Automated Speech Recognition in Controller Communications applied to Workload Measurement

José Manuel Cordero, Natalia Rodríguez , José Miguel de Pablo
CRIDA, ATM R&D Reference Center
Madrid, Spain
{jmcordero, jmdepablo, nrodriguez}@crida.es

Manuel Dorado
AENA, Simulation Division
Madrid, Spain
mdorado@aena.es

***Abstract*—This paper is focused on the description of an Automated Speech Recognition system in Air Traffic Control environment and its application in the automated measurement of controller workload for demand and capacity balance purposes. For years there have been several attempts to properly apply Automated Speech Recognition technologies in the ATC domain, but the results obtained have been quite frequently not positive enough. AENA, the Spanish Air Navigation Service Provider, started a project with the objective to develop a prototype able to recognize voice communications from the controller to the pilot and integrate it on a Real-Time Simulator, closing the loop by generating the answers from the pseudopilots. After several iterations and testing many different technological and commercial Speech Recognition solutions (most of which proved to be inefficient for ATC communications), an ad-hoc system able to successfully recognize the content of the controller communication has been developed, being able to automatically transcribe it and determine the associated controller event. This system is currently obtaining high detection rates, which have allowed its integration with real ATC communications for automated controller voice events detection, that are later used for automated controller workload measurement. This innovative application of Speech Recognition systems to the ATC operation voice communications will be described in this paper.**

Keywords- air traffic control; voice; speech recognition; workload; human factors; controller events; capacity management

I. INTRODUCTION

In the continuous search of applying potentially useful new technologies to the Air Traffic Control (ATC) domain, one of the most complex ones historically has been the Speech Recognition (SR), that even though numerous attempts has shown not very promising results [1]. While this may look discouraging from an operational point of view, in the Simulation and Training domains some positive and useful results have been obtained, especially those related with the use of contextual knowledge [2, 3]. Although there are limitations because of the strong need of the system to have contextual information available (which is not always possible) and the need to keep the phraseology reasonably similar to predefined standard models (not very realistic out of the simulation environment), these results proved the existence of an appliance field for Speech Recognition technologies in ATC, and a sufficient maturity level in them for simulation applications.

The prototype presented here, named VOICE, tries to go one step further by applying in an innovative way these technologies to the operational domain, particularly from an Air Traffic Controller (ATCO) workload measurement perspective on the scope of capacity management in DCB (Demand Capacity Balance) tools, with rather positive results both in detection rates, usability and robustness that are presented in this paper.

Workload, at a microscopic level, is widely recognized as one of the key factors affecting controller's performance and, thereby, system capacity. However, as workload measurement cannot be performed through direct means, it has to be inferred or estimated, for which many methods have been developed during the years, covering from a very early queuing analysis [4] and controller's physiological variables monitoring [5] to more recent indicators of traffic complexity and, especially, its relationship to controller's workload [6-9], through diverse models of controller's behavior, among which some of the most extended are those based on the multiple cognitive resources model and theory of Wickens, with focus in controller events [10].

In order to be able to determine in a complete way the controller events occurring during operation, a wide set of information from different sources has to be collected, correlated and interpreted, achieving a better and richer detection as more sources are considered so that every potential controller event could be detected. This includes all the information related to the interaction of the controller with the ATC System interface, from simple use to voice communications (either controller-pilot or controller-controller). Usually controller's voice communications are not accessible for this monitoring/measurement purposes, which together with the low performance results showed by the speech recognition systems applied to ATC communications until the moment makes not convenient the inclusion of this source of information in the workload model. Thus, a proper and valuable integration of the controller's voice communication in the workload model would necessarily require of two key aspects: access to the voice communications, and an efficient ATC speech recognition system with high detection rates.

Extensive studies have been done in the past investigating the relationship between ATC communication events and associated taskload/workload with focus in the number and

duration of controller-pilot communications [11-13], and more recently with a network dynamics-based approach [14], all of them leaving apart the content of the transmissions. Part of the difficulty for the inclusion of even this elementary communication data in systematic workload calculation is the considerable amount of time and labor required to merely determine the number and duration of communications manually, largely increased if semantic analysis was required. Thereby, for the consideration of controller communications in workload measurement, in terms of usability, it must be required that any processing done with them, either in number and duration or semantic analysis, is automated.

Trying to extract and process all the possible information from the ATC System for workload measurement through controller events detection, AENA (Spanish Air Navigation Service Provider) designed and developed in 2008 a prototype with speech recognition capabilities integrated in a Real-Time Simulation Platform, also able to extract from it every interaction of the controller with the ATC System HMI (human-machine interface) during any given exercise [15].

The voice communications obtained during a set of simulation exercises were analyzed both grammatically and semantically, and later correlated with information extracted from the platform, obtaining a resulting set of controller events usable for workload calculation. The controllers involved in the simulation used natural speech as in normal ATC operation.

Conclusions from that study pointed out that frequently controller events could be detected in a redundant way either from interaction with the ATC platform (explicit or implicit action patterns) or by voice communications, and emphasized that several actions and events were only obtainable by voice analysis (i.e., inter-sector coordination). When the controller event was detectable through both sources of information a cross check was done between them, showing than for a set of events the detection through speech recognition provides more accurate and reliable results.

Voice communications recognition in ATC environment is thereby identified as a fundamental source of controller events identification. However, the use of widely extended Automated Speech Recognition (ASR) tools for the automated on-line analysis of controller communications didn't provide acceptable results in terms of word detection rate (detailed later), showing great dependence of the contextual information provided by the ATC System and consequently very poor performance when this contextual information was unavailable. The very specific ATC vocabulary and syntax, as well as the variety of accents, speakers and communication channels with different audio characteristics, made difficult to obtain acceptable detection rates with commercial Speech Recognition systems, even after extensive training and adaptation to the ATC environment. Wide state-of-the-art analysis was performed and several Commercial-off-the-shelf (COTS) applications were used, with results varying from 0% to 10% of word detection rate (WDR) when no contextual information was provided. Semantic analysis and search of keywords made possible to increase the detection rate through voice communication in terms of controller events (not words) up to 20% average. Clearly, the application to operational usage required far better detection rates.

Consequently, for systematic and useful communication analysis it became fundamental to count on a reliable ATC Speech Recognition system, able to provide automated transcription of the controller communications as well as their duration with high word detection rates and, specifically, high controller event detection rates so that a high percentage of ATC voice communications could be correctly analyzed later.

At that moment, such a system didn't exist in ATC environment, and all the attempts made to obtain it from a direct adaptation of the existing ASR COTS had been (and still seem to be, as a part of the technological surveillance in this area carried inside the VOICE project) unsuccessful, mainly because of the particularities of ATC communications, that will be later described in this paper.

Establishing the objective of evolving and enhancing the initial VOICE prototype described to a fully ATC operation-usable one with the characteristics previously explained (high detection rates, automation, context information-independency), this paper addresses (i) the solution developed in AENA that resulted in an innovative ATC Speech Recognition system able to reliably detect controller events from controller communications, (ii) its automated application in operational environment for workload calculation, and (iii) the presentation of the results obtained.

II. PROTOTYPE DESCRIPTION

A. ATC Voice Communications Characteristics

In the ATC domain, before the upcoming of digital communications between controllers and pilots (and controller themselves), voice communication via radio was the primary means for traffic control. In fact, nowadays it is still the only communication method between controllers and pilots in most control centers.

Communication activity in this paper is defined as the act in which the ATC controller sends a transmission either to an aircraft in its sector or to another controller in a different sector, considering also the contents of the transmission. Empty transmissions are not considered as a communication activity in this model, as they generate no controller event or workload.

At this point a distinction must be made between voice and speech recognition: voice recognition is commonly used in this field as related to command and control of the system through voice: pronouncing commands in a determined way results in the understanding and execution of those commands by the system; however, speech recognition is related to the natural language, with no restriction to what the controller can say, or to how he has to speak. So, ATC controller voice communications are considered in this paper in terms of Automated Speech Recognition and natural language, not including command and control of the system (thereby, not affecting operation or workload, as forcing the controller to vary its usual behavior or speech would influence the workload because of the mental effort required to do it).

While a large set of different communication channels are included in the Controller Working Position (CWP), the ones affecting controller workload are those which contain input and output controller voice communications, which are three: controller headset output channel (PJ, Panel Jack), radio channel (RD) being an input line through which the controller receives the aircraft communications in the sector radio frequency, and coordination telephone line (TL), a bidirectional channel specifically used to coordinate control actions between adjacent sectors/controllers. The first one, a purely output channel, contains the highest percentage of communications originated by the controller and is emitted through the headset microphone, either to pilots or rarely to other controllers; whereas the coordination line is occasionally used and always to communicate between controllers. There is, thus, one source of controller-pilot communications (PJ) and two sources of controller-controller communications (PJ, TL).

The radio channel has specific characteristics mainly due to its noisy nature, which significantly reduces the signal-noise ratio (SNR) for this communications respect to the SNR observed in the PJ and TL channels. This, obviously, has a negative impact in the detection rates observed in this channel. As the essential controller event information is not contained in this channel, it will be used only to determine the duration and number of communications that the controller receives through this channel, not for controller events detection.

A key aspect of the ATC communications analyzed in this case of study, as they correspond to Spanish airspace in operational environment, is that they could be either in Spanish, English, or even occasionally a Spanish-English mix in the same sentence. This particular feature is caused for being both Spanish and English ICAO (International Civil Aviation Organization) languages, so that their use is allowed in Spanish airspace. An additional feature of the problem, which adds a remarkable complexity to the speech recognition module, is the fact that English speakers in these communications are non native. In practice, this implies that the speech recognizer must be able to detect communications in any of the two languages without having previous notice of which one of them is actually being spoken.

On the other hand, even though ATC communications should be theoretically according to regulated phraseology, empirical observation of real controller communications shows that strict standard phraseology is not consistently used (which has also been stated in [17]), one of the main reasons why detection models based on standard ATC regulated phraseology have provided low detection rates out of the simulation environment. Thereby, the phraseology needs to be enhanced to include real expressions and forms used in real operation; while the speech recognition models design, calibration and training get complicated by this fact, once correctly done the detection rates in real operation communications show great improvement respect to the basic standard phraseology. As a result, controller event detection in this prototype fits to standard and extended phraseology, being both defined as a unique, greater set, able to be enhanced by flexible addition of new words and expressions when required.

B. ATC Speech Recognition System Architecture

The ATC Speech Recognition prototype developed works according to the block diagram shown in Figure 1.

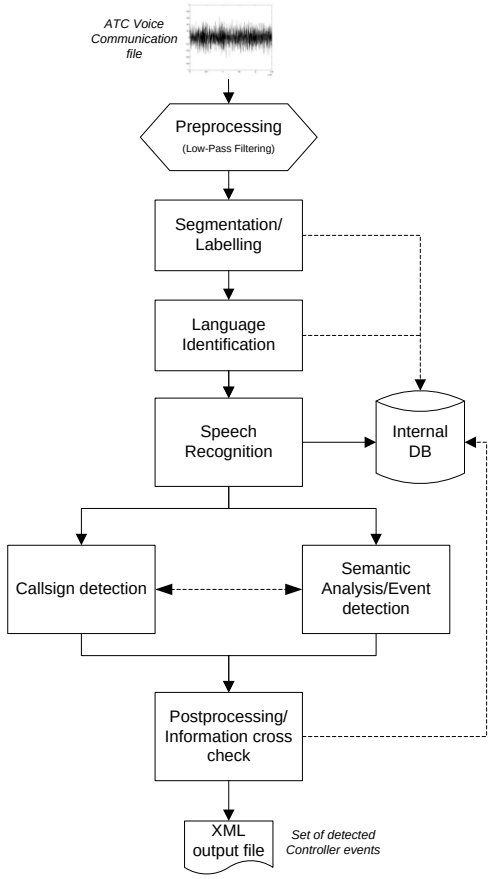

Figure 1. System block diagram.

The signal is filtered for noise reduction, and then segmented by removing silences between individual ATC voice communications, which are extracted separately so that small audio files corresponding to these individual communications are obtained, stored in an internal database and labeled by a reference index. In this stage several system parameters can be adjusted for better segmentation, such as silence, pause o mumbling time.

Determination of the language used in each communication (English or Spanish) is done next to the segmented audio files, based on Bayesian classifier techniques adapted to this specific communications [19]. The observed average language detection rate is 99.2%, being slightly better for Spanish communications (99.6%) than for English ones (97.4%).

The Speech Recognition module then provides a suggested transcription, based in a cutting-edge Speech Recognizer from European Media Laboratory GmbH (EML), enhanced and optimized for ATC communications in a collaborative way with AENA, as well as extensively trained with around 500 transcribed hours of real operational ATC communications. The Speech Recognition is based on a multi-mode Hidden Markov Model (HMM), specifically trained according to the phraseology, vocabulary, syntax and structure of ATC voice

communications. As a key feature, this module makes no use of contextual information.

This speech recognition module is, in fact, subdivided into two models which perform effective recognition:

Language Model: contains the phraseology to be used (in Spanish and English), with usual syntax and grammar structure (i.e., call sign composition rules). Logical relationships, probabilistic modeling and language restrictions are included here, allowing reusability of the model in different ATC centers or individuals, making the system almost speaker independent.

Acoustic model: takes into account physical characteristic of the speech as a sound, characterizing particularities of the speaker accent, speed, etc…, as well as models the physical channel so that changes between channels in different centers are compensated and do not affect the speech recognition module.

The speech recognition module outputs a transcription of the audio segment, and sends it to two different submodules that work in a coordinated way, but applying completely different detection algorithms: the call sign detection and the event detection submodules.

The call sign detection submodule looks into the transcription for a determined segment and starts searching for candidate call sign structures (which can be quite diverse). As an example, Figure 2 illustrates the case of a real ATC communication and the candidate call signs that might be considered; the difficulty of determining the call sign segment and the rest of the communication segment (that will determine the type of controller event) is remarkable as the number of digits in the cal sign it is unknown.

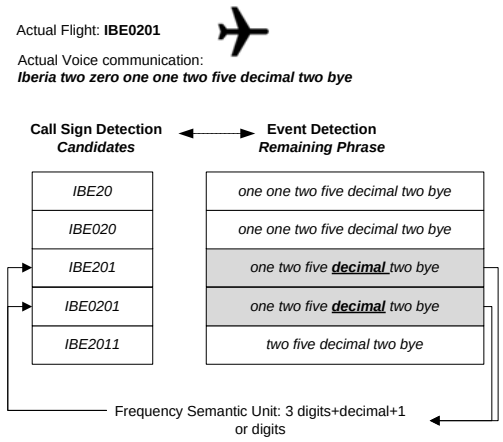

Figure 1. Example of detection.

Consequently, the two detection submodules must work in a coordinated way: they both receive the same transcription and search for different semantic units (call signs, in one case, and predefined structures such as levels, speeds, frequencies, waypoints, greetings, farewells, etc… in the other) looking for keywords and composing candidate sentence structures according to the previously defined grammar rules and phraseology. In the example before, "decimal" is a keyword that unequivocally identifies a frequency, whose defined structure is three digits+*decimal* (optional token, always implies a frequency when detected)+one/two digits; so, once the decimal keyword is identified the system search for a frequency semantic unit structure around the keyword, and the three digits before the "decimal" are assigned a very high probability index to be part of the frequency and, thereby, the call sign is either IBE201 or IBE0201 (which doesn't make a difference to the system). This example illustrates non-standard phraseology detection problems, and how interoperability between both submodules is essential to maximize results.

It is important to remark here that the system has been isolated from the ATC system data and flight plan information on purpose, with the objective of having the best possible ATC speech recognition system without the use of contextual information. Working with access to the system and flight plan data would have allowed knowing the candidate call signs in a sector in advance, thereby increasing detection rate in an easy way just searching for the known call signs in the transcription input. However, the system was designed and developed taking into account a possible use in places not connected to the ATC system, so this restriction was made. Connection to the existing ATC platform is possible for enhanced results.

Finally, after a call sign and a controller event (if existing in the phrase) have been found, they are included in an output file that enters the post-processing module. Here, information is checked in order to fill the gaps that might have not been detected, especially call signs, correlating them with other detected ones, trying to find relationships according to established typical behaviors of a flight inside a sector. Additional filtering is also performed, checking if a same event for a same call sign in a determined short time is in fact a new event or a repetition (what would impact on effective controller workload calculation). After this a final output file is obtained for a sector in a configurable period of time (typically one hour), containing a set of controller events detected from controller voice communications, that can be used for workload measurement or other further analysis.

C. System training

The information used to train and test the Speech Recognition system has been obtained from operational blind recordings (meaning that no date, time or associated sector name are known for them) corresponding to real controller communications in normal activity. No simulation data has been included in the training and testing of the system.

The audio data needed to be manually transcribed in order to reliably feed the language model of the ATC Speech Recognition system. This process was extremely costly, as typically a raw (i.e., including silences) ATC voice communication recording of one hour needs a transcription effort of 8-10 man-hours. If it is considered that typically in an hour of ATC voice communication, after removing silences, there is only an average of 10-15 minutes of controllers voice activity, it takes approximately 40 man-hours in average to get a net (i.e., without silences) hour of ATC voice communications transcribed. This process had to be done in a completely manual way because of the low efficiency of available speech recognition tools available.

So, in order to increase efficiency and reduce operator effort a transcription module was developed, which takes advantage of the existing Speech Recognition engine, providing automated segmentation and labeling, as well as suggesting a transcription for each individual communication, leaving to the operator the decision of correcting or validating the suggested transcription. The necessary effort to transcribe one net hour has been reduced, progressively with the improvement of the Speech Recognition engine and thereby of the suggested transcription, from the initial 40 man-hours up to 7 average man-hours, with this semi-automation on transcription. A secondary immediate application of this transcription supporting module is in the case of ATM incidents that need transcription, and for which the developed tool provides noticeable effort reduction.

The ATC Speech Recognition system has been trained with an increasing amount of transcribed hours in an iterative process, showing that the controller event detection rate, as well as the word detection rate, increases following a logarithmic behavior. The current system contains a trained core of 100 transcribed net hours including en-route and approach, without silences, corresponding to approximately 500 hours of raw controller communications. A key aspect in the later detection of pauses, hesitations and elongations (common in natural speech) is managed here by a correct labeling of them during the training phase.

D. Automation Architecture

The described ATC Speech Recognition prototype is designed to be included into a wider architecture that performs automated analysis and transcription of ATC voice communications. The input recordings are provided by a digital voice recording system through a VoIP (Voice over IP) network in ADPCM 32 coding (with a suitable audio quality in terms of Mean Opinion Score, marking 4.05 in a 1-5 scale [28, 29]), which can be integrated with any external system through a C++ API (Application Programming Interface). Figure 3 illustrates the overall automation architecture.

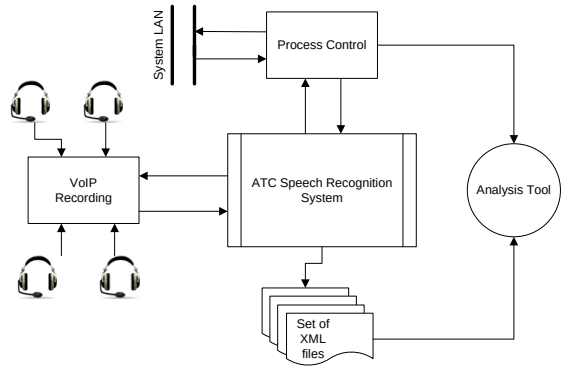

Figure 2. Automated ATC Speech Recognition Architecture.

Integration into an operative environment brings additional difficulties, as it is not known in advance which sector is assigned to which CWP, and for how long. The need of obtaining this information implies integration and analysis of internal ATC system data, through connection to the internal network of the ATC system. Once determined the objective CWP, the specific audio recordings needed (specifying required channels and time period) are requested to the digital voice recording system, and then processed. A set of files (typically one per sector and hour) is obtained, containing all the detected communications, including associated controller events (codified) detailing date, time, sector, communication duration, and associated transcription. An example of the contents of these output files is shown in Figure 4.

```
- <event>
    <name>Av</name>
    <callsign>JKK5690</callsign>
    <UCE>PAU</UCE>
    <date>12/09/2011</date>
    <time>17:02:50</time>
    <duration>4,96</duration>
    <coord>No</coord>
    <message> spanair five six naina zero continue climbing flight level
      two three zero </message>
  </event>
- <event>
    <name>Csv</name>
    <callsign>JKK4233</callsign>
    <UCE>PAU</UCE>
    <date>12/09/2011</date>
    <time>17:03:20</time>
    <duration>5,35</duration>
    <coord>No</coord>
    <message> spanair four two tri tri contact madrid one tri four tri
      five bye </message>
  </event>
```

Figure 3. Example of controller events in XML output file.

This set of files can be processed later with any analytic tool desired, to obtain whatever parameters wished, either by sector, day, time or event flight behavior either for a sector or for all the ACC airspace. However, the overall system has been designed for the use of the detected controller events in workload calculation. Thus, these files are typically used as an input of a Workload Measurement based on events, thereby providing a wide range of operational workload measures in a fully automated way that may allow DCB (Demand Capacity Balance) tools for enhanced capacity management.

The ATC communication processing time unit is considered to be 24 hours (configurable), after which no processing queue is expected to remain in the system. Thus, the processing capacity of the automated speech recognition system is calculated considering the average number of CWP open-hours during a day, and keeping a latent remaining capacity of 15% to deal with peak days.

III. RESULTS

The basis of the described system is, obviously, the capacity to provide reliable transcriptions of real ATC communications; otherwise, the whole system would provide values that would be of no use for further analysis, since they could not be considered as reliable sources of information.

As it was stated in the introduction, up to the best of the authors' knowledge there haven't been found any Speech Recognition systems, providing acceptable detection rates when tested with real (non-simulation) ATC communications, especially with no contextual information provided (flight plans, call signs, etc…). Focus, consequently, has been put on developing and training a new system able to recognize ATC voice communications with a high detection rate, on a double perspective: word detection rate, firstly (for automated transcription, measuring the speech recognition functionality),

and controller event detection rate, secondly (essential for controller workload/sector capacity automated analysis).

For the results provided below, the system has been tested with 3 sets of 10 hours of en-route communications which haven't been used in the system training and 3 sets of 10 approach hours in the same circumstances, all of them from real operation. Four indicators have been defined:

Word Detection Rate (WDR): Measurement of the ability of the ATC speech recognizer, not taking into account whether the recognized words were considered keywords or not. It is defined as the relation between the number of correctly detected words and the total number of words pronounced.

Event Detection Rate (EDR): With the transcription obtained, event detection is later done; however, the False Positives (FP) need to be taken into account, understood as the detection of a controller event when really there was none, as they imply unreal controller workload. In this context, EDR is defined as the number of correctly detected events divided by the sum of total existing events and the number of false positives. An event is considered to be correctly detected only when both the associated flight call sign and the event categorization are correct.

False Positive Rate (FPR): Used as an index of the reliability of the system, it is reflected as a standalone indicator to try to provide good values of EDR without taking too many risks in event detection. So, the FPR is defined as the relation between the total amount of False Positives obtained and the total number of real events (not counting the False Positives).

Event Detection Rate without call sign ($EDR_{no\_callsign}$): In this case, the index is similar to the EDR but, exceptionally, an event is considered to be correct simply by obtaining correct event categorization, not taking into account the call sign detection. This index is mainly used to measure the efficiency of the event detection algorithms, as the call sign detection gets highly affected by poor transcription results, while event detection algorithm (excluding the call sign segment) makes use of several techniques to enrich the speech recognition results.

These four indicators can be used in a single analysis unit (one hour, or interval, in one sector), or aggregated. The results presented here take into account global values of detection across all the events and communications in the 60 hours analyzed for testing. Table I below details the overall results obtained:

TABLE I. DETECTION RATES

	WDR	EDR	FPR	$EDR_{no\_callsign}$
En-route	67.3%	74.6%	5.8%	95.9%
Approach	69.8%	72.5%	5.3%	91.2%
Overall	68.9%	73.5%	5.6%	93.4%

These values are obtained analyzing 6591 controller events, with their corresponding call signs. 3137 controller events were corresponding to en-route recordings while 3454 controller events were from approach sectors recordings. The results evaluation has been done by human listeners with a solid ATC background.

On the light of these results, it seems clear that the WDR, being acceptable, still needs more refining and training to be suitable for fully automated transcription, but can be considered as a useful tool in semi-automatic transcription, providing a suggested transcription to the operator that will reduce the necessary effort for these tasks.

In terms of the main objective of this system (ATC controller voice events detection) the results are very promising, showing both strengths and areas of improvement. Focusing on the EDR, it is observed that detection rate is consistently over 70% both in en-route and approach communications. As it has been said, the event detection takes advantage of syntactical and grammar structures defined according to the field analysis of ATC voice communications, as well as of the presence or absence of keywords, to compensate for the lower values of WDR. These values must be interpreted taking into account that, as stated, an event is only considered correct when both event categorization and call sign detection are marked as correct.

This is the main reason why the $EDR_{no\_callsign}$ indicator was introduced: to provide a quantitative approach of how well the event detection submodule was performing if call signs were not considered. In fact, the values obtained just in controller event categorization are above 90%, especially for en-route sectors, thereby in a range that allows operational deployment.

During all the prototype development, call sign detection proved to be extremely challenging, even though it was initially not expected to be. The variety of ways to refer to the same call sign/flight, the use of airline aliases when communicating with them (e.g. Speedbird for British Airways), the existence of unexpected and previously unreferenced military call signs and, especially, the way of working of the system, in complete isolation from the ATC system (without access, intentionally, to any traffic information), has dropped the overall EDR to 73.5%, as opposed to the over 93% obtained when call sign detection is not considered. It has been checked that this difference of almost 20% is mainly caused by the lack of correct call sign detection (partial detections may be achieved in some cases, but are not considered, as evaluation of correct or wrong has been done in a strictly binary way). Cases of correct call sign detection and erroneous event allocation are marginal.

Solutions have been applied to allow better call sign detection (emulating controller situational awareness), especially the appliance of Call Sign Similarity (CSS) Rules [26], and the post-processing methods described in the system architecture section, trying to emulate the visual information that the controller has on screen when emitting the voice command to the pilot (e.g., when the controller refers in a voice communication to a flight only by the airline name or alias without mentioning the call sign numbers because there is no other flight in the sector of the same airline). However, the approach of keeping the ATC Speech Recognition system isolated from the traffic information was still maintained until the end of the prototyping. Only after it, initial testing have been done with the prototype having access to the traffic data

when doing the ATC speech recognition, and initial results show that EDR meets $EDR_{no\_callsign}$, even increasing both of them as better percentage of *"call sign-rest of the phrase"* segmentation is achieved. Work on this line is on-going.

Especial relevance must be given to the fact that the high $EDR/EDR_{no\_callsign}$ values are achieved while keeping FPR under acceptable limits (in the $EDR/EDR_{no\_callsign}$ calculations, false positives are already penalizing); however, works to reduce FPR are in progress, focused in enhancing the event detection algorithm. Connection to the ATC System as well as improvements in the WDR is expected to indirectly contribute to reduce the FPR.

Additionally, in terms of robustness the system has been tested under heavy processing load and queuing conditions, being able to successfully process more than 600 hours of raw (before silence removal and segmentation) ATC voice communications. This testing, under strong production conditions, was the last step before deployment in operation environment that is currently performed, regularly providing controller events from voice communications in an automated way (maturity in this aspect is a key factor in any automated system and the speech recognition engines provide additional challenges when working on heavy load conditions).

The ATC Speech Recognition prototype also includes an API to allow connection and automated operation either standalone or operated by other external systems, if required. However, the standalone end-users orientation has also been considered and a specific HMI has been developed for processing specific sessions by a human operator, as illustrated in Figure 6 below.

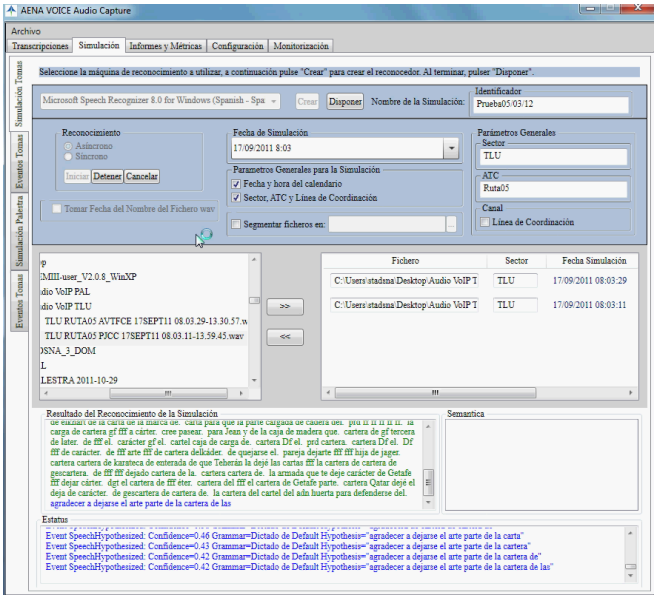

Figure 4. ATC Speech Recognition HMI.

A secondary but valuable result regarding manual transcription activities by an operator, which (as previously stated) without any help would typically need 8-10 hours to transcribe 1 raw hour (i.e., including silences) can be reduced by the use of this semi-automated tool to less than 2 hours in average, as the activities of synchronization, segmentation, coding and timing are automatically performed by the tool, also providing a suggested transcription (with the same WDR that the system provides, as uses the same speech recognition models) that the operator can directly accept or modify. Thereby, a remarkable increase in efficiency of transcription activity has been achieved, reducing the operator times.

Finally, a consideration on the WDR and $EDR_{no\_callsign}$ evolution during the several stages of the system development is that they show a clear asymptotic behavior, being more complicated to get noticeable improvements, especially on $EDR_{no\_callsign}$. During the last year new several new event detection strategies have been tested, with an improvement in $EDR_{no\_callsign}$ of less than 2%. Evolution to an objective 99% in this event detection rate is currently on its last stage of development, showing promising preliminary results.

IV. CONCLUSIONS

A lot of models, approaches and calculation methods have been historically applied to the controller workload calculation, as a key factor in determining and managing sector capacity in, for example, DCB tools. The system presented here proposes an automated approach to controller operational workload estimation, based on a continuous reliable ATC Speech Recognition system able to systematically detect controller events that will be later used for workload estimation. This system is able to interpret real operation (non-simulator) ATC voice communications both in en-route and approach with high word and detection rates, and consistently detect a wide range of controller events.

No commercial or open source Speech Recognition system available had proved suitable for ATC needs in operational environment, as determined by a wide state-of-the-art analysis, obtaining poor or even null detection rates, being the main successful application the integration of these Speech Recognition technologies in simulation environment. Thereby, for accurate analysis of ATC voice communications in an operational environment (more demanding than the simulation one) the development of a new system able to recognize, transcribe and understand what is been said was needed, emulating in an automated way what a human listener would do.

Semantic interpretation is done in such a way that controller events are "understood", detected and, finally, exported in a format able to be used by any event-based workload or complexity calculation system, with high event detection rates.

Evolution of the ATC Speech Recognition system, in particular, has proved to need a great training effort, both in terms of real ATC communications transcription to feed the core of the system and in establishing rich Hidden Markov Models logical relationships, as well as determining and tuning the system in the shape of acoustic and language models, which greatly affect the obtained word detection rate.

Surprisingly, great difficulties appeared in the correct detection of call signs, quite critical as in the presented model an event needs to be associated to a correct call sign in order to be validated; otherwise, when not concerned about call signs,

the detection rate increases to more than 90% (validated by a human listeners), which is considered a very positive rate.

The system is currently deployed allowing automated ATC event detection.

It has been decided to develop the system in complete isolation of the ATC Platforms, thereby taking no advantage of the contextual information and flight plan data contained in it, as active or pre-active call signs in the sector, in an innovative strategy that has resulted in an easily exportable system, not linked to a particular ATC Platform, and able to provide good detection results in any environment (either with contextual information available or not). It is anticipated that connection with a real ATC Platform will generate a notorious increase in call signs detection, also increasing the overall detection rates of the system, and is already in progress.

New algorithms are already under design to improve the event and call sign detections, learning from the mistakes found, but also taking into account the asymptotic behavior that the event detection rate shows and the fact that any change must be carefully analyzed and tested as it may have a negative effect in the detection of other events. This fact, together with the very positive controller event detection rate obtained, is considered as a maturity indicator of the system presented.

In the next future, extension to Tower controller-pilot voice communications is planned, while additional work shall also focus on obtaining similar detection rates in the radio channel (for the communications originated by the pilots that the controller hears) based on a specific acoustic model. On a parallel development line, cross check with other controller event potential sources, such as flight plan or radar data, will be done in order to achieve the widest possible controller event detection.

REFERENCES

[1] G. Churcher, "Speech Recognition for Air Traffic Control", Technical Report, Leeds University, School of Computer Studies, 1996.

[2] D. Schäfer, "Context-Sensitive Speech Recognition in the Air Traffic Control Simulation", 4th USA/Europe Air Traffic Management R&D Seminar, Santa Fe, 2001.

[3] R. Bolczak and J. Celio, "Accommodating ATC system evolution through advanced training techniques". American Institute of Aeronautics and Astronautics 5th Aviation, Technology, Integration, and Operations Conference, 2005.

[4] D.K. Schmidt, "A queuing analysis of the air traffic controller's workload", IEEE Transactions on Systems Man and Cybernetics, vol. SMC-8, no. 6, pp. 492-498, 1978.

[5] P. Averty, S. Athènes, C. Collet and A. Dittmar, "Evaluating a new index of mental workload in real ATC situation using psychophysiological measures", 21st Digital Avionics Systems Conference, 2002.

[6] B. Sridhar, K.S. Sheth, and S. Grabbe, "Airspace complexity and its application in air traffic management", 2nd USA/Europe Air Traffic Management R&D Seminar, 1998,

[7] C. Manning, and E. Pfleiderer, "Relationship of sector activity and sector complexity to air traffic controller taskload", FAA Oklahoma Technical Report, 2006.

[8] M. Cano, P. Sánchez-Escalonilla, M. Dorado, "Complexity Analysis in the next generation of Air Traffic management system", 26th Digital Avionics System Conference, 2007.

[9] S. Athènes, P. Averty, S. Puechmorel, D. Delahaye, and C. Collet, "ATC complexity and controller workload: trying to bridge the gap", HCI-Aero, 2002.

[10] C.D. Wickens, Engineering Psychology and Human Performance, Harper-Collins, 1992.

[11] C. Manning, and E. Pfleiderer, "Relationships between measures of air traffic controller voice communications, taskload, and traffic complexity", 5th USA/Europe Air Traffic Management R&D Seminar, 2003.

[12] C. Manning, S. Mills, C. Fox, E. Pfleiderer, and H. Mogilka, "The relationship between air traffic control communication events and measures of controller taskload and workload", 4th USA/Europe Air Traffic Management R&D Seminar, Santa Fe, 2001.

[13] K. Cardosi, "Time required for transmission of time-critical air traffic control messages in an en-route environment", International Journal of Aviation Psychology, vol.7, pp.171-182, 1993.

[14] Y. Wang, M. Hu, P. Bellot, F. Vormer, and V. Duong, "Spatial, temporal, and grouping behaviors in controller communication activities", 9th USA/Europe Air Traffic Management R&D Seminar, Berlin, 2011.

[15] J.B. Raja, J.M. Cordero, and J.M. de Pablo, "Reconocimiento y síntesis de voz en sistemas automatizados de control de tráfico aéreo", Congreso de Ingeniería de Transporte, Spain, 2008.

[16] J.M. Cordero, M. Dorado and J.M. de Pablo, "Automated speech recognition in ATC environment", ATACCS' 2012, London, 2012.

[17] H. Said, "Pilots & Air traffic controllers phraseology study", International Air Transport Association (IATA), 2011

[18] J. Rakas and S.Yang, "Analysis of multiple open message transactions and controller-pilot miscommunications", 7th USA/ Europe Air Traffic Management R&D Seminar, Barcelona, 2007.

[19] F.Fernández, R. de Córdoba, J. Ferreiros, V. Sama, and L.F. D'Haro, "Language identification techniques based on full recognition in an air traffic control task", International Conference in Spoken Language Processing, 2004.

[20] B. H. Juang, L. R. Rabiner, "Hidden Markov models for speech recognition" in Technometrics, vol. 33, No. 3, pp. 251-272, 1991.

[21] C.Geacar, D.Ion, and A.Stoica, "Language modeling in air traffic control", Bucharest Politechnic University Science Bulletin, Series D, Vol. 74, Issue 4, 2012.

[22] H. Hering, "Technical analysis of ATC controller to pilot voice communication with regard to automatic speech recognition system", EEC Note No. 01/2001, EUROCONTROL Experimental Centre, Paris, France, 2001.

[23] H. Gish, M. Siu, and R. Rohlicek, "Segregation of speakers for speech recognition and speaker identification", International Conference on Acoustics, Speech, and Signal Processing, 1991.

[24] S. Furui, "50 years of progress in speech and speaker recognition", Proceedings of SPECOM, pp. 1-9, 2005.

[25] H. David, "Speech generation and recognition: state of the art application to ATC studies at EEC", EEC Note No. 26/1996, EUROCONTROL Experimental Centre, Paris, France, 1996.

[26] EUROCONTROL Call Sign Similarity user Group, "Call Sign Similarity Rules", April 2010.

[27] K. Cardosi, "An analysis of en-route controller-pilot voice communications", Volpe National Transportation System Center, DOT/FAA/RD-93/11, US Deparment of Transportation, FAA, 1993.

[28] L. Besacier, C. Bergamini, D. Vaufreydaz, and E. Castelli, "The effect of speech and audio compression on speech recognition performance", IEEE 4th Workshop on Multimedia Signal Processing, pp. 301-306, 2001.

[29] International Telecommunication Union-Telecommunication Standardization Sector (ITU-T) Recommendation P.800, "Methods for subjective determination of transmission quality", 1996.

[30] J. Proakis, Digital Communications, 3rd ed., Mc Graw-Hill, 1995.