | *CaE= 1 / 3 = 33% (purple)* | *CaRj =0/3 = 0%* |

**Table 9: Metric Definition for Callsign Recognition Performance**

Table 10 provides the callsign recognition performance for all annotated ATCo utterances.

| Day and Work Station | UttCnt | Wrong Csgn | Invented Csgn | No Extraction | No Csgn missed | Break Break |
|---|---|---|---|---|---|---|
| 2020-03-12__Sector1_001 | 56 | 3 | 1 | 1 | 1 | 0 |
| 2020-07-21__Sector1_001 | 22 | 1 | 0 | 1 | 0 | 0 |
| 2020-07-24__Sector1_001 | 18 | 2 | 0 | 1 | 1 | 0 |
| 2020-07-24__Sector1_002 | 13 | 0 | 0 | 0 | 0 | 0 |
| 2020-07-31__Sector1_001 | 18 | 0 | 0 | 0 | 0 | 0 |
| 2020-07-31__Sector1_002 | 10 | 0 | 0 | 0 | 0 | 0 |
| 2020-08-01__Sector1_001 | 7 | 0 | 0 | 0 | 0 | 0 |
| 2020-08-02__Sector1_001 | 188 | 3 | 1 | 0 | 2 | 0 |
| 2020-08-04__Sector1_002 | No manual annotation for ATCos done at all | | | | | |
| 2020-08-05__Sector1_001 | 23 | 1 | 0 | 0 | 1 | 0 |
| 2020-08-07__Sector1_001 | 11 | 0 | 0 | 0 | 0 | 0 |
| 2020-09-02__Sector1_001 | 10 | 0 | 0 | 0 | 0 | 0 |
| 2020-09-02__Sector1_002 | 8 | 0 | 0 | 0 | 0 | 0 |
| Sum | 384 | 10 | 2 | 3 | 5 | 0 |
| Rates | 97.4% | CaR | CaE | 1.8% | CaRj | 0.8% |

Table 10: Accuracy of automatic callsign extraction/recognition rates for ATCo utterances

Column “UttCnt” contains the number of considered utterances, i.e. wave files. Column “Wrong Csgn” shows the number of cases, in which a callsign was extracted from the utterance, but the callsign was wrong. It is the sum of the following four columns. An example is the utterance “euro trans three” just consisting of these 3 words. Extracted was “BCS3”, but the correct extraction would have been “BCS3998”, which would be possible, because the “BCS3998” is the only euro trans at that time. “Invented Csgn” counts the number of cases, in which a callsign was recognized, which was not said. “No Extraction” counts the number of cases, in which “NO_CALLSIGN” was extracted, but this is wrong, because a callsign was provided. “No Csgn missed” counts the number of cases, in which a callsign was extracted, but no callsign was said. “Break Break” counts the number of cases, in which more than one different callsign was said by the ATCo, but only one was extracted, e.g. in “lufthansa alfa bravo descend flight level six zero break break **speed bird four alfa nine** call you back stand by”.

The following Table 11 corresponds to Table 6, but it contains the numbers only for pilot utterances.

| Day and Work Station | # gold | RcR | ErR | RjR | # Words | Unknown Cl | Unknown 2 |
|---|---|---|---|---|---|---|---|
| 2020-03-12__Sector1_001 | 55 | 98.2% | 0.0% | 1.8% | 324 | 45 | 2 |
| 2020-07-21__Sector1_001 | 46 | 82.6% | 6.5% | 10.9% | 235 | 43 | 3 |
| 2020-07-24__Sector1_001 | 40 | 90.0% | 7.5% | 2.5% | 230 | 47 | 7 |
| 2020-07-24__Sector1_002 | 44 | 79.5% | 2.3% | 20.5% | 219 | 34 | 2 |
| 2020-07-31__Sector1_001 | 67 | 82.1% | 7.5% | 10.4% | 381 | 71 | 10 |
| 2020-07-31__Sector1_002 | 39 | 89.7% | 5.1% | 5.1% | 185 | 25 | 3 |
| 2020-08-01__Sector1_001 | 12 | 50.0% | 0.0% | 50.0% | 65 | 21 | 1 |
| 2020-08-02__Sector1_001 | 331 | 87.9% | 2.4% | 10.0% | 1780 | 315 | 25 |
| 2020-08-04__Sector1_002 | 6 | 83.3% | 0.0% | 16.7% | 34 | 5 | 0 |
| 2020-08-05__Sector1_001 | 17 | 64.7% | 29.4% | 11.8% | 135 | 39 | 2 |
| 2020-08-07__Sector1_001 | 24 | 70.8% | 16.7% | 12.5% | 177 | 81 | 0 |
| 2020-09-02__Sector1_001 | 54 | 87.0% | 5.6% | 9.3% | 280 | 54 | 3 |
| 2020-09-02__Sector1_002 | 31 | 90.3% | 9.7% | 3.2% | 171 | 34 | 4 |
| Sum / Average for Pilots | 766 | 85.9% | 4.8% | 9.9% | 4216 | 814 | 62 |
| Sum / Average for ATCos | 752 | 93.9% | 5.7% | 2.3% | 4026 | 665 | 41 |

**Table 11: Accuracy of automatic command extraction for Pilot utterances**

The table shows the challenges on pilot side. The extraction rate on ATCo side is much better than on pilot side.

Table 12 corresponds to Table 10 and shows the callsign recognition/extraction rate for pilot utterances.

| Day and Work Station | UttCnt | Wrong Csgn | Invented Csgn | No Extraction | No Csgn missed | Break Break |
|---|---|---|---|---|---|---|
| 2020-03-12__Sector1_001 | 56 | 3 | 1 | 1 | 1 | 0 |
| 2020-07-21__Sector1_001 | 22 | 1 | 0 | 1 | 0 | 0 |
| 2020-07-24__Sector1_001 | 18 | 2 | 0 | 1 | 1 | 0 |
| 2020-07-24__Sector1_002 | 38 | 0 | 0 | 0 | 0 | 0 |
| 2020-07-31__Sector1_001 | 25 | 2 | 0 | 0 | 2 | 0 |
| 2020-07-31__Sector1_002 | 14 | 0 | 0 | 0 | 0 | 0 |
| 2020-08-01__Sector1_001 | 16 | 1 | 0 | 1 | 0 | 0 |
| 2020-08-02__Sector1_001 | 29 | 0 | 0 | 0 | 0 | 0 |
| 2020-08-04__Sector1_002 | 15 | 0 | 0 | 0 | 0 | 0 |
| 2020-08-05__Sector1_001 | 5 | 2 | 0 | 2 | 0 | 0 |
| 2020-08-07__Sector1_001 | 204 | 5 | 1 | 3 | 1 | 0 |
| 2020-09-02__Sector1_001 | 3 | 0 | 0 | 0 | 0 | 0 |
| 2020-09-02__Sector1_002 | 10 | 0 | 0 | 0 | 0 | 0 |
| Sum for Pilots | 403 | 11 | 1 | 6 | 4 | 0 |
| Rates for Pilots | 97.3% | CaRR | CaER | 1.2% | RejRR | 1.5% |
| Sum for ATCos | 384 | 10 | 2 | 3 | 5 | 0 |
| Rates for ATCos | 97.4% | CaR | CaE | 1.8% | CaRj | 0.8% |

Table 12: Accuracy of automatic callsign extraction/recognition rates for Pilot utterances

Table 12 and Table 10 also show that the callsign extraction/recognition rates for pilots and ATCos have already the same performance. The improve challenges are on command extraction rates for pilots.

# 5 Current Performance of Automatic Annotation on automatic transcriptions

The previous chapter has shown the current performance of command extraction/recognition on manually transcribed utterance. Good performance on the output of a speech recognizer is, however, more important.

We, therefore, automatically transcribed all manually transcribed data again automatically by a software provided by BUT during April 2021.

The Word Error Rates (WER) calculated are shown in the following two tables Table 14 and Table 13.

| | |
|---|---|
| ATCo WER | 8.0% |
| Pilot WER | 12.2% |
| Total | 10.2% |

Table 14: Average Word Error Rates for Pilot and ATCo, if calculated on all files

| | |
|---|---|
| ATCo WER | 9.2% |
| Pilot WER | 18.3% |
| Total | 14.0% |

Table 15: Average Word Error Rates for Pilot and ATCo, if calculated only on validation files

WER provided in Table 16 are more realistic with respect to expected rates in the future, because these recordings were excluded from the training data.

The following calculation were created on 2021-05-10. The number of annotated files has increased by roughly a factor of two, compared to the results shown in the previous chapter.

| Day and Work Station | # gold | RcR | ErR | RjR | # Words | Unknown Cl | Unknown 2 |
|---|---|---|---|---|---|---|---|
| 2020-03-12__Sector1_001 | 81 | 92.6% | 6.2% | 8.6% | 453 | 119 | 2 |
| 2020-07-21__Sector1_001 | 42 | 92.9% | 2.4% | 4.8% | 229 | 55 | 1 |
| 2020-07-24__Sector1_001 | 61 | 37.7% | 6.6% | 59.0% | 383 | 169 | 6 |
| 2020-07-24__Sector1_002 | 42 | 40.5% | 2.4% | 61.9% | 305 | 156 | 4 |
| 2020-07-31__Sector1_001 | 66 | 31.8% | 3.0% | 66.7% | 398 | 215 | 8 |
| 2020-07-31__Sector1_002 | 30 | 26.7% | 0.0% | 73.3% | 152 | 67 | 5 |
| 2020-08-01__Sector1_001 | 49 | 79.6% | 6.1% | 20.4% | 301 | 93 | 3 |
| 2020-08-02__Sector1_001 | 355 | 85.9% | 4.5% | 11.0% | 1653 | 266 | 15 |
| 2020-08-04__Sector1_002 | 5 | 80.0% | 0.0% | 20.0% | 32 | 7 | 0 |

| Day and Work Station | # gold | RcR | ErR | RjR | # Words | Unknown Cl | Unknown 2 |
|---|---|---|---|---|---|---|---|
| 2020-08-05__Sector1_001 | 81 | 51.9% | 1.2% | 50.6% | 464 | 168 | 4 |
| 2020-08-07__Sector1_001 | 55 | 49.1% | 10.9% | 45.5% | 324 | 148 | 6 |
| 2020-09-02__Sector1_001 | 48 | 45.8% | 33.3% | 31.3% | 450 | 244 | 5 |
| 2020-09-02__Sector1_002 | 27 | 44.4% | 14.8% | 44.4% | 182 | 69 | 7 |
| Sum / Average for ATCos | 942 | 67.3% | 6.3% | 29.7% | 5326 | 1776 | 66 |

**Table 17: Accuracy of automatic command extraction for ATCo utterances, when the input does not result from manual transcription but from automatic transcription**

| Day and Work Station | # gold | RcR | ErR | RjR | # Words | Unknown Cl | Unknown 2 |
|---|---|---|---|---|---|---|---|
| 2020-03-12__Sector1_001 | 55 | 81.8% | 1.8% | 16.4% | 305 | 58 | 17 |
| 2020-07-21__Sector1_001 | 46 | 63.0% | 15.2% | 28.3% | 255 | 62 | 11 |
| 2020-07-24__Sector1_001 | 167 | 71.3% | 4.8% | 26.9% | 820 | 233 | 21 |
| 2020-07-24__Sector1_002 | 142 | 72.5% | 2.1% | 26.1% | 654 | 171 | 23 |
| 2020-07-31__Sector1_001 | 208 | 72.1% | 2.9% | 26.9% | 1029 | 260 | 34 |
| 2020-07-31__Sector1_002 | 68 | 55.9% | 7.4% | 36.8% | 319 | 99 | 11 |
| 2020-08-01__Sector1_001 | 155 | 76.1% | 7.7% | 17.4% | 715 | 147 | 26 |
| 2020-08-02__Sector1_001 | 331 | 68.9% | 9.1% | 24.8% | 1747 | 408 | 64 |
| 2020-08-04__Sector1_002 | 38 | 44.7% | 28.9% | 28.9% | 219 | 48 | 9 |
| 2020-08-05__Sector1_001 | 141 | 76.6% | 5.0% | 20.6% | 661 | 163 | 18 |
| 2020-08-07__Sector1_001 | 98 | 74.5% | 0.0% | 25.5% | 496 | 139 | 12 |
| 2020-09-02__Sector1_001 | 192 | 67.2% | 9.9% | 27.1% | 898 | 184 | 29 |
| 2020-09-02__Sector1_002 | 92 | 70.7% | 13.0% | 19.6% | 499 | 123 | 22 |
| Sum / Average for Pilots | 1733 | 68.9% | 8.3% | 25.0% | 8617 | 2095 | 297 |

**Table 18: Accuracy of automatic command extraction for Pilot utterances, when the input does not result from manual transcription but from automatic transcription**

We calculated also the performance when the predicted callsigns extracted from the surveillance data were not provided for both bases, i.e. when commands are extracted from manually transcribed and also from automatically transcribed utterances.

| | | ATCO | | | Pilot | | |
|---|---|---|---|---|---|---|---|
| | | RcR | ErR | RjR | RcR | ErR | RjR |
| Manual | With Callsigns | 92.57% | 6.90% | 3.40% | 89.84% | 3.75% | 7.33% |
| | No Callsigns | 83.55% | 10.40% | 9.66% | 76.05% | 8.77% | 16.79% |
| Auto-matic | With Callsigns | 67.30% | 6.26% | 29.72% | 70.51% | 6.98% | 24.75% |
| | No Callsigns | 58.70% | 8.17% | 36.73% | 56.20% | 12.81% | 33.76% |

**Table 19: Performance of automatic command extraction for ATCo and Pilot utterances, when manual transcribed (manual) versus automatically transcribed (automatic) and when callsign information is provided (With Callsigns) versus when no callsign information from command prediction is provided (No Callsigns)**

Table 19 shows that the extraction recognition performance for ATCO commands goes down from 92.57% to 67.3% when the extraction has only input automatic transcriptions. The decrease is not so high for the pilot side, although the WER is much worse for pilot utterances. One reason is that all utterances (also those included in training set) are considered here. The recognition error rates on ATCO do not increase and on pilot side only small from 3.75% to 6.98%.

Table 19 also shows that the recognition rate RcR goes down for both pilot and ATCo utterance, when no callsign information is provided any more. This is also true for the recognition error rate ErR. The results also do not depend on whether the input comes from manual or automatic transcriptions.

The following Table 20 provides the performance when only the callsigns are considered.

| | | ATCO | | | Pilot | | |
|---|---|---|---|---|---|---|---|
| | | CaR | CaE | CaRj | CaR | CaE | CaRj |
| Manual | With Callsigns | 97.2% | 1.7% | 1.1% | 97.3% | 0.8% | 1.8% |
| | No Callsigns | 88.5% | 5.8% | 5.8% | 82.6% | 7.1% | 10.3% |
| Auto-matic | With Callsigns | 78.0% | 3.6% | 18.3% | 82.7% | 4.0% | 13.3% |
| | No Callsigns | 68.4% | 7.2% | 24.3% | 66.6% | 12.4% | 20.9% |

**Table 20: Performance of automatic callsign extraction for ATCo and Pilot utterances, when manual transcribed (manual) versus automatically transcribed (automatic) and when callsign information of available callsigns is provided (With Callsigns) versus when no callsign information from command prediction is provided (No Callsigns)**

The results are as before. Automatic transcriptions result in a performance decrease and callsign information of the callsigns from the surveillance data are of decisive importance.

# 6 Next Steps

Currently roughly 90 minutes of the manual transcribed data are already manual annotated. The rest of the 7.5 hours is already automatically annotated.

During the following next months, more and more data will be manual checked, so that more and more validation data is available also for quality assurance of command extraction.

The main challenge, however, is to improve extraction rate on the automatic transcriptions, i.e. the output of the speech to text block.

## 6.1 Checking for unused word sequences

The following Table 21 shows an excerpt, of which word sequences were not used for (automatic) extracting commands. In the left part, we see the extractions from the manual transcriptions (from so called jcor files) and in the right part from automatically generated transcriptions (from so called jtxt files) being created beginning of April 2021.

| | |
|---|---|
| arrange your flight: 3 | |
| arrival: 5 | |
| | arrival: 6 |
| | artic: 150 |
| | artic eagle: 146 |
| | artic eagle seven: 89 |
| | artic eagle seven five: 32 |
| | artic eagle seven five seven: 13 |
| | artic eagle seven five three: 3 |
| | artic eagle seven five two: 14 |
| | artic eagle seven five two that: 3 |
| | … |
| | artic eagle seven two: 14 |
| | artic eagle seven two one: 10 |
| | artic eagle seven two zero: 4 |
| | artic eagle three: 12 |
| | artic eagle three four: 4 |
| | artic eagle three three: 6 |
| ascot: 10 | ascot: 11 |
| at: 66 | at: 67 |
| at all: 3 | at all: 3 |
| at keflavik: 10 | at keflavik: 11 |
| at one: 3 | at one: 4 |
| at the: 4 | at the: 4 |
| at the moment: 3 | at the moment: 3 |
| at zero: 4 | at zero: 4 |

**Table 21: Occurrence count, how often some word sequences were observed**

We see e.g. that the word sequence “at Keflavik” is 10 times and 11 times resp. not used for extracting a command. This could be a hint, that information is lost independent of automatic or manual transcription is used.

On the other hand, we also observe that the word sequences “artic” or “artic eagle” are not used, when the word sequences result from automatic transcriptions. The reason here is, that the airline name is “arctic eagle” with a “c” and not “artic eagle”. Something systematically went wrong in Speech-

to-Text" transformation. These problems need to be corrected in Speech-to-Text transformation or if not possible in the Text-to-Concept (command extraction) block.

## 6.2 Improving wrong extraction

For 90 minutes, already the gold annotations are specified. This, however, does not mean that these command sequences are already correctly extracted. Otherwise the command extraction rates should be 100% and the command extraction error rates should be 0%, which is not the case, as shown in the previous chapters.

We show an example:

```
 delta  two four reykjavik roger good morning identified
  csgn csgn csgn      valu  type type    type       type
Gold Commands:
   DAL24 STATION REYK_RADAR
   DAL24 GREETING
   DAL24 INIT_RESPONSE
Extracted Commands:
   DAL24 STATION REYK_RADAR // plaus: 0.6
   DAL24 AFFIRM
   DAL24 GREETING
   DAL24 INIT_RESPONSE
```

AFFIRM is extracted, but not expected. "roger" should, however, also result in "AFFIRM". In this case the gold (manual) annotation needs to be corrected.

The next example shows a wrong extraction

```
        are  you able flight level three seven zero
       type type unkn   unit  unit  valu  valu valu
     Gold Commands:
        NO_CALLSIGN REPORT_MISCELLANEOUS
     Extracted Commands:
        NO_CALLSIGN REPORT_MISCELLANEOUS
        NO_CALLSIGN ALTITUDE 370 FL // plaus: 0.6
```

A clearance to "ALTITUDE 370 FL" is extracted, but this is wrong. It is just a question and not a real clearance. Therefore, here the extraction must be improved.

In "okay i just gonna check you gonna stay five miles behind the delta two four your heading for the same destination i am just gonna see with norway control if you are converting or splitting up" the callsign DAL24 is accidently extracted. It is not a command for the DAL24. It should be "NO_CALLSIGN NO_CONCEPT".

In the following weeks the wrong extractions will be systematically analysed. Either the gold annotation will be corrected or which costs more effort, the command extraction needs to be improved, which will result not only in changing the extraction from one utterance, but for many cases.

Some of them will be correct, other extraction after the "improvement" will not be correct. Systematic checking again is needed.

The following tables Table 22 to Table 26 provide the extraction performances for all types currently considered for Isavia airspace.

Types, which are marked with yellow colour in column "Type" did not occur. It is assumed that they also will not occur in the remaining data. The will be excluded in the next versions from command extraction, which will slightly speed up extraction process and slightly reduce the extraction error rate, provided that they really do not appear.

| Type | Total | User-Rec | GoldPr | Err | Rej | Relevance | Rec-Rate | GoldErr | Diagnostics |
|---|---|---|---|---|---|---|---|---|---|
| AFFIRM | 63 | 54 | 0 | 0 | 2 | 2.2% | 85.7% | 0.0% | bad extractic |
| ALTITUDE | 318 | 276 | 34 | 28 | 10 | 11.1% | 86.8% | 10.7% | bad extractic |
| ALTITUDE BETWEEN | 0 | 0 | 5 | 0 | 0 | 0.0% | | 100.0% | missing implementation |
| CALL_YOU_BACK | 22 | 21 | 1 | 0 | 0 | 0.8% | 95.5% | 4.5% | bad extractic |
| CANCEL MISCELLANEOUS | 0 | 0 | 0 | 0 | 0 | 0.0% | | | |
| CLEARED APPROACH | 0 | 0 | 1 | 0 | 0 | 0.0% | | 100.0% | |
| CLEARED RNAV | 0 | 0 | 1 | 0 | 0 | 0.0% | | 100.0% | |
| CLEARED TO | 0 | 0 | 1 | 0 | 0 | 0.0% | | 100.0% | |
| CLEARED VIA | 0 | 0 | 0 | 0 | 0 | 0.0% | | | |
| CLIMB | 164 | 158 | 9 | 4 | 0 | 5.7% | 96.3% | 5.5% | bad extractic |
| CLIMB BETWEEN | 0 | 0 | 0 | 0 | 0 | 0.0% | | | |
| CONFIRM_ACCEPT | 0 | 0 | 0 | 0 | 0 | 0.0% | | | |
| CONTACT | 159 | 145 | 11 | 8 | 2 | 5.5% | 91.2% | 6.9% | bad extractic |
| CONTACT_FREQUENCY | 242 | 229 | 9 | 5 | 2 | 8.4% | 94.6% | 3.7% | bad extractic |
| CONTINUE PRESENT_HEAI | 0 | 0 | 0 | 0 | 0 | 0.0% | | | |
| CORRECTION | 4 | 4 | 3 | 0 | 0 | 0.1% | 100.0% | 75.0% | |
| CPDLC | 101 | 99 | 1 | 1 | 0 | 3.5% | 98.0% | 1.0% | |

**Table 22: Extraction Performance for selected types (letters A-C)**

Column "Total" shows how often the type occurs in already annotated commands from the pilot and from the ATCo. Column "User-Rec" specifies how often the type was fully recognized correctly, which includes "callsign fully correct", "type and second type fully correct", "unit correct", "qualifier correct" and also that the "condition is fully correct".

| Type | Total | User-Rec | GoldPr | Err | Rej | Relevance | Rec-Rate | GoldErr | Diagnosti |
|---|---|---|---|---|---|---|---|---|---|
| DESCEND | 63 | 54 | 21 | 5 | 1 | 2.2% | 85.7% | 33.3% | bad extractic |
| DESCEND BETWEEN | 0 | 0 | 0 | 0 | 0 | 0.0% | | | |
| DIRECT_TO | 210 | 165 | 25 | 30 | 13 | 7.3% | 78.6% | 11.9% | bad extractic |
| DISREGARD | 1 | 1 | 0 | 0 | 0 | 0.0% | 100.0% | 0.0% | |
| EXPECT ILS | 0 | 0 | 0 | 0 | 0 | 0.0% | | | |
| EXPECT RUNWAY | 0 | 0 | 0 | 0 | 0 | 0.0% | | | |
| EXPECT_ROUTE | 0 | 0 | 0 | 0 | 0 | 0.0% | | | |
| EXPEDITE_PASSING | 0 | 0 | 0 | 0 | 0 | 0.0% | | | |
| FAREWELL | 166 | 159 | 3 | 1 | 4 | 5.8% | 95.8% | 1.8% | bad extractic |
| FOLLOW_ROUTE | 0 | 0 | 3 | 0 | 0 | 0.0% | | 100.0% | |
| GREETING | 344 | 337 | 3 | 1 | 0 | 12.0% | 98.0% | 0.9% | bad extractic |
| HEADING | 0 | 0 | 1 | 0 | 0 | 0.0% | | 100.0% | |
| HEADING LEFT | 1 | 1 | 0 | 0 | 0 | 0.0% | 100.0% | 0.0% | |
| HEADING RIGHT | 1 | 1 | 0 | 0 | 0 | 0.0% | 100.0% | 0.0% | |
| HEADING none | 11 | 4 | 0 | 6 | 1 | 0.4% | 36.4% | 0.0% | bad extractic |
| HOLDING | 2 | 1 | 2 | 1 | 0 | 0.1% | 50.0% | 100.0% | |
| INCREASE BETWEEN | 0 | 0 | 0 | 0 | 0 | 0.0% | | | |
| INFORMATION ACTIVE_R\ | 10 | 7 | 3 | 1 | 2 | 0.3% | 70.0% | 30.0% | |
| INFORMATION MISCELLAI | 0 | 0 | 3 | 0 | 0 | 0.0% | | 100.0% | |
| INFORMATION QNH | 41 | 40 | 1 | 0 | 0 | 1.4% | 97.6% | 2.4% | bad extractic |
| INFORMATION TRAFFIC | 10 | 7 | 1 | 1 | 1 | 0.3% | 70.0% | 10.0% | |
| INFORMATION WINDDIRE | 2 | 0 | 1 | 1 | 1 | 0.1% | 0.0% | 50.0% | |
| INFORMATION WINDSPEE | 2 | 1 | 0 | 1 | 0 | 0.1% | 50.0% | 0.0% | |
| INIT_RESPONSE | 145 | 141 | 1 | 2 | 1 | 5.1% | 97.2% | 0.7% | bad extractic |

**Table 23: Extraction Performance for selected types (letters D-I)**

"Err" counts the numbers of errors for this type, which means here that the type is extracted, but the command does not appear in the gold annotations, either it does not appear at call, or the callsign was wrong or the type etc.

"Rej" counts the rejections for this type, which means here that the type is extracted, but the command does not appear in the gold annotations, either it does not appear at call, or the callsign was wrong or the type etc. In contrast "Err" the extracted command is rejected, because its type is either "NO_CONCEPT" or "NO_CALLSIGN" was extracted.

Column "GoldPtr" counts the numbers of errors, when this type appears in the gold annotations, but the type was not extracted at all from this utterance and the number of extractions is less than the number of gold annotation in this utterance.

| Type | Total | User-Rec | GoldPr | Err | Rej | Relevance | Rec-Rate | GoldErr | Diagnostic |
|---|---|---|---|---|---|---|---|---|---|
| LEAVE_FREQUENCY | 0 | 0 | 0 | 0 | 0 | 0.0% | | | |
| LEAVE_HOLDING | 0 | 0 | 0 | 0 | 0 | 0.0% | | | |
| MAINTAIN ALTITUDE | 38 | 28 | 11 | 8 | 1 | 1.3% | 73.7% | 28.9% | bad extractic |
| MAINTAIN HEADING | 0 | 0 | 0 | 0 | 0 | 0.0% | | | |
| MAINTAIN PRESENT_SPEE | 0 | 0 | 0 | 0 | 0 | 0.0% | | | |
| MAINTAIN SPEED | 0 | 0 | 0 | 0 | 0 | 0.0% | | | |
| NAVIGATION_OWN | 2 | 2 | 0 | 0 | 0 | 0.1% | 100.0% | 0.0% | |
| NEGATIVE | 7 | 7 | 0 | 0 | 0 | 0.2% | 100.0% | 0.0% | |
| NO_CONCEPT | 176 | 147 | 17 | 0 | 21 | 6.1% | 83.5% | 9.7% | bad extractic |
| NO_SPEED_RESTRICTIONS | 5 | 5 | 0 | 0 | 0 | 0.2% | 100.0% | 0.0% | |
| ORBIT | 1 | 0 | 0 | 1 | 0 | 0.0% | 0.0% | 0.0% | |

**Table 24: Extraction Performance for selected types (letters K-Q)**

Column "Relevance" counts the relevance of this type. The type "MAINTAIN ALTITYPE" appears 38 times, which means that 38 makes 1.3% of call commands.

"Rec-Rate" is the "Command Recognition rate for this type, column "User-Rec" divided by column "Total". "GoldErr" is column "GoldPtr" divided by column "Total".

| Type | Total | User-Rec | GoldPr | Err | Rej | Relevance | Rec-Rate | GoldErr | Diagnostics | |
|---|---|---|---|---|---|---|---|---|---|---|
| RATE_OF_CLIMB | 0 | 0 | 0 | 0 | 0 | 0.0% | | | | |
| RATE_OF_CLIMB BETWEEN | 0 | 0 | 0 | 0 | 0 | 0.0% | | | | |
| RATE_OF_CLIMB EXPEDITE | 0 | 0 | 0 | 0 | 0 | 0.0% | | | | |
| RATE_OF_CLIMB MAX | 0 | 0 | 0 | 0 | 0 | 0.0% | | | | |
| RATE_OF_CLIMB OWN | 0 | 0 | 0 | 0 | 0 | 0.0% | | | | |
| RATE_OF_DESCENT | 0 | 0 | 0 | 0 | 0 | 0.0% | | | | |
| RATE_OF_DESCENT BETWI | 0 | 0 | 0 | 0 | 0 | 0.0% | | | | |
| RATE_OF_DESCENT EXPED | 0 | 0 | 0 | 0 | 0 | 0.0% | | | | |
| RATE_OF_DESCENT MAX | 0 | 0 | 0 | 0 | 0 | 0.0% | | | | |
| RATE_OF_DESCENT OWN | 0 | 0 | 0 | 0 | 0 | 0.0% | | | | |
| REDUCE BETWEEN | 0 | 0 | 0 | 0 | 0 | 0.0% | | | | |
| REDUCE_BY | 0 | 0 | 0 | 0 | 0 | 0.0% | | | | |
| REPORT CROSSING | 0 | 0 | 0 | 0 | 0 | 0.0% | | | | |
| REPORT LEVEL_PASSING | 0 | 0 | 5 | 0 | 0 | 0.0% | | 100.0% | | missing implementation |
| REPORT PASSING | 2 | 0 | 1 | 2 | 0 | 0.1% | 0.0% | 50.0% | | |
| REPORT POINT | 0 | 0 | 0 | 0 | 0 | 0.0% | | | | |
| REPORT_MISCELLANEOUS | 27 | 22 | 4 | 3 | 0 | 0.9% | 81.5% | 14.8% | bad extractic | |
| REPORT_NOW ALTITUDE | 1 | 1 | 0 | 0 | 0 | 0.0% | 100.0% | 0.0% | | |
| REPORT_NOW FLIGHT_LEV | 0 | 0 | 1 | 0 | 0 | 0.0% | | 100.0% | | |
| REPORT_NOW HEADING | 0 | 0 | 0 | 0 | 0 | 0.0% | | | | |
| REPORT_NOW POSITION | 0 | 0 | 0 | 0 | 0 | 0.0% | | | | |
| REPORT_NOW SPEED | 0 | 0 | 0 | 0 | 0 | 0.0% | | | | |
| REPORT_NOW SPEED_MA | 0 | 0 | 0 | 0 | 0 | 0.0% | | | | |
| RESUME_NORMAL_SPEED | 1 | 1 | 0 | 0 | 0 | 0.0% | 100.0% | 0.0% | | |

**Table 25: Extraction Performance for selected types (letter R)**

Column "diagnostics" shows a hint for future work. "bad extraction" means here, that the recognition performance is not good enough, which is the case, when column "Rec-Rate" is below 92% and the number of occurrences of this type in column "Total" is above 10. The column contains "missing implementation", when this command type is never extracted (column "Total" is 0) and it was

annotated, i.e. “GoldErr“ column is greater 3. If we mark the column “Diagnostics” with yellow, these types will be improved in the near future.

| Type | Total | User-Rec | GoldPr | Err | Rej | Relevance | Rec-Rate | GoldErr | Diagnos |
|---|---|---|---|---|---|---|---|---|---|
| SPEED BETWEEN | 0 | 0 | 0 | 0 | 0 | 0.0% | | | |
| SQUAWK | 42 | 39 | 11 | 1 | 0 | 1.5% | 92.9% | 26.2% | bad extractic |
| STATION | 454 | 441 | 5 | 3 | 2 | 15.8% | 97.1% | 1.1% | bad extractic |
| STOP_ALTITUDE | 0 | 0 | 0 | 0 | 0 | 0.0% | | | |
| STOP_CLIMB | 0 | 0 | 0 | 0 | 0 | 0.0% | | | |
| STOP_DESCEND | 0 | 0 | 0 | 0 | 0 | 0.0% | | | |
| TRANSITION | 1 | 0 | 1 | 1 | 0 | 0.0% | 0.0% | 100.0% | |
| TURN LEFT | 0 | 0 | 1 | 0 | 0 | 0.0% | | 100.0% | |
| TURN RIGHT | 0 | 0 | 0 | 0 | 0 | 0.0% | | | |
| TURN_BY LEFT | 2 | 2 | 0 | 0 | 0 | 0.1% | 100.0% | 0.0% | |
| TURN_BY RIGHT | 2 | 2 | 0 | 0 | 0 | 0.1% | 100.0% | 0.0% | |
| UNRECOGNIZED_CONCEP | 0 | 0 | 0 | 0 | 77 | 0.0% | | | |
| VERTICAL_RATE | 0 | 0 | 0 | 0 | 0 | 0.0% | | | |
| VERTICAL_RATE BETWEEN | 0 | 0 | 0 | 0 | 0 | 0.0% | | | |
| VERTICAL_RATE EXPEDITE | 0 | 0 | 0 | 0 | 0 | 0.0% | | | |
| VERTICAL_RATE MAX | 0 | 0 | 0 | 0 | 0 | 0.0% | | | |
| VERTICAL_RATE OWN | 0 | 0 | 0 | 0 | 0 | 0.0% | | | |
| Sums | 2843 | 2602 | 201 | 115 | 141 | 99.0% | | | |

**Table 26: Extraction Performance for selected types (letters S-Z and sums)**

The types “ALTITUDE BETWEEN”, “DESCEND BETWEEN” and “CLIMB BETWEEN” are currently not modelled at all. They occur, i.e. for the following utterances with the keyword “block”

```
reykjavik control myflug seven seven alfa request flight level  one seven zero block  one eight zero
     valu    valu   csgn  csgn  csgn csgn    type   type  type valu  valu valu  unkn unkn  unkn unkn
                                                                              ######################

         Gold Commands:
  MYA77A PILOT STATION REYK_RADAR
  MYA77A PILOT REQUEST ALTITUDE BETWEEN 170 180 FL
Extracted Commands:
  MYA77A PILOT STATION REYK_RADAR // plaus: 0.6
  MYA77A PILOT REQUEST ALTITUDE 170 FL
```

The following utterances show current problems with the type CLIMB “after passing rixun climb flight level three seven zero”. Here the condition “WHEN PASSING RIXUN” is missing. And in “climbing  two nine  two nine zero  to gunpa” nothing is recognized. It is a wrong command. The correct phraseology would have been “climbing two nine **correction** two nine zero to gunpa”.

Many of the problems with DIRECT_TO are related to lat/long coordinates as in “direct  six three north three zero west”, which should result in “DIRECT_TO 63N_30W none”, which is currently not suppoted.

The errors resulting from “MAINTAIN ALTITUDE” are due to the fact, that here often the wrong unit is extracted as e.g. in “maintaining level three nine zero” which currently results in “MAINTAIN ALTITUDE 390 none” which is of course wrong. It must be “MAINTAIN ALTITUDE 390 FL”.

Table 27 only considers command types which are either significant or at least occur at least 10 times.

| Type | Total | User-Rec | Err | Relevance | Rec-Rate | Delta Rec | Delta Rec |
|---|---|---|---|---|---|---|---|
| AFFIRM | 67 | 34 | 2 | 2.3% | 50.7% | -20 | -35.0% |
| ALTITUDE | 310 | 235 | 22 | 10.8% | 75.8% | -41 | -11.0% |
| CALL_YOU_BACK | 21 | 14 | 0 | 0.7% | 66.7% | -7 | -28.8% |
| CLIMB | 162 | 130 | 8 | 5.6% | 80.2% | -28 | -16.1% |
| CONTACT | 148 | 105 | 8 | 5.2% | 70.9% | -40 | -20.2% |
| CONTACT_FREQUENCY | 236 | 189 | 4 | 8.2% | 80.1% | -40 | -14.5% |
| CORRECTION | 5 | 3 | 0 | 0.2% | 60.0% | -1 | -40.0% |
| CPDLC | 102 | 95 | 1 | 3.6% | 93.1% | -4 | -4.9% |
| DESCEND | 65 | 40 | 5 | 2.3% | 61.5% | -14 | -24.2% |
| DIRECT_TO | 212 | 123 | 38 | 7.4% | 58.0% | -42 | -20.6% |
| FAREWELL | 110 | 90 | 5 | 3.8% | 81.8% | -69 | -14.0% |
| GREETING | 215 | 199 | 4 | 7.5% | 92.6% | -138 | -5.4% |
| HEADING none | 11 | 4 | 6 | 0.4% | 36.4% | 0 | 0.0% |
| INFORMATION ACTIVE_RWY | 12 | 7 | 2 | 0.4% | 58.3% | 0 | -11.7% |
| INFORMATION QNH | 37 | 21 | 0 | 1.3% | 56.8% | -19 | -40.8% |
| INFORMATION TRAFFIC | 11 | 4 | 2 | 0.4% | 36.4% | -3 | -33.6% |
| INIT_RESPONSE | 143 | 121 | 1 | 5.0% | 84.6% | -20 | -12.6% |
| MAINTAIN ALTITUDE | 36 | 23 | 9 | 1.3% | 63.9% | -5 | -9.8% |
| NEGATIVE | 7 | 5 | 0 | 0.2% | 71.4% | -2 | -28.6% |
| NO_CONCEPT | 194 | 115 | 0 | 6.8% | 59.3% | -32 | -24.2% |
| REPORT_MISCELLANEOUS | 25 | 16 | 2 | 0.9% | 64.0% | -6 | -17.5% |
| SQUAWK | 36 | 31 | 2 | 1.3% | 86.1% | -8 | -6.7% |
| STATION | 454 | 373 | 8 | 15.8% | 82.2% | -68 | -15.0% |
| Sums | 2643 | 1990 | 134 | 92.1% | | -612 | |

**Table 27: Extraction Performance for selected types on manual and automatic transcriptions**

We compare the command recognition rates from manual transcriptions with those from automatic transcriptions. In the last two columns we see difference between “UserRec” for automatic transcriptions and those for manual transcription and also the difference in “Rec-Rate”. We mark in light brown the types, for which we see big differences, i.e. at least 10 commands less.

## 6.3 Unclassified word sequences

Additional to the already mentioned steps for improvement we will concentrate on word sequences, which were currently not used for classification. This includes manual and automatic transcriptions with emphasis first on manual transcriptions.

The following extraction shows such an example in which “bravo india romeo delta” is accidently recognized as a callsign.

```
          are  you able  to  log  on  to cpdlc bravo india romeo delta
         type type unkn unkn unkn unkn unkn  unkn  csgn  csgn  csgn  csgn
                ##############################
       Gold Commands:
          NO_CALLSIGN REPORT_MISCELLANEOUS
       Extracted Commands:
          BIRD REPORT_MISCELLANEOUS
```

In the following example “request” is not used for command extraction, which results in the wrong command extraction.

```
          request direct romeo kilo india three  one intermediate  fix  for runway three  one
             unkn   type  valu valu  valu  valu valu          unkn unkn unkn   valu  valu valu
         ########                                  ########################
         Gold Commands:
            NO_CALLSIGN PILOT REQUEST DIRECT_TO RKI31 none
            NO_CALLSIGN PILOT REQUEST CLEARED APPROACH RW31_BIRK
         Extracted Commands:
            NO_CALLSIGN PILOT DIRECT_TO RKI31 none
            NO_CALLSIGN PILOT INFORMATION ACTIVE_RWY RW31_BIRK // plaus: 0.6
```

## 6.4 Requirements from D1-1

The Operational Concept Document D1-1 [2], update in D6-2 [3] contains many transcription examples from Islandic airspace together with the suggested annotations.

The following lines show the remaining challenges, i.e. which transcription examples are currently not correctly extracted:

Shown is always the challenge, i.e. what is currently not correctly implemented, the transcription, i.e. the word sequence, the classification at word level for each word. If more than one consecutive unknown classification occurs, we also mark with "#" characters. At the end the current extracted annotation is also shown.

The Priority shows when it is planned to start the implementation of these annotations. Prio 1 has the highest priority. It should be implemented during the next weeks. Prio 2 is planned for the next two months and Prio 3 was never or very seldom observed in the present transcribed data. So it might be postponed until the end of the project. Prio 4 shows example for which currently it is not clear how to model / annotate the utterance. More examples would be needed.

```
<method name="libASR::IsaviaFreqUsedExamples::climbAndBlockExtraction()">
   CLIMB BETWEEN is the challenge; Prio 1
    iceair zero zero seven climb   to flight level three twenty block three sixty
      csgn csgn csgn  csgn  type type   unit  unit  valu   valu  unkn  unkn  unkn
                                                          ##################
   ICE007 CLIMB 320 FL

<!--  name="libASR::IsaviaFreqUsedExamples::climbAndBlockExtraction()" --></method>
<method name="libASR::IsaviaFreqUsedExamples::climbToReachByEndpointExtraction()">
   WHEN TIME 1200 is the challenge; Prio 2
    austrian  two three four climb   to reach flight level three ninety   at  one  two zero zero
        csgn csgn  csgn csgn  type type  unkn   unit  unit  valu   valu unkn unkn unkn unkn unkn
                                  ######                        ########################
   AUA234 CLIMB 390 FL

<!--  name="libASR::IsaviaFreqUsedExamples::climbToReachByEndpointExtraction()" --></method>
<method name="libASR::IsaviaFreqUsedExamples::climbToReachBeforeSignificantPointPt()">
   UNTIL PASSING 018W is the challenge; Prio 1
    american five five five climb   to reach flight level three ninety   by  one eight west
        csgn csgn csgn csgn  type type  unkn   unit  unit  valu   valu unkn unkn  unkn unkn
                                 ######                       ####################
   AAL555 CLIMB 390 FL

<!--  name="libASR::IsaviaFreqUsedExamples::climbToReachBeforeSignificantPointPt()" --></method>
<method name="libASR::IsaviaFreqUsedExamples::climbAfterPassingExtraction()">
   AFTER PASSING 026-30W   is the challenge; Prio 1
    faxi five  six after passing  two  six three zero west climb flight level three nine zero
    csgn csgn csgn  unkn    unkn unkn unkn  unkn unkn unkn  type   unit  unit  valu valu valu
            ########################################
   FXI56 CLIMB 390 FL

<!--  name="libASR::IsaviaFreqUsedExamples::climbAfterPassingExtraction()" --></method>
<method name="libASR::IsaviaFreqUsedExamples::climbAfterPassingExtraction()">
   WHEN TIME 1303   is the challenge; Prio 2
    executive three alfa bravo   at  one three zero three climb flight level four five zero
         csgn  csgn csgn  csgn unkn unkn  unkn unkn  unkn  type   unit  unit valu valu valu
                          ##########################
   EXT3AB CLIMB 450 FL

<!--  name="libASR::IsaviaFreqUsedExamples::climbAfterPassingExtraction()" --></method>
<method name="libASR::IsaviaFreqUsedExamples::descendBelowCAExtraction()">
   WHEN PASSING MY  is the challenge; Prio 2
    pegasus zero zero  two   at mike yankee descend below controlled airspace
```

```
          csgn csgn csgn csgn unkn unkn    unkn     type  type       type     type
                              #################
      MVM002 DESCEND none none

   <!--  name="libASR::IsaviaFreqUsedExamples::descendBelowCAExtraction()" --></method>
   <method name="libASR::IsaviaFreqUsedExamples::climbViaExtraction()">
      2nd command FOLLOW ROUTE SID is the challenge, Prio 2
       india  fox india climb  via  sid   to flight level  180
        unkn unkn  unkn  unkn unkn unkn unkn   unkn  unkn unkn
      ########################################################
      NO_CALLSIGN NO_CONCEPT

   <!--  name="libASR::IsaviaFreqUsedExamples::climbViaExtraction()" --></method>
   <method name="libASR::IsaviaFreqUsedExamples::cancelLevelRestrClimbExtraction()">
      FOLLOW SID, NO_ALTI_RESTRICTIONS with condition are the challenges; Prio 3
       west  jet  six  six  six climb  via  sid   to flight level three four zero cancel level restrictions   at kilo  fox
india
       csgn csgn csgn csgn csgn  type unkn unkn unkn   unit  unit  valu valu valu   unkn  unkn         unkn unkn unkn unkn  unkn
                                 ###############                          ##############################################
      WJA666 CLIMB 340 FL

   <!--  name="libASR::IsaviaFreqUsedExamples::cancelLevelRestrClimbExtraction()" --></method>
   <method name="libASR::IsaviaFreqUsedExamples::cancelSpeedRestrClimbExtraction()">
      FOLLOW SID, NO_ALTI_RESTRICTIONS with condition WHEN PASSING 100 FL are the challenges; Prio 3
       survey nine bravo climb  via  sid   to flight level  two five zero cancel speed restrictions   at flight level  one zero
zero
        csgn csgn  csgn  type unkn unkn unkn   unit  unit valu valu valu   unkn  unkn         unkn unkn   unit  unit valu valu
valu
                           ###############                          ##############################
      SUY9B CLIMB 250 FL
      SUY9B ALTITUDE 100 FL // plaus: 0.6

   <!--  name="libASR::IsaviaFreqUsedExamples::cancelSpeedRestrClimbExtraction()" --></method>
   <method name="libASR::IsaviaFreqUsedExamples::cancelSpeedRestrDescendExtraction()">
      NO_SPEED_RESTRICTIONS with condition WHEN PASSING VM are the challenges; Prio 3
        top swiss  one bravo descend  via star   to flight level  one zero zero cancel speed restrictions   at victor mike
       unkn  unkn csgn  csgn    type unkn unkn unkn   unit  unit valu valu valu   unkn  unkn         unkn unkn   valu valu
      ###########                ###############                          ##############################
      EZS1B DESCEND 100 FL
      EZS1B DIRECT_TO VM none // plaus: 0.6

   <!--  name="libASR::IsaviaFreqUsedExamples::cancelSpeedRestrDescendExtraction()" --></method>
   <method name="libASR::IsaviaFreqUsedExamples::cancelLevelRestrDescendExtraction()">
      NO_ALTI_RESTRICTIONS with condition WHEN PASSING KFV 120 NM; Prio 3
       romeo yankee romeo  one  two  two descend  via star   to flight level  one zero zero cancel level restrictions  one  two
zero miles from kilo  fox victor
       csgn   csgn  csgn csgn csgn csgn    type unkn unkn unkn   unit  unit valu valu valu   unkn  unit         unkn valu valu
valu  unkn unkn valu valu   valu
                                          ###############                                  #######       #############
###########
      RYR122 DESCEND 100 FL
      RYR122 ALTITUDE 120 FL // plaus: 0.6
      RYR122 DIRECT_TO KFV none // plaus: 0.6

   <!--  name="libASR::IsaviaFreqUsedExamples::cancelLevelRestrDescendExtraction()" --></method>
   <method name="libASR::IsaviaFreqUsedExamples2::reclearedMach()">
      speed value as mach is the challenge, Prio 2
        air france zero three three recleard mach zero eight zero   in reykjavik area
       csgn   csgn csgn  csgn  csgn     unkn unkn valu  valu valu unkn      valu unkn
                                 ##############               #####          #####
      AFR033 HEADING 080 none // plaus: 0.6
      AFR033 STATION REYK_RADAR // plaus: 0.6

   <!--  name="libASR::IsaviaFreqUsedExamples2::reclearedMach()" --></method>
   <method name="libASR::IsaviaFreqUsedExamples2::reclearedDirectToWithCondition()">
      DIRECT_TO 61N_030W AFTER PASSING 64N_020W and ignoring rest    is the challenge, Prio 1
       scandinavian nine four  six recleard after passing  six four north  two zero west  via  six  one north three zero west
rest   of clearance unchanged
              csgn csgn csgn csgn     unkn  unkn    unkn unkn unkn  unkn unkn unkn unkn unkn unkn unkn  unkn  unkn unkn unkn
unkn unkn      unkn      unkn

#########################################################################################################################
      SAS946 NO_CONCEPT

   <!--  name="libASR::IsaviaFreqUsedExamples2::reclearedDirectToWithCondition()" --></method>
   <method name="libASR::IsaviaFreqUsedExamples2::enterCAAtLevel()">
      ENTER_CA WHEN PASSING 190 FL    is the challenge, Prio 3
       greenland  air  two four  six enter controlled airspace   at flight level  one nine zero
            unkn unkn csgn csgn csgn  unkn       unkn     unkn unkn   unit  unit valu valu valu
      ###############              ################################
      GRL246 ALTITUDE 190 FL // plaus: 0.6

   <!--  name="libASR::IsaviaFreqUsedExamples2::enterCAAtLevel()" --></method>
   <method name="libASR::IsaviaFreqUsedExamples2::enterCAAtTime()">
      ENTER_CA WHEN TIME 1245 OR_LATER    is the challenge, Prio 3
       faxi three zero zero enter controlled airspace   at time  one  two four five   or later
       csgn  csgn csgn csgn  unkn       unkn     unkn unkn unkn unkn unkn unkn unkn  unkn
                        ###################################################################
      FXI300 NO_CONCEPT

   <!--  name="libASR::IsaviaFreqUsedExamples2::enterCAAtTime()" --></method>
   <method name="libASR::IsaviaFreqUsedExamples2::leaveCAAtLevel()">
      LEAVE_CA WHEN PASSING 120 FL    is the challenge, Prio 3
       november  two charlie tango leave controlled airspace   at flight level  one  two zero
           csgn csgn    csgn  csgn  unkn       unkn     unkn unkn   unit  unit valu valu valu
```

```
                                ##############################
   N2CT ALTITUDE 120 FL // plaus: 0.6

<!--  name="libASR::IsaviaFreqUsedExamples2::leaveCAAtLevel()" --></method>
<method name="libASR::IsaviaFreqUsedExamples2::leaveCAAtTime()">
   LEAVE_CA WHEN TIME 1919 OR_EARLIER    is the challenge, Prio 3
    danish  two rescue leave controlled airspace   at time  one nine  one nine   or earlier
      csgn csgn   unkn  unkn       unkn     unkn unkn unkn unkn unkn unkn unkn unkn    unkn
              ###########################################################################
   DTR2 NO_CONCEPT

<!--  name="libASR::IsaviaFreqUsedExamples2::leaveCAAtTime()" --></method>
<method name="libASR::IsaviaFreqUsedExamples2::flightPlanRouteToExtraction()">
   DIRECT_TO 69N_020W WHEN PASSING YDP, flight plan route ignored   is the challenge, Prio 2
    connie  two tango from yankee delta papa flight plan route   to  six nine north  two zero west
      csgn csgn  csgn unkn   valu  valu valu   unkn unkn  unkn unkn unkn unkn  unkn unkn unkn unkn
                  #####                ######################################################
   CKS2T DIRECT_TO YDP none // plaus: 0.6

<!--  name="libASR::IsaviaFreqUsedExamples2::flightPlanRouteToExtraction()" --></method>
<method name="libASR::IsaviaFreqUsedExamples2::maintainAltitudeUntilPassing()">
   UNTIL PASSING ING,   is the challenge, Prio 2
    austrian five five maintain flight level three  two zero   to india november golf
        csgn csgn csgn     type   unit  unit  valu valu valu unkn  valu     valu valu
   AUA55 MAINTAIN ALTITUDE 320 FL
   AUA55 DIRECT_TO ING none // plaus: 0.6

<!--  name="libASR::IsaviaFreqUsedExamples2::maintainAltitudeUntilPassing()" --></method>
<method name="libASR::IsaviaFreqUsedExamples2::maintainAltitudeUntilPassingXXft()">
   UNTIL PASSING ES,   is the challenge, Prio 2
    tango  fox  fox oscar x-ray maintain three thousand feet until echo sierra
     csgn csgn csgn  csgn  csgn     type  valu     valu unit  unkn valu   valu
   TFFOX MAINTAIN ALTITUDE 3000 ft
   TFFOX DIRECT_TO ES none // plaus: 0.6

<!--  name="libASR::IsaviaFreqUsedExamples2::maintainAltitudeUntilPassingXXft()" --></method>
<method name="libASR::IsaviaFreqUsedExamples2::maintainAltitudeUntilPassingLatLong()">
   UNTIL PASSING 62-30N,   is the challenge, Prio 1
     klm nine nine maintain flight level three nine zero until passing  six  two three zero north
    csgn csgn csgn     type   unit  unit  valu valu valu  unkn    unkn unkn unkn  unkn unkn  unkn
                                                  ########################################
   KLM99 MAINTAIN ALTITUDE 390 FL

<!--  name="libASR::IsaviaFreqUsedExamples2::maintainAltitudeUntilPassingLatLong()" --></method>
<method name="libASR::IsaviaFreqUsedExamples2::maintainAltitudeUntilPassingCA()">
   UNTIL PASSING CA,   is the challenge, Prio 3
    faxi five  six maintain flight level  one nine zero while   in controlled airspace
    csgn csgn csgn     type   unit  unit valu valu valu  unkn unkn       unkn     unkn
                                                 ##############################
   FXI56 MAINTAIN ALTITUDE 190 FL

<!--  name="libASR::IsaviaFreqUsedExamples2::maintainAltitudeUntilPassingCA()" --></method>
<method name="libASR::IsaviaFreqUsedExamples2::climbToCrossLatLong()">
   CLIMB 380 FL WHEN PASSING 019W,   is the challenge, Prio 3
    pakistan  six   oh  six climb   to cross nineteen west   at   or above flight level three eight zero
        csgn csgn unkn unkn  unkn unkn  unkn     unkn unkn unkn unkn  unkn   unit  unit  valu  valu valu
              ##########################################################
   PIA606 ALTITUDE 380 FL // plaus: 0.6

<!--  name="libASR::IsaviaFreqUsedExamples2::climbToCrossLatLong()" --></method>
<method name="libASR::IsaviaFreqUsedExamples2::directToWaypointAtAltitude()">
   WHEN ALTITUDE 250 FL OR_BELOW,   is the challenge, Prio 3
    world  one  one cross kilo  fox victor   at   or below flight level  two five zero
     csgn csgn csgn  unkn valu valu   valu unkn unkn  unkn   unit  unit valu valu valu
                ######              ################
   WOA11 DIRECT_TO KFV none // plaus: 0.6
   WOA11 ALTITUDE 250 FL // plaus: 0.6

<!--  name="libASR::IsaviaFreqUsedExamples2::directToWaypointAtAltitude()" --></method>
<method name="libASR::IsaviaFreqUsedExamples2::climbBetweenTwoAltitudes()">
   CLIMB BETWEEN 350 410,   is the challenge, Prio 1
     air france  two  two cruise climb between flight level three five zero  and flight level four  one zero
    csgn   csgn csgn csgn   unkn  type    unkn   unit  unit  valu valu valu unkn   unit  unit valu valu valu
                       #######     ########                              #####
   AFR22 CLIMB 350 FL
   AFR22 ALTITUDE 410 FL // plaus: 0.6

<!--  name="libASR::IsaviaFreqUsedExamples2::climbBetweenTwoAltitudes()" --></method>
<method name="libASR::IsaviaFreqUsedExamples2::directToLatLongAtAlt()">
   DIRECT_TO 012W WHEN ALTITUDE 350 FL OR_BELOW   is the challenge, Prio 2
    speed bird  two four five cross  one  two west   at flight level three five zero   or below
     csgn csgn csgn csgn csgn  unkn unkn unkn unkn unkn   unit  unit  valu valu valu qual  qual
                          ############################
   BAW245 ALTITUDE 350 FL OR_BELOW // plaus: 0.6

<!--  name="libASR::IsaviaFreqUsedExamples2::directToLatLongAtAlt()" --></method>
<method name="libASR::IsaviaFreqUsedExamples2::reportMiscellaneousExtraction()">
   REPORT_MISCELLANEOUS instead DIRECT_TO   is the challenge, Prio 3
    dynasty five five confirm established   on  the track between ecco sierra  and alfa kilo india
       csgn csgn csgn    unkn        unkn unkn unkn  unkn    unkn unkn   unkn unkn valu valu  valu
                  ############################################################
   CAL55 DIRECT_TO AKI none // plaus: 0.6

<!--  name="libASR::IsaviaFreqUsedExamples2::reportMiscellaneousExtraction()" --></method>
<method name="libASR::IsaviaFreqUsedExamplesHoldClearance::holdWaypointAtAltAndExpectFurtherClrnce()">
```

```
      HOLDING NASBU is the challenge
       iceair  two bravo cleared   to nasbu descend flight level nine zero   to hold expect further clearance   at  one  two
one five
         csgn csgn  csgn    type type  valu    type   unit  unit valu valu unkn unkn   unkn    unkn      unkn unkn unkn unkn
unkn unkn

######################################################
      ICE2B DIRECT_TO NASBU none
      ICE2B DESCEND 90 FL

   <!--  name="libASR::IsaviaFreqUsedExamplesHoldClearance::holdWaypointAtAltAndExpectFurtherClrnce()" --></method>
   <method name="libASR::IsaviaFreqUsedExamplesHoldClearance::holdTrackAndLeaveHolding()">
      HOLDING TRACK 340 RIGHT and LEAVE HOLDING WHEN DURATION 2 min    are the challenges, Prio 3
       norland zero  one cleared   to golf romeo maintain flight level zero eight zero hold inbound track three four zero right
hand pattern outbound time  two minutes expect approach clearance   at  two three five nine
        csgn csgn csgn    type type valu  valu     type   unit  unit valu  valu valu unkn    unkn  unkn  unkn unkn unkn  unkn
unkn    unkn     unkn unkn unkn    unkn   unkn     unkn      unkn unkn unkn  unkn unkn unkn

#########################################################################################################################
########
      FNA01 DIRECT_TO GR none
      FNA01 MAINTAIN ALTITUDE 80 FL

   <!--  name="libASR::IsaviaFreqUsedExamplesHoldClearance::holdTrackAndLeaveHolding()" --></method>
   <method name="libASR::IsaviaFreqUsedExamplesHoldClearance::holdNoneLeftHandPattern()">
      how to model not fully clear, Prio 4
       scandinavian  one seven double eight cleared   to  the  one five  two radial   of  the kilo  fox victor  vor   at  two
three  dme  fix climb flight level  two  two zero hold left hand pattern expect further clearance   at time  two  two  two  two
             csgn csgn  csgn   csgn  csgn    unkn unkn unkn unkn unkn unkn   unkn unkn unkn valu valu   valu unkn unkn unkn
unkn unkn unkn  type   unit  unit valu valu valu type unkn unkn    unkn   unkn    unkn      unkn unkn unkn unkn unkn unkn unkn
                                        ################################################
##############################
#########################################################################
      SAS1788 CLIMB 220 FL // plaus: 0.8
      SAS1788 HOLDING KFV // plaus: 0.8

   <!--  name="libASR::IsaviaFreqUsedExamplesHoldClearance::holdNoneLeftHandPattern()" --></method>
   <method name="libASR::IsaviaFreqUsedExamplesHoldClearance::holdNoneBetweenTwoDME()">
      how to model not fully clear, Prio 4
       condor  six zero cleared   to  the  two  two zero radial   of  the india november golf  vor   at  one five  dme  fix
maintain flight level three  two zero hold between  one five  and three five  dme expect further clearance   at  one five five
zero
         csgn csgn csgn    unkn unkn unkn unkn unkn unkn   unkn unkn unkn  valu     valu valu unkn unkn unkn unkn unkn unkn
type   unit  unit  valu valu valu type    unkn unkn unkn unkn  unkn unkn unkn   unkn    unkn      unkn unkn unkn unkn unkn unkn
                    ################################################              ##############################
#############################################################################################
      CFG60 MAINTAIN ALTITUDE 320 FL
      CFG60 HOLDING ING

   <!--  name="libASR::IsaviaFreqUsedExamplesHoldClearance::holdNoneBetweenTwoDME()" --></method>
   <method name="libASR::IsaviaFreqUsedExamplesSpeedControl::maintainSpeedUntilFL()">
      UNTIL PASSING 100 FL  is the challenge, Prio 3
        air india seven maintain  two seven zero knots until passing flight level  one zero zero
       csgn  csgn  csgn     type valu  valu valu  unit  unkn    unkn   unit  unit valu valu valu
                                              ###############
      AIC7 MAINTAIN SPEED 270 kt
      AIC7 ALTITUDE 100 FL // plaus: 0.6

   <!--  name="libASR::IsaviaFreqUsedExamplesSpeedControl::maintainSpeedUntilFL()" --></method>
   <method name="libASR::IsaviaFreqUsedExamplesSpeedControl::maintainSpeedOrGreaterUntilFL()">
      UNTIL PASSING 300 FL  is the challenge, Prio 3
       papa papa victor triple seven maintain three hundred knots   or greater until passing flight level three zero zero
       unkn csgn   csgn   csgn  csgn     type  valu    valu  unit qual    qual  unkn    unkn   unit  unit  valu valu valu
      #####                                                          ###############
      PPV777 MAINTAIN SPEED 300 kt OR_GREATER
      PPV777 ALTITUDE 300 FL // plaus: 0.6

   <!--  name="libASR::IsaviaFreqUsedExamplesSpeedControl::maintainSpeedOrGreaterUntilFL()" --></method>
   <method name="libASR::IsaviaFreqUsedExamplesSpeedControl::speedWhenSpeedConversion()">
      WHEN SPEED_CONVERSION  is the challenge, Prio 3
       iceair  two three three speed upon conversion three zero zero knots
         csgn csgn  csgn  csgn  type unkn       unkn  valu valu valu  unit
                                  ################
      ICE233 SPEED 300 kt

   <!--  name="libASR::IsaviaFreqUsedExamplesSpeedControl::speedWhenSpeedConversion()" --></method>
   <method name="libASR::IsaviaFreqUsedExamplesSpeedControl::maintainSpeedUntilLatLong()">
      0.8 MA and UNTIL PASSING 020W  are the challenge, Prio 3
       midnight nine nine maintain mach zero eight zero   or greater until twenty west
           csgn csgn csgn     type unkn valu  valu valu unkn    unkn  unkn   unkn unkn
                              #####               ###############################
      MDT99 MAINTAIN ALTITUDE 80 none

   <!--  name="libASR::IsaviaFreqUsedExamplesSpeedControl::maintainSpeedUntilLatLong()" --></method>
   <method name="libASR::IsaviaFreqUsedExamplesSpeedControl::maintainSpeedUntilWaypoint()">
      0.82 MA and UNTIL PASSING HO  are the challenge, Prio 3
       netjet eight maintain mach eight  two   or less until hotel oscar
         csgn  csgn     unkn unkn  unkn unkn unkn unkn  unkn  unkn  unkn
                  ####################################################
      NJT8 NO_CONCEPT

   <!--  name="libASR::IsaviaFreqUsedExamplesSpeedControl::maintainSpeedUntilWaypoint()" --></method>
   <method name="libASR::IsaviaFreqUsedExamplesTrafficInfo::reportAltitudeInfoMisc()">
      ALTITUDE 6000 ft must be avoided, it is just an information of new transition altitude, Prio 3
```

```
       lucky  air  six  six  low altitude warning check your altitudeimmediately  qnh   is  low nine seven  two   at bildudalur
airport minimum flight altitude   is  six thousand feet
        csgn csgn csgn csgn unkn     unkn    unkn  unkn unkn               unkn type unkn unkn valu  valu valu unkn       unkn
unkn    unkn   unkn     unkn unkn valu     valu unit
                        ####################################################                                    ##########
################################################
      LKE66 INFORMATION QNH 972
      LKE66 ALTITUDE 6000 ft // plaus: 0.6

   <!--  name="libASR::IsaviaFreqUsedExamplesTrafficInfo::reportAltitudeInfoMisc()" --></method>
   <method name="libASR::IsaviaFreqUsedExamplesVectoringInstructions::continuePresentHeadingWhenCondition()">
       WHEN PASSING ATSIX is challenge, Prio 2
       wizz  air  two seven five continue present heading after passing atsix
       csgn csgn csgn  csgn csgn     type    type    type  unkn    unkn  unkn
                                                 ####################
      WZZ275 CONTINUE PRESENT_HEADING

   <!--  name="libASR::IsaviaFreqUsedExamplesVectoringInstructions::continuePresentHeadingWhenCondition()" --></method>
   <method name="libASR::IsaviaFreqUsedExamplesVectoringInstructions::leftHeadingForDelay()">
      INFORMATION TRAFFIC for 'for delay' is the challenge, Prio 3
       greenland  air five alfa turn left heading  two seven zero  for delay
            unkn unkn csgn csgn type qual    type valu  valu valu unkn  unkn
      ###############                                    ############
      GRL5A HEADING 270 LEFT

   <!--  name="libASR::IsaviaFreqUsedExamplesVectoringInstructions::leftHeadingForDelay()" --></method>
   <method name="libASR::IsaviaFreqUsedExamplesVectoringInstructions::turnRightByXDegrees()">
      REPORT HEADING and REPORT_MISCELLANEOUS  is the challenge; Prio 3
       november  one  one turn right  ten degrees report  new heading
           csgn csgn csgn unkn  qual valu    unkn   type unkn    unkn
                        #####          ########       #############
      N11 TURN_BY 10 RIGHT
      N11 REPORT_MISCELLANEOUS

   <!--  name="libASR::IsaviaFreqUsedExamplesVectoringInstructions::turnRightByXDegrees()" --></method>
   <method name="libASR::IsaviaFreqUsedExamplesVectoringInstructions::turnLeftXDegreesInfoTraffic()">
      INFORMATION TRAFFIC for 'due traffic' is the challenge, Prio 3
       kilo  fox bravo turn left  ten degrees  due traffic
       csgn csgn  csgn unkn qual valu    unkn unkn    type
                    #####          #############
      KFB TURN_BY 10 LEFT
      KFB INFORMATION TRAFFIC none

   <!--  name="libASR::IsaviaFreqUsedExamplesVectoringInstructions::turnLeftXDegreesInfoTraffic()" --></method>
   <method name="libASR::IsaviaFreqUsedExamplesVectoringInstructions::headToAndDirectWhenAble()">
      WHEN ABLE  for DIRECT_TO   is the challenge, Prio 3
       nordland eight eight  fly heading  two eight zero when able proceed direct rapax
           csgn  csgn  csgn type    type valu  valu valu unkn unkn    type   type  valu
                                                 ##########
      NWS88 HEADING 280 none
      NWS88 DIRECT_TO RAPAX none

   <!--  name="libASR::IsaviaFreqUsedExamplesVectoringInstructions::headToAndDirectWhenAble()" --></method>
   <method name="libASR::IsaviaFreqUsedExamplesVectoringInstructions::resumeOwnNaviXMilesFromY()">
      WHEN PASSING KFV 45 NM as condition for NAVIGATION_OWN  is the challenge, Piro 3
       golf echo charlie seven eight four five miles from kilo  fox victor resume  own navigation direct gunpa
       csgn csgn    csgn  csgn  csgn unkn unkn  unkn unkn valu valu   valu   type type       type   type  valu
                                 ####################
      GEC78 DIRECT_TO KFV none // plaus: 0.6
      GEC78 NAVIGATION_OWN
      GEC78 DIRECT_TO GUNPA none

   <!--  name="libASR::IsaviaFreqUsedExamplesVectoringInstructions::resumeOwnNaviXMilesFromY()" --></method>
   <method name="libASR::IsaviaFreqUsedExamplesVectoringInstructions::resumeOwnNaviMagneticTrackUntil()">
      new type MAGNETIC_TRACK 115 with condition UNTIL DISTANCE 100 NM  is the challenge, Prio 3
       india charlie echo four seven radar vectoring terminated resume  own navigation direct november bravo magnetic track
one  one five distance  one hundred miles
        csgn    csgn csgn csgn  csgn  unkn      unkn       unkn   type type       type   type     valu  valu     unkn  unkn
unkn unkn unkn     unkn unkn    unkn  unkn
                                 ############################
#########################################################
      ICE47 NAVIGATION_OWN
      ICE47 DIRECT_TO NB none

   <!--  name="libASR::IsaviaFreqUsedExamplesVectoringInstructions::resumeOwnNaviMagneticTrackUntil()" --></method>
   <method name="libASR::IsaviaFreqUsedExamplesVectoringInstructions::resumeOwnNaviMagneticTrackUntil2()">
      new type MAGNETIC_TRACK 015 with condition UNTIL DISTANCE 1236 NM  is the challenge, Prio 3
        pat three resume  own navigation direct mykenes magnetic track zero  one five distance  one  two three  six miles
       csgn  csgn   type type       type   unkn    unkn     unkn  unkn valu valu valu     unkn unkn unkn  unkn unkn  unkn
                                    ##############################              ###################################
      PAT3 NAVIGATION_OWN
      PAT3 HEADING 015 none // plaus: 0.6
```

.

The first and the last row of an example show the name of the corresponding test, i.e. for quality assurance all examples are implemented as unit test. After each code change, each test is automatically executed and in case of a broken test, the changed code needs to be reset.

The following figure shows the current status of test execution. 1044 tests are currently implemented for quality assurance of command extraction. 47 tests are disabled and the 46 failed tests will also be disabled again, after having created this report.

# 7 References

[1] H. Helmke; A. Prasad, T. S. Simiganoschi, J. Harfmann: HAAWAII project: D2.3 One Month of Surveillance and Voice Data from Isavia, version 1.00, 4th March 2021.

[2] H. ARILÍUSSON, T. SIMIGANOSCHI, H. HELMKE, J. HARFMANN: HAAWAII project: D1.1: Operational Concept Document, version 01.00.00, 19. August 2020.

[3] H. ARILÍUSSON, T. SIMIGANOSCHI, H. HELMKE, J. HARFMANN: HAAWAII project: D6.2: Updated Operational Concept Document, version 01.50.00, intermediate version, 16. May 2021.

[4] H. Helmke, M. Kleinert, O. Ohneiser, H. Ehr, S. Shetty, "Machine Learning of Air Traffic Controller Command Extraction Models for Speech Recognition Applications," IEEE/AIAA 39th Digital Avionics Systems Conference (DASC), San Antonio, TX, USA, 2020.

[5] H. Helmke, M. Slotty, M. Poiger, D. F. Herrer, O. Ohneiser et al., "Ontology for transcription of ATC speech commands of SESAR 2020 solution PJ.16-04," in IEEE/AIAA 37th Digital Avionics Systems Conference (DASC). London, United Kingdom, 2018.


HAAWAII


SESAR
JOINT UNDERTAKING


DLR


BRNO
UNIVERSITY
OF TECHNOLOGY


idiap
RESEARCH INSTITUTE


austro
CONTROL


isavia ANS
Air Navigation Services


NATS


CROATIA
CONTROL