

Online/offline transcription

By

Ruben Gaspar – IT-CDA-IC

Agenda

- Background and Scope
- Vendors
 - AI-Media
 - S2T
 - Amberscript
 - Mozilla DeepSpeech
 - MLLP
- Conclusions

All my notes can be found at

<https://codimd.web.cern.ch/8LsIM46XSsK1hhuJZtRQOQ?view>

Scope

- Looking into Online transcription on the frame of Video and Webcast services
- Starting to look at automatic transcription in the framework of lecture processing (CES – Micala workflow)
- Digital memory is a different problematic, much more complex in general. Though there are clear interconnections
- **Functionality** we are looking for:
 - Trying to cover both **online** and **offline** automatic transcription. **Equally important.**
 - Trying to get also a tool that can **translate** a transcript
 - Trying to get a tool that can provide **voice synthetization**

AI-Media

- ASR provided by Google, Speechmatics,.. → several models
 - [Contextual help \(topic adaptation\)](#) can be provided [depending on the engine](#) used, e.g. google accepts a dictionary
- [No REST API](#) or Interface to test provided
 - For my test I needed to send the files to James (AI-Media salesman)
 - A demo that can be watched [here](#) a WebUI that would implement the workflow requested. A REST api will be also provided
- [EEG encoders](#), they can add transcription at encoding time. They also have a [WOWZA module](#) for transcription
- Human translation possible (expensive)
- Voice synthetization possible.
- WebUI available.
- ASR about 10 Euros/hour
- Online: only accepts two languages. No more details.

WER (version2)	WER	Indicoid
57.3	62.8	539453
36.5	48.4	536271
43.4	43.4	506114

Amberscript

- Specialized on minor languages e.g. German, Dutch, French... they do also English
- [Integration](#) with [Openstack](#) (PR in github)
- API available, No topic adaptation
- Human translation only: expensive 1.9Euros/minute.
- No voice synthetization
- No online transcription
- WebUI player for transcript corrections
- 20 Euros per hour.
 - They sell packages of 20hours that can be used in 12 months

WER (latest)	Indicoid
36.2	536271
37.4	506114
58.3	539453

Mozilla's DeepSpeech

- Open Source, just about speech recognition.
- It looks to me the way to go in the future
 - Using the community to improve ASR models: <https://voice.mozilla.org/en>
- Tested on my Mac, not really good results:

WER	Indicoid
78.1	539453
69.6	536271
67.5	506114

S2T WIPO

- Tested by JY
- They provide a WebUI player for transcript corrections.
- Looks very hacky (my impression) to setup. Best with a workflow that could split GPU and CPU phases.
- No API, No queue system (state machine needs to be implemented)
- Alpha for online transcription (In my opinion they are far away)
- No voice synthetization
- Code is free for CERN
- Consultancy agreement: 20k/year for 1 engineer month work to do tasks like update the model.
- Despite it may work only with CPU's, it's not the way WIPO works so it would be wise to run with GPU servers.

Machine Learning Language Procession group (MLLP)

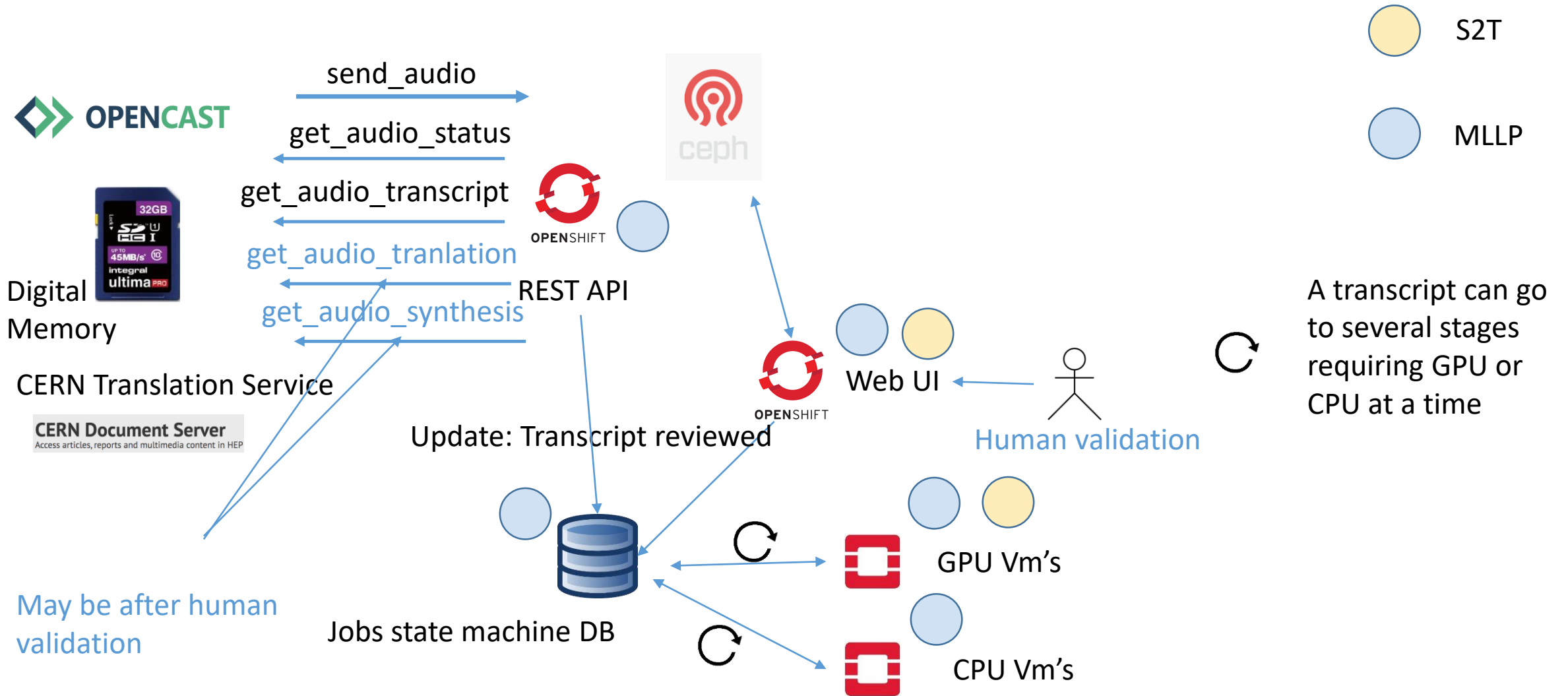
- More than 10 years working on Speech recognition
 - Very well ranked on International competitions (see codimd)
 - Working on ASR for several EU projects
 - Behind all transcripts for Polimedia (MOOC/lectures from UPV)
- Transcriptions, best ones using Maria's lectures (not a trained model).
 - Show example form Medical convention: <https://ttp.mllp.upv.es/s/apyJO>
 - Topic adaptation (pdf, text files)
- Translations: including French, Spanish, German,...
- Synthesis of voice: explain Elena's use case: <https://ttp.mllp.upv.es/ttsdemo/>

WER (latest)	Indicoid
36.2	536271
37.4	506114
58.3	539453

MLLP (2)

- Rich REST API and Python library
 - Not based on Kaldi libs but their own development.
- WebUI for lectures captions correction: <https://ttp.mllp.upv.es>
 - Integrated with Paella plugin
- Online transcription – demo (<https://ttp.mllp.upv.es>)!
- Way to go further with them:
 - **Engage on a Pilot**: they will create CERN model, check integration with our SSO, Paella plugin integration of the WebUI, etc. 5k Euros/6 months
 - **If Pilot is successful**, how we could continue:
 - UPV Cloud: **pay as you go** 3 Euros/hour
 - UPV Cloud: dedicated servers, flat rate, full-time engineer: support, model adaptation, etc 55k Euros/year
- Research project for Digital Memory: [See JY video at MLLP site](#) (2019-12-11_17-58-43)

Weblecture processing workflow



Conclusions

- **First choice: MLLP**
 - Most complete set of **features covering all** our nowadays and future needs
 - Pilot should help to create a CERN ASR model, develop **topic adaptation** with RNN and validate integration with CERN environment
 - **Research project** for Digital Memory use cases
- **Second choice: S2T + MLLP**
 - S2T for offline and translation: it will need quite some development in my opinion
 - Online based on MLLP: cloud based solution
- **Time frame, depends on the options**
 - PJAS considering COVID also considering that the person starts on July 2020 remotely.

Conclusions: about MLLP

- Cloud solution at first. We could evolve to a on premise setup
- Spanish Research group compliant with <https://gdpr.eu/checklist/> but not certified
- Data processing agreement <https://gdpr.eu/data-processing-agreement/>
 - Offline data can be removed (hard option) via API
 - Online data, nothing is retained.

model	WER evaluation	WebUI	API	Transcription (ASR)	Translation	Online transcription	Voice Synthesis	Cost
Amber	2nd	Yes	Yes	Yes (En, Fr)	Manual	No	No	20Euros/hour
AI-Media	3rd	Yes	Yes	Yes (En, Fr)	Automatic	Yes	Yes	9.81 Euros/hour
S2T	N/A	Yes	No	Yes (En)	Automatic	No (alpha stage)	No	Code for free, 20k for 1 month consultancy
MLLP	1st	Yes	Yes	Yes (En)	Automatic	Yes	Yes	3 Euros/hours, or 55k per year with a dedicated engineer