Flight call sign identification on a Controller Working Position

Raquel Garcia Lasheras[1], *Adrián Fabio*[1], *Fernando Celorrio*[1], *Juan Albarrán*[2], *Nadal Ceñal*[3], *Carlos Pinto de Oliveira*[4], *Cristina Bárcena Martín*[2], *Julián Chaves*[3], *Mhamed Fillal*[3]

[1]CRIDA A.I.E. Spain, [2]ENAIRE, Spain, [3]INDRA, Spain, [4] EML Speech Technology GmbH, Germany

rglasheras@e-crida.enaire.es, afabio@e-crida.enaire.es, fccamara@e-crida.enaire.es, jaalbarran@enaire.es, ncenal@indra.es, carlos.pintodeoliveira@eml.org, cpbarcena@enaire.es, jchaves@indra.es, mfillal@indra.es

Abstract

In the Air Traffic Management, ATM, environment, Air Traffic Controllers, ATCos, and Flight Crew, FC, communicate via voice to exchange different type of data such as commands, readbacks (confirmation of the command) and information related to the air traffic environment. Speech recognition can be used in these voice exchanges to support ATCos in their work: each time a flight identification or call sign is mentioned by the controller or the pilot, the flight is recognized through Automatic Speech Recognition, ASR, and the call sign is highlighted on the ATCo screen to increase her situational awareness and safety.

This paper presents the work that is being performed within SESAR2020 founded solution PJ.10-W2-96 ASR in call sign recognition via voice by Enaire, Indra, and Crida using ASR models developed jointly by EML Speech Technology GmbH (EML) and Crida. The paper describes the ATCo speech environment and presents the main requirements impacting the design, the implementation performed, and the initial outcomes obtained using real operation communications.

Index Terms: speech recognition, human-computer interaction, situational awareness, air traffic management, air traffic controller, flight call sign, ASR.

1. Introduction

ATCos work with a Controller Working Position, CWP, which displays all the information needed to support them in performing the safe, orderly, and efficient management of flights. On the CWP, flights are presented as radar tracks with an associated label indicating as a minimum the flight identification or call sign, current flight level, current speed, and next point of the route.

While performing their tasks, ATCos must communicate with flight crews to provide them with commands and information. This communication can be performed via voice or via datalink.

Communication between ATCos and FC follows the standard defined in ICAO [1]. This standard states that when communications are initiated by ATCos they must:

- Start by the identification or call sign of the flight being addressed;
- Continue by issuing the command with its qualifiers or information.

Example:

Iberia three four two descend flight level two five zero

Control commands safety-related parts must always be acknowledged by the FC whose answer:

- Starts by the command with its qualifiers;
- Ends with the identification or call sign of flight.

This answer is known as readback and it is vital for ensuring mutual understanding between the FC and the ATCo of the intended plan for the aircraft. ICAO [2] requires "Flight Crew shall read back to the air traffic controller safety-related parts of ATC clearances and instructions which are transmitted by voice"

The answer to the previous command would be:

Descending to flight level two five zero, Iberia three four two

When the FC initiates communication with the ATCos they will start the communication with the call sign and follow it with the necessary information. FC can initiate communications for several reasons:

- FCs always have to call Air Traffic Control when they are about to enter a new Air Traffic Service, ATS, unit or sector; they make a call prior to the boundary between both airspaces. FC communicates with the ATCo to make her aware of its presence, and confirm that voice communication is feasible for emergency use. In this communication FC will typically greet the ATCo and provide some information related to the flight. Example: *Good morning Ryanair nine zero three five flight level three hundred*
- FC usually starts communications any time with ATCos to request modifying vertical / horizontal trajectories and / or speed to fly at the optimum performance of the aircraft.
- Another important reason to initiate a call from the FC is requesting to modify their flight level, route, speed or any other flight condition because of adverse weather like encountering cumulonimbus, severe turbulence, icing etc:

Air Europa six alfa bravo requesting flight level four zero zero due to severe turbulence.

2. ASR in ATM communication environment

The ATM community has investigated ASR mainly using communications from controller utterances [3], [4], [5]. This is due mainly to the fact that the ASR is seen as a means to free the controller from the necessity to manually introduce commands on the CWP, but also because of the characteristics of controllers and pilot communications.

2.1. Basic features in communications initiated by the controller

The voice signal used for speech recognition from controllers' voice utterances is extracted directly from the jack of the controller. This signal has a low degree of noise.

Controllers' language is English or the local language of the ground station [6].

Usually, controllers of an Air Navigation Service Provider will have similar accents when speaking.

The percentage of women/men in air traffic control differs from one country to another. In Spain or France the percentage is around 33% women [7], [8].

As presented in [9], communications from controller to flight crew can be standardized as:

Call id + command + qualifier 1 + qualifier 2.

2.2. Basic features in communications initiated by flight crew

The voice signal used for speech recognition from flight crew voice utterances is extracted from radio communications. The quality of these communications is highly dependent on:

- The distance of the aircraft to the receiving radio station.
- The signal to noise ratio, SNR, can vary from 10 dB to-5 dB [10].
- The quality of the signal transportation from the radio station to the air traffic control facility where the signal is analysed.

Flight crew language is English or the local language of the ground station [5].

FCs have very different accents usually, but not always, related to the flight company country. Countries that are in the routes of international flights have even higher rates of different accents.

Communications from flight crew to controller can, similarly to the controller's ones, be decomposed as: Call id + command + qualifier 1 + qualifier 2.

Or if it is a readback:

Command + qualifier 1 + qualifier 2 + Call id

2.3. Other environment information

Each ATCo has a list of flights that either are in his sector, are about to enter in it or are of interest (e.g. because they fly near the sector border). The information is provided by a flight data processor, FDP, that ensures that the list of flights is updated with new incorporations and cancellations once the flight is no longer of interest.

The level of automation is having continuous improvements and enhancements introducing new functions to assist the ATCo for better situational awareness and a reduction in workload supporting them to focus attention when and where needed. Within these new functions it is the ASR Project that requires a new HMI presentation; The new methods of interaction have to be compatible with the other systems and subsystems within the CWP to benefit controller's duties.

3. Call sign illumination

In current operations, a common feature in the communications procedures is the presence of the flight identification or call sign.

Call sign recognition and illumination is considered as a quick win by Enaire that can be implemented in any CWP as they are equipped with radar surveillance service that can display the radar track and call sign identification regardless of the unit where they are installed: En-route / Terminal maneuvering area, TMA, or in Tower, TWR, units.

The expected benefits are:

- Increase of ATCos' situational awareness by quickly identifying new flights entering the sector or flight crews requesting actions from ATCos.
- Provide a safety check to the ATCo by illuminating the call sign coming from an ATCo utterance ensuring they are addressing the proper aircraft.
- Familiarisation of ATCos with speech recognition technology in the ATM environment before its possible incorporation in functions with higher automation.

3.1. Basic ASR engine requirements for call sign illumination

Due to the ATM application in which voice recognition is going to be used, there are three outstanding requirements:

- The Voice Recognition System, VRS, shall be able to function without connection to sources external to the Area Control Centre, ACC.
- The call sign illumination must be produced as soon as possible once the communication has started.
- The ASR engine shall be able to process the utterance in English and local language, when local languages are allowed.

The first requirement limits the available ASR engines, as it must be autonomous. Enaire considers flight management as a strategic field, and therefore, an ACC must be able to provide its service even if it is isolated. This is a requirement set by Enaire that may not be shared by other ANSP.

The second requirement implies that the ASR engine must be able to perform in streaming and provide partial transcriptions. As in the most useful use cases the call sign is at the beginning of the phrase, the time to the initiation of these partial transcriptions is also critical. In project PJ.10-W2-96 ASR [11], the requirement has been established in one second after the ATCo has said the call sign. This value needs to be validated.

3.2. VRS requirements for call sign illumination

The fact that a mistake in call sign illumination can mislead the ATCo which in turn may provoke an accident puts in place a new set of requirements on the VRS:

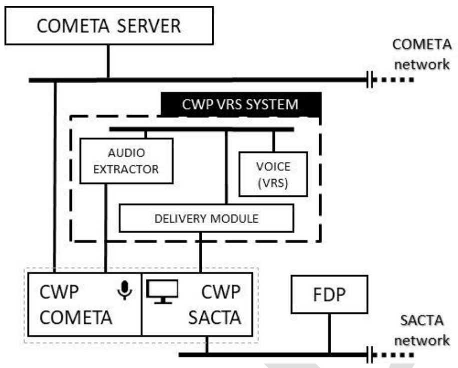
- It is preferable not to have a call sign illumination rather than a wrong call sign illumination.
- The VRS will use the sector flight list from the CWP to improve its performance.

4. Solution Architecture

To investigate the benefits of call sign illumination, a VRS has been integrated in an Enaire's CWP. The VRS is Voice, a recognition system developed by Crida using EML's ASR engines. The ASR models contained have been developed jointly by EML and Crida [12].

Enaire's ATM system is SACTA (Air Traffic Control Automatic System) developed by Indra. The communication system that processes audio signals has recently been upgraded to COMETA (Integrated Voice IP Communication into SACTA). Figure 1 presents the architecture used. The audio is extracted by the audio extractor and sent to Voice for speech and event recognition. The delivery module then sends the event (call sign highlight) information to the SACTA CWP. The SACTA CWP also sends the environment information to Voice via the delivery module.

Figure 1: *System Architecture*



The VRS module uses environment information to improve recognition rate and allows the system to perform a safety check on the correct identification of the flight.

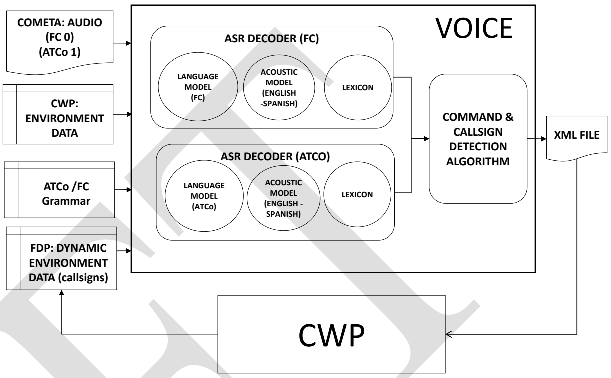
COMETA processes the audio signal following the aeronautical standard [13]. The raw audio is extracted and provided to the VRS. COMETA distinguishes between controller and FC communications. The signal is tagged with a flag indicating the source, FC (0) or ATCo (1), Figure 2. The ATCo can be in charge of one or several frequencies for radio reception (RX) and transmission (TX) depending on the sector configuration. E.g one planner controller may listen not only to the frequency of her sector, but also to the frequency of the neighbouring TMA sector to increase the situational awareness of departing flights. The system is linked to the frequency that the controller transmits (TX).

The list of possible flights of interest can be provided to the VRS from two different sources.

- The FDP has the list of flights that are of interest for the ACC (composed by several sectors). The FDP ensures that the list of flights in each CWP is updated with new incorporations or cancellations once the flight is no longer of interest.
- The CWP has the list of flights that are of interest for the sector. This list is smaller than the previous one, but some flights may not be covered, for example, last minute flights deviated due to weather.

After performing a cost/benefit analysis regarding the size of each file, the system implications, and the number of flights that may be impacted, Enaire decided that the CWP would be the one providing the list. This list is provided to the dynamic lexicon to enhance the call sign detection and to the detection algorithm. It is updated dynamically.

Figure 2: *VRS Architecture*



4.1. Speech models

ASR is provided to "Voice" application with "EML Transcription Server". The recognition engines are set for real-time streaming transcriptions, using latest state-of-the-art technologies in ASR such as bidirectional long-short term memory (BILSTM) neural networks [14] for acoustic modelling, voice activity detection (VAD) [15] and dynamic lexicon update feature. Due to the different characteristics of the communications two different recognition models have been developed. Both use the same multilingual (English and Spanish) acoustic model trained with 1000 hours of recordings (out of ATM domain) and then adapted with approximately 100 hours of ATM domain recordings. One class-based language model for controller communications (ATCo), and one class-based language model for flight crew communications (FC). The ATCo language model is a mature model that has been trained with operational transcriptions (approximately 90k) gathered along several years of collaboration between Crida and EML. The FC model is a new model that has been adapted with transcriptions from 12 h of operational recordings. "EML Transcription Server" dynamic lexicon update feature enables language model class entries to be defined in runtime without having to restart the recognition engines, allowing "Voice" application to update, for example, the list of call signs for a given sector and time.

4.2. Call sign detection algorithm

The call sign detection algorithm analyses the text extracted by the ASR decoder and classifies the words according to the most probable value. Once a sequence is classified as probable call sign, it is compared against the list of possible call signs. If there is a match, a file with the information is created and sent to the Human Machine Interface HMI that highlights the radar track during a configurable time, currently set as 4 seconds. The HMI is able to highlight up to 5 call signs simultaneously, as more than one communication can be performed during these 4 seconds.

The algorithm includes the call sign rules defined by ICAO [1] and is able to detect a flight indicative independently of the method or language used by the controller (pilot) to address it. These methods include:

- The radio call sign e.g. Beeline/ Cactus
- The company name e.g. Brussels Airlines/ US Airways
- ICAO designator using aeronautical alphabet. E.g. Bravo Echo Lima (BEL)/Alpha Whiskey Eco (AWE).
- All the possible modes to pronounce a number. E.g. one zero zero, ten zero, one hundred.

ATCos may pronounce more than one call sign in one utterance, e.g. because they give instructions to different aircraft or because they are informing a flight about a traffic that may influence them..

ATCos and FCs are allowed to refer only to partial call signs once the first communication is established and there is no possibility of confusion.

In the current implementation only complete call signs are recognized. The list of call signs is used to ensure only possible call signs are detected.

5. Initial outcome

Enaire, Indra and Crida performed in November 2020 a dry-run where ATCos analysed the partial implementation of the speech recognition with special emphasis on call sign recognition. The dry-run was a real time simulation simulating two sectors at Madrid ACC (Spain). Due to the Covid-19 restrictions, one operator acted as ATCo issuing the instructions and two pseudo-pilots managed the aircraft. The session was transmitted online to several ATCos, operational and technical staff. The online attendants provided comments and asked to perform different actions during the exercise execution. These were transmitted to the operator and pseudo-pilots.

The dry-run was followed by a statistical analysis of real operational recordings. A gold standard was created with 449 operational recordings from ATCo and FC utterances. These recordings were extracted from 3 different Spanish sectors. Several of the recordings were disregarded in the final analysis because they were just noise or did not contain a call sign.

Call sign recognition rates obtained from the analyses appear on Table 1. The row controller/pilot's indicates the number of call signs contained and detected in controllers'/pilot utterances. The row first call/request is a subset of the pilot utterances where the FC initiates the communication i.e. the first time a flight enters a sector or a request from the pilot not expected by the controller. As previously indicated, these communications are of special interest as they imply a change to the controller's attention focus. No false recognition was performed.

Table 1: *Call sign recognition*

	Call signs	**Detected Call signs**	**Detection rate**
Controller	143	125	87.4%
Pilot	158	77	48.7%
First call/request	65	38	58.5%

Table 2 indicates the utterances where only partial call signs have been pronounced and the correct transcription performed by the ASR. As previously indicated, the algorithm has not implemented this detection so it was done manually.

Table 2: *Short call sign recognition*

	Short Call signs	**Transcribed Call signs**	**Detection rate**
Controller	3	2	66.7%
Pilot	12	7	58.3%

Other technical feedback obtained:

- The call sign illumination took 3 seconds since the initialization of the utterance until the illumination. If the call sign was at the end of the phrase the reaction time was lower, 1.4 seconds.
- The call sign recognition rate was better for ATCos than for FC.

The subjective feedback from the ATCos was that:

- The call sign illumination was useful. They indicated the reaction time should be lower.
- They were satisfied with the call sign recognition pattern i.e. being able to address the flight using several different approaches.
- They would like to be able to independently switch on and off the controller and pilot recognition modules.

6. Conclusions and next steps

The call sign illumination has been well received by the different stakeholders involved in the exercise. The controller model has higher recognition rates than the pilot model. This was already expected due to the different maturity of both voice models and the inherent aspects of pilot utterances with worse signal/noise relation and higher number of different accents when compared with the controller model. To foster controller situational awareness, it is key that first call recognition rates are high and model reaction times are low.

The utterances with short-call sign are usually readbacks from pilots which initially do not improve controller situational awareness although they may provide a safety check to ensure the proper flight is giving the readback. Future implementations where controller utterances are used to automatically implement the command on the CWP will require the recognition of partial call signs.

Taking these considerations into account the project will perform the following steps along 2021:

- Improve the ASR pilot decoder model with new training hours.
- Improve the VOICE prototype to lower the initial reaction time.
- Perform a cost/benefit analysis on short-call sign recognition algorithm, taking into account the improved recognition rate against the false positive rate and the final use of the algorithm.

7. Acknowledgements

This project has received funding from the SESAR Joint Undertaking under the European Union's Horizon 2020 research and innovation programme under grant agreement No 874464.

8. References

[1] Procedures for Air Navigation Services (PANS) - Air Traffic Management Doc 4444. ICAO. 16th Edition, 2016.

[2] Annex 11 – Air Traffic Services. ICAO. 15th Edition, 2018. Para 3.7.3.1

[3] H. Helmke, J. Rataj, T. Mühlhausen, O. Ohneiser, H. Ehr,M. Kleinert, Y. Oualil, M. Schulder, and D. Klakow, “Assistant-based speech recognition for ATM applications,” inProc. of 11thUSA/Europe Air Traffic Management Research and DevelopmentSeminar (ATM 2015), Jun. 2015.

[4] V Nhan Nguyen, Holone, H. “Possibilities, challenges and the state of the art of automatic speech recognition in air traffic control”, Int. J. Comput. Inf. Eng.2015,9, 1940–1949.

[5] T. Pellegrini, J. Farinas, E. Delpech, and F. Lancelot, “The Airbus Air Traffic Control Speech Recognition 2018 Challenge: Towards ATC Automatic Transcription and Call Sign Detection,” in Interspeech 2019, 2019, pp. 2993–2997.

[6] Annex 10. Aeronautical Telecommunications. ICAO. Edition 2001. 2001.Para 5.21.2

[7] Estudio sobre el empleo real de las mujeres en el sector aeronáutico. Fly News June 2017

[8] Delpech, E.; Laignelet, M.; Pimm, C.; Raynal, C.; Trzos, M.; Arnold, A.; Pronto, D. “A Real-life, French-accented Corpus of Air Traffic Control Communications.” In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), Miyazaki, Japan, 7–12 May 2018.18.

[9] 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

[10] Segura, J.C.; Ehrette, T.; Potamianos, A.; Fohr, D.; Illina, I.; Breton, P. A.; Clot, V.; Gemello, R.; Matassoni,M.; Maragos, P, “The HIWIRE Database, A Noisy and Non-Native English Speech Corpus for Cockpit Communication.” 2007

[11] SESAR PJ.10-W2-96 ASR Initial Technical Specification. V00.01.00 . 2020

[12] J.M. Cordero, M. Dorado, and J.M. de Pablo, “Automated speech recognition in ATC environment,” Proceedings of the 2nd International Conference on Application and Theory of Automation in Command and Control Systems (ATACCS‘12), IRIT Press, Toulouse, France, 2012,pp. 46–53.

[13] EUROCAE ED-137 Interoperability Standards for VOIP ATM components. Ed 2012

[14] V. Fischer, O. Ghahabi, and S. Kunzmann, “Recent improvements to neural network based acoustic modeling in the EML real-time transcription platform.” in Proc. ESSV, 2018, pp. 38-45.

[15] O. Ghahabi, W. Zhou and V. Fischer, “A robust voice activity detection for real-time automatic speech recognition,” in Proc. ESSV, 2018, pp. 85-91.