

Automated Speech Recognition in Controller Communications applied to Workload Measurement



ineco

Third SESAR Innovation Days
27th November 2013
Stockholm, Sweden

José Manuel Cordero –CRIDA
jmcordero@e-crida.aena.es

José Miguel de Pablo - CRIDA
Manuel Dorado – AENA
Natalia Rodríguez-CRIDA

- Final objective of the system:

- *Workload measurement*
- *in an automated way*
- *in operation environment*

Automated
Speech
Recognition
in ATM
as a medium,
not an end

- Dual approach:

- Automated voice controller events detection
- Underlying technology (ATC semantic speech recognition)



"Unlocks" the way for ASR
applications in operation

Setting the scenario: what is this all about?

- ASR in ATM has proved to be **very challenging**
- Various reasons:
 - Immaturity of natural speech recognition technology
 - Separation from standard ICAO phraseology
 - Multilingual
 - Need of a highly reliable system (*less than that may even increase workload*)
 - Difficult to access to real ATC communications
 - High user expectations (*and growing!*)
- Applications mainly in Simulation environment, until recently

Setting the scenario: A long story short

- AENA: Initial research around 2006
 - Pseudo pilots scheme in real-time simulation environment
- Extremely difficult in initial stages to achieve effective speech recognition
 - COTS didn't provide acceptable detection rates (under 30%)
 - For simulation purposes, the integration with the ATC Platform allowed to mitigate the problem (however, speech recognition itself was poor)
- Decision to make a “non-contextual information” approach

- What does it mean "*non contextual information*"?
 - No integration with ATC Platform, so no information of Flights to help on detection (*standalone ASR*)

+	-
-Independency from ATC Platform (easy adaptation) -Usable (as a service) in many other applications -Better ASR	-Increases difficulty of detection (wider constellation) -Requires more training/modelling to get similar results

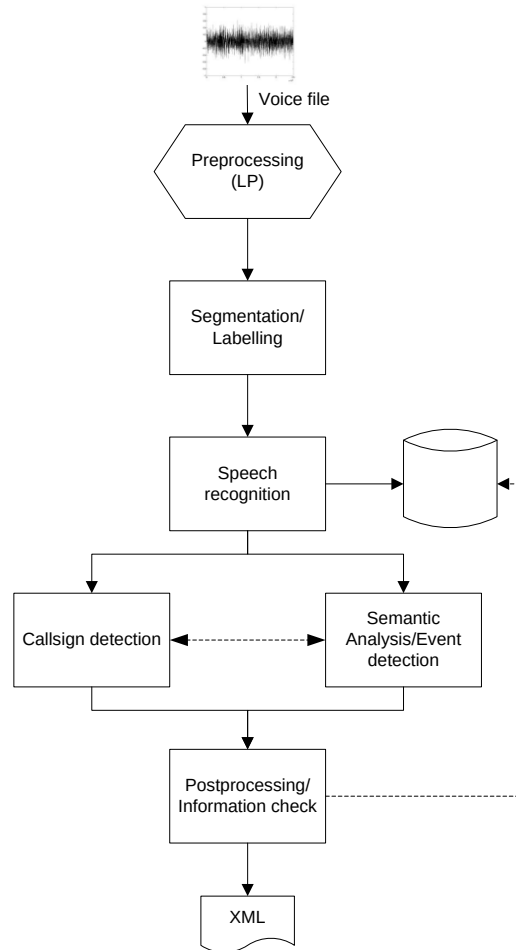
- However, ATM logic is inside the detection model (even the scenario can be included).

Setting the scenario: On the other side, some strenghts



- Wide set of real ATC communications available
- Close collaboration with operational staff (trainers)
 - Validation/calibration
 - Event model refinement
 - Language interpretation
- Reoriented objective: **Workload estimation by voice recognition (in real operation recordings)**
 - Calculation through detected controller events
 - Voice is an essential source of information

The underlying technology: ATC Event Detection Functional Scheme

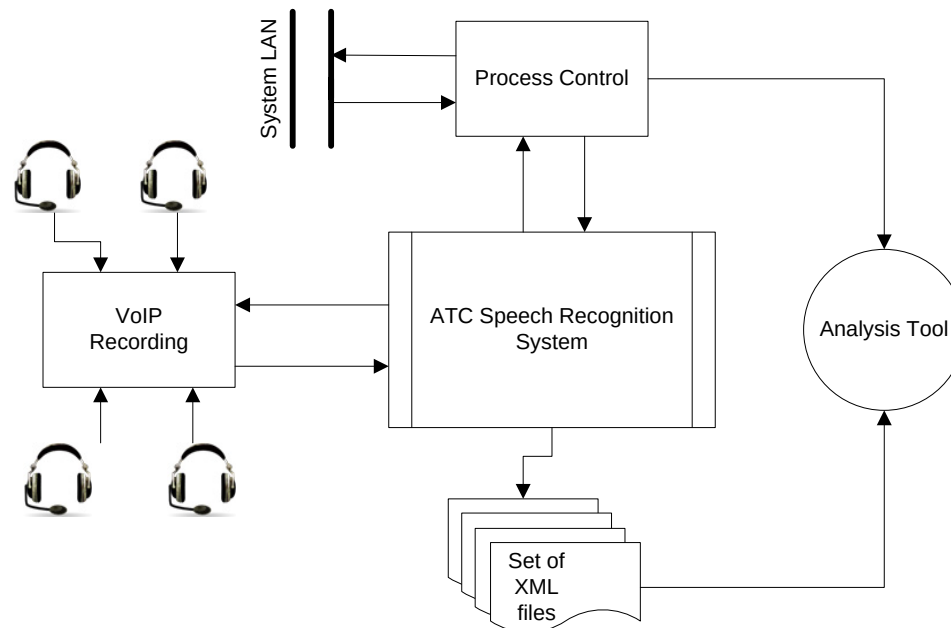


- Preprocessing
- Segmentation/ Labelling (silence removal, ...)
- Speech recognition (HMM)
 - Language Model
 - Acoustic Model
 - Extensive training
- CS detection/Event detection-> Algorithms, keywords+logic
- Postprocessing/Refining
- XML (Output)

The *enabler*: System Training

- The ASR Module needs to be trained with transcriptions (from real ATC communications)
- Transcriptions are very time-consuming and done manually -> Transcription-aid strategy
- Current prototype contains **more than 100 net (no-silence transcribed hours)** (both en-route and TMA), with 100% reliability (human check), corresponding to aprox. 500 raw hours (with silences)
- *Evolution strategy: limit 100% accuracy manual transcriptions, use those with automated confidence index >95% -> Improvement in WDR*

- Sector configuration in CWP to be extracted from the ATC system
- VoIP recording (NICE System)
- Double Workload calculation based on controller events (Wickens/MWM and NORVASE)



The output

- 1 XML file per sector, per hour, combining channels: **set of events**

```
- <event>
  <name>Av</name>
  <callsign>ANE8390</callsign>
  <UCE>Ruta 15</UCE>
  <date>17/09/2011</date>
  <time>08:11:33</time>
  <duracion>4</duracion>
  <coord>No</coord>
  <message>air nostrum ocho tres nueve cero descienda a nivel de vuelo
    uno cinco cero</message>
</event>
- <event>
  <name>Av</name>
  <callsign>ANE8390</callsign>
  <UCE>Ruta 15</UCE>
  <date>17/09/2011</date>
  <time>08:11:46</time>
  <duracion>3</duracion>
  <coord>No</coord>
  <message>air nostrum ocho tres nueve cero esta establecido cuarenta
    y cinco millas</message>
</event>
```

Level Change Event

Communication
Automated
Transcription

Some numbers (results)-Feb 2013

- Metrics for an ASR applied to Workload estimation:
 - **WDR: Word Detection Rate.** Is more usual to find WER (*Word Error Rate = substitutions+deletions+insertions/total real words*), $WDR = 100 - WER$
 - **EDR= Event Detection Rate** (*An event is considered correct when type of event and CS are OK*)
 - **EDR_{no callsign}= Event Detection Rate without callsign** (*Only considers event categorisation*)
 - **FPR= False Positives Rate**

Some numbers (results)-Feb 2013

- Results obtained from a set of **60 raw hours** not included in the training of the system (control group) (**6591 events**)

	WDR	EDR	FPR	EDR _{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%

- Rates better than any other known product applied to ATC communications, **continuously evolving**
- Later on, the workload calculation can be performed using diverse methodologies

A look to workload measurement

- After the events are obtained, they are cross-checked with those detected from pure FP and radar data.
- A set of events for a period of time is determined, and send to two different workload calculation modules:
 - MWM (MultiWorkload Model, based on Wickens cognitive workload model
 - NORVASE (Sector Validation Normative), based on Spanish normative

A look to workload measurement

- NORVASE is particularly relevant as automation of the workload measuring process allows a bigger number of samples for all sectors, thus increasing the accuracy of the measure (versus manual takes, very limited and selective).
 - More workload samples
 - More sectors measured
 - Fully automated
 - Cost efficient

Which events necessarily need voice?

- Focus is put in three of them, where voice analysis is key for effective event detection:
 - i) **Direct/heading determination**
 - From simple radar data analysis is difficult to determine
 - Voice is the most reliable source
 - Mistake in this determination has a big impact in workload
 - ii) **Effective sector exit**
 - Radar data allows geographical exit, but not frequency transfer
 - Key as the moment when the event happens is relevant for workload
 - Only obtainable through voice analysis
 - iii) **Inter-sector coordination**
 - Unavailable from any other source

Events Detected (rates Feb 2013)

Event code ^a	Event Description	EDR _{no_callsign}	Com. duration (s)		Ocurr En-route	Ocurr TMA
			μ	σ		
CTEv	Sector Entry Communication	96.2%	3.1	1.8	33.26%	17.88%
Csv	Sector change Communication to Pilot	98.5%	3.9	1.8	32.42%	19.87%
Dv	Direct Communication	92.1%	2.7	0.9	2.01%	0.33%
Xv	Heading Type 1 Communication	90.1%	3.7	1	0.13%	0.00%
Sv	Heading Type 2 Communication	91.5%	4.6	2.9	0.13%	28.48%
Vv	Speed change Communication	94%	3.3	0.9	1,34%	7.28%
Av	Level change Communication	96.6%	1.8	2	17.38%	9.38%
Cov	Inter-sector controller-controller coordination	79.7%	7.8	7.6	8.56%	2.32%
Ac3.4.11v	Clearance or intruction Communication	93%	3.3	1.3	0.87%	0.00%
Ac7v	ILS Authorization Communication	91.3%	3.6	1.4	0.53%	3.64%
Ac13.1v	STAR assignment Communication	90%	3.9	2.5	N/A	0.53%
Ac9v	Essential information Comunnication	80%	5.8	5.3	2.41%	8.83%
H1v	Holding stack Communication	87.2%	2.3	0.8	0.40%	1.10%
CRv	Clearance/authorization Correction communication	88.8%	2.4	1.1	0.67%	0.33%

Model optimised for en-route detection

En-route: 43,25% events voice detection has a key role for workload

TMA: 51% events voice detection has a key role for workload

What's next?

- As stated, underlying technology unlocks and enables the way for new applications
- SESAR Exercise EXE-04-07.01-VP-003, *"Resolving Complexity by dynamic management of airspace"*
- V2 exercise, OFA05.03.04
- Voice will be analysed on-line for complexity indicators calculation, using the same ASR technology described.

- Automated measurement of controller workload, based on ASR, in operation environment
- Dual approach: Application and enabler
- Set of events provided, for later WL calculation in two modules
- Nice EDR (around 75%), very good EDR if not considering callsigns (over 90%)

Conclusions (II)

- Voice information key element in >40% events en-route, >50% events in TMA
- Need to evolve some detection algorithms (especially callsigns)
- Plan to include Airports (2014)
- Final integration with ATC system for virtual 100% EDR (2014-2015)
- Other applications now feasible (*even virtual pseudo-pilots*)

Thank you!

Any questions?



Centro de Referencia I+D+i ATM

