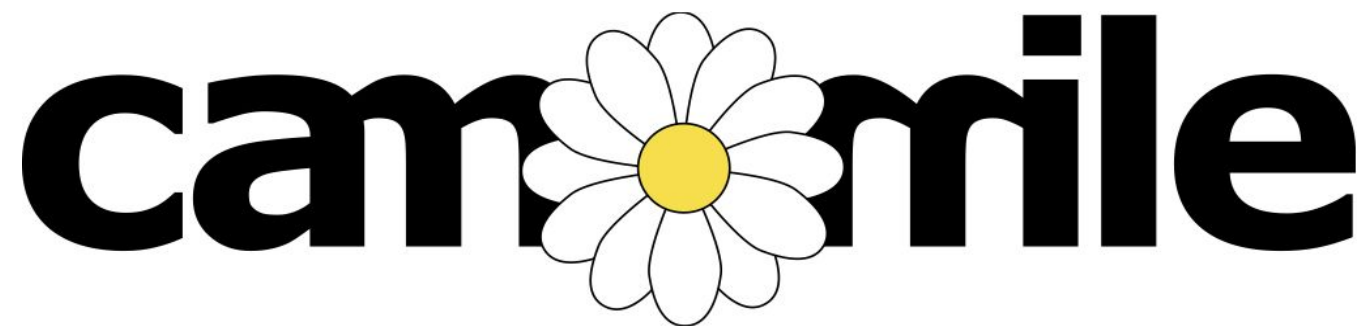
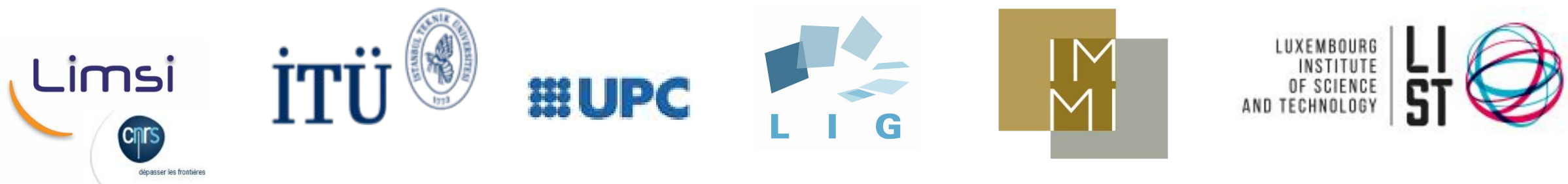




CHIST-ERA Project Seminar

Bern, April 28, 2016





Project Summary

- Collaborative **A**nnotation of multi-**M**odal, **M**ulti-**L**ingual and multi-m**E**dia documents
- CHIST-ERA call 2011 D2K
Oct. 1, 2012 to Sep. 30, 2015 ... Nov. 30, 2016
- 6 participants from 4 countries
France, Luxembourg, Spain, Turkey

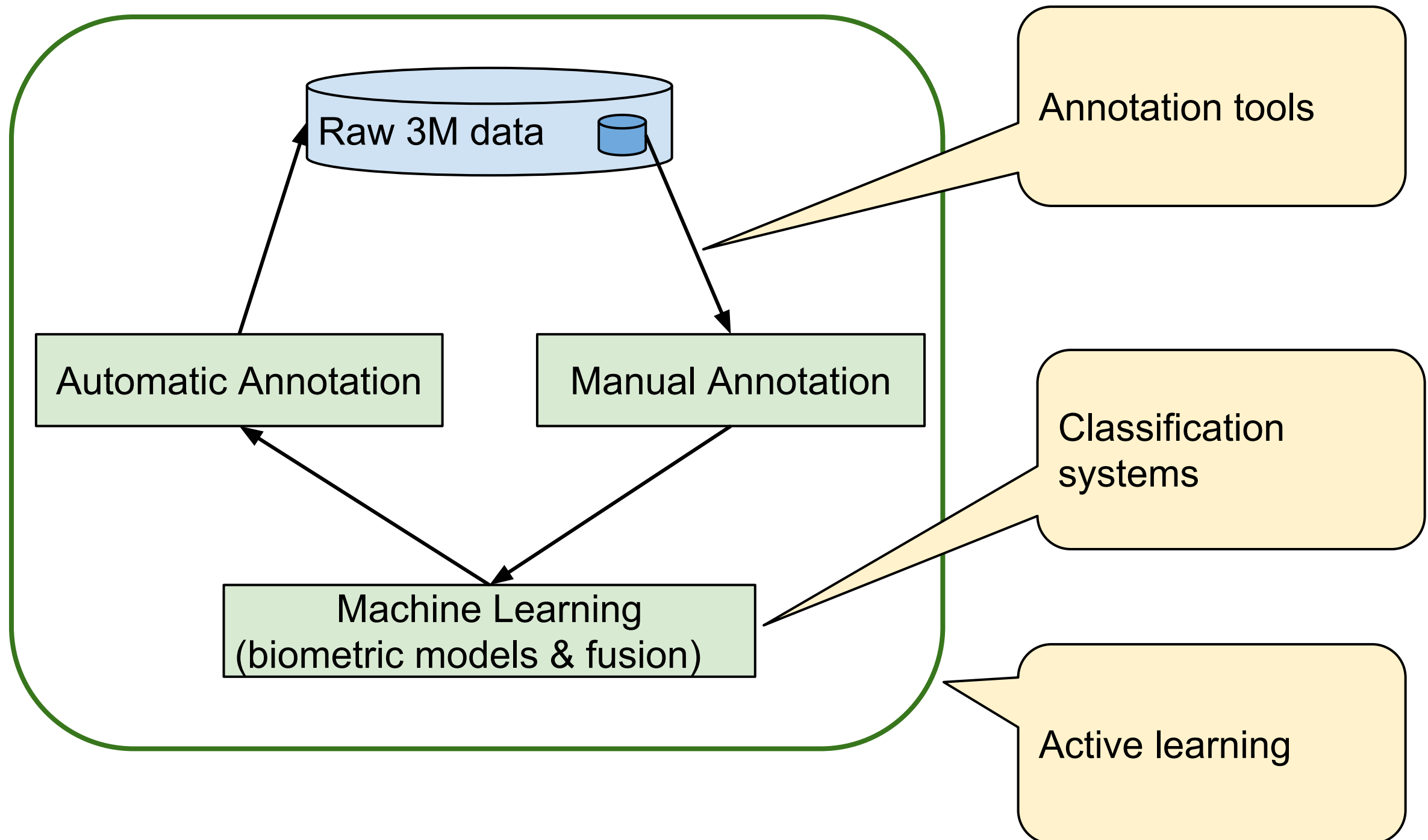


Scientific Background

Scientific Background

- Large amount of heterogeneous, human-generated data available on the web
 - collection and analysis for research (social sciences, linguistics, economics, artificial intelligence & computer science)
- Growing diffusion of embedded, perceptual systems
 - Pattern recognition and machine learning
- Statistical approaches rely on large amount of annotated data
 - human annotation costly
 - several modalities with different levels of abstraction
- Focus on 3M data (multimodal, multimedia, multilingual)
 - share an object of study between scientific domains

Scientific Background





Challenges and Impact

Key challenges and objectives

- Challenges

- Collaborative annotation framework
 - remote annotations combined locally
- Annotation of 3M data
 - improve state-of-the-art automatic systems
 - integrated framework for the different modalities
- Combine automatic and manual annotations
 - output automatic treatments as help for manual annotation

- Objectives

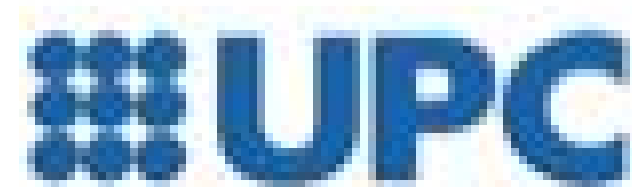
- Prototype of an experimental platform
- Proof of concept on different use cases

Goals summary

- Go to multilingual / multimodal annotation
 - Annotation for various modalities
 - Distributed environment
 - Collaborative annotation in parallel / in sequence / through crowdsourcing
- Coordination with international initiatives
 - ELDA, LDC, META-SHARE, ISO, OAC...
 - Through an international advisory committee
- Target users for collaborative annotation
 - Researchers in machine learning working on 3M data
 - Institutions aiming at producing manual annotation of 3M data
- Unifying use case: evaluation on multimodal person identification



Consortium



- Spoken Language Processing group
 - 17 academic + support staff, 7 contractuels and post-doct, 15 PhD
- Research topics
 - **Speaker characterization in a multimodal context**
 - Affective and social dimensions of spoken interactions
 - Perception and automatic processing of variation in speech
 - Robust analysis of spoken language and dialog systems
 - Translation and machine learning
 - Speech recognition
 - Language resources

- Oldest technical university in Turkey (1773)
 - The first faculty on computer and informatics in Turkey
 - Accepts students from top 1% of central university entrance exam
- Involved faculty members:
 - Assoc. Prof. Hazım K. Ekenel (Principal Investigator – Computer Vision)
 - Prof. Zehra Çataltepe (Machine Learning)
 - Asst. Prof. Cenk Demiroğlu (Ozyegin Univ., Speech Processing)



- Universitat Politècnica de Catalunya
- 2 UPC labs
 - Image and video processing group: 11 staff + 1 research assistant + 10 PhD
 - Speech processing: 10 staff + 10 PhD students
- Research lines
 - Video Indexing: Region-based semantic annotation, Face ID, Generic object detection, QbE, Text extraction and recognition
 - 3D scene analysis and coding: Multi-camera reconstruction, body analysis, Tracking, 3DTV coding
 - Speaker, Language, and Emotions recognition, Event detection (audio), Diarization
 - Multilingual and robust speech recognition, Text-to-speech conversion, Speech to Speech Translation



Consortium CNRS - LIG (Grenoble)



- 165 academic and research staff, 260 doctoral and post-doctoral students, 65 support staff
- Two research teams involved (G. Quénot, L. Besacier)
 - GETALP: Study Group for Speech and Natural Language Translation/Processing
 - MRIM: Multimedia Information Modeling and Retrieval Group
- **Related research topics** (QUAERO consortium & REPERE challenge)
 - Multimodal person recognition in TV broadcast
 - Multimedia semantic indexing
 - Corpus annotation (Collaborative annotation)
 - Evaluation (TRECVID, REPERE)
 - Active learning



Consortium CNRS - IMMI (Orsay)



- **Institute for Multilingual and Multimedia Information**
 - French-German Institute (IMMI-Labs)
involving LIMSI-CNRS (Paris) + RWTH (Aachen) + KIT (Karlsruhe)
 - Launched in the framework of QUAERO program
 - Closed on Jan. 1, 2016 - activities in the project transferred to LIMSI
- **Participation in other projects**
 - EC FLaReNet and META-NET, ANR Bulb, DGA Papyrus
- **Contributions in CAMOMILE**
 - Organization of the International Advisory Committee
 - Link with ELRA, LDC, META-SHARE...

- Luxembourg Institute of Science and Technologie (LIST)

- Created 1/1/2015 from Public Research Centers Lippmann & Tudor
- Applied Research and Technology Transfer: funding from research programs (international, national) & contract research (private, public)
- Within Environmental Research & Innovation department, eScience Research Unit applies Visual Analytics for scientific discovery and innovative data-driven solutions.

- In Camomile Consortium

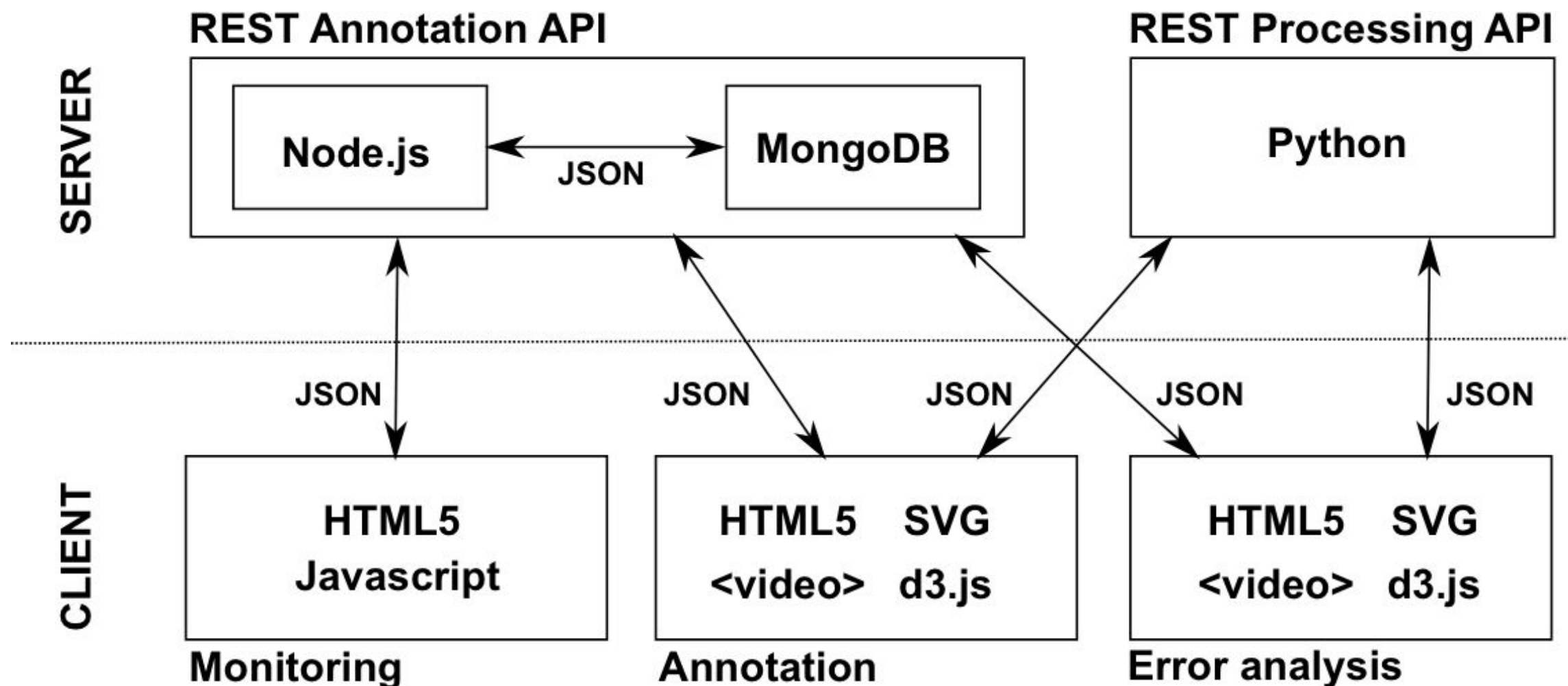
- Expert for a visualization platform that focuses on interpretation, interaction and heterogeneous data
- Task leader for Cooperative Annotation Monitoring (WP4)
- Apply the know-how of generic tools developed at LIST: Calluna (advanced visualization) & Cadral (data mining)



Main scientific results

Collaborative Annotation Framework

Interoperability through web services with controlled access to share and perform collaborative annotations



Collaborative Annotation Framework



- **OPEN-SOURCE** server and client code on GitHub
- online Slate documentation

- prerequisite
- Authentication
- User
- Group
- Corpus
 - Create a corpus
 - Get list of corpus
 - Get a corpus
 - Update a corpus
 - Delete a corpus
 - Add media into a corpus
 - Create a layer into a corpus
 - Get list of medias into a corpus
 - Get list of layers into a corpus
 - get ACL of a corpus
 - update user ACL of a corpus
 - update group ACL of a corpus
 - remove user from ACL of a c...
 - remove group from ACL of a ...

Create a layer into a corpus

Restriction: user/group rights is 'O' or 'W' for this corpus

HTTP REQUEST

POST `http://example.com/corpus/<id_corpus>/layer`

QUERY PARAMETERS

Parameter	Type	Description
id_corpus	String	The corpus identifier (required)

DATA PARAMETERS

Parameter	Type	Description
name	String	Name of the layer (required, unique in the corpus)
description	free	A description of the layer
fragment_typ	String	The fragment type (required)
data_typ	String	The data type (required)
annotations	List	list of annotations (OR , ,) to add to the new layer

shell python javascript

```
curl -i -X POST http://example.com/corpus/<id_corpus>/layer --data '{"name":<name>, "description":<description>, "fragment_type":<fragment_typ>, "data_type":<data_typ>, "annotations":[{"media_name":<media_name>, "fragment":<fragment>, "data":<data>}]}' -H "Content-Type: application/json" --cookie "cookies.txt"
```

The above command returns JSON structured like this:

```
{
  "ACL": {
    "users": {
      "5305d1a044dc07e805000003": "0"
    }
  },
  "_id": "545387f8777a800000766bf3",
  "description": {
    "abc": "def"
  },
  "history": [
    {
      "date": "2014-10-31T13:00:40.604Z",
      "id_user": "5305d1a044dc07e805000003",

```

camomile-project / camomile-server

Camomile REST API <http://camomile-project.github.io/camomile-server>

302 commits 3 branches 5 releases

branch: master camomile-server / +

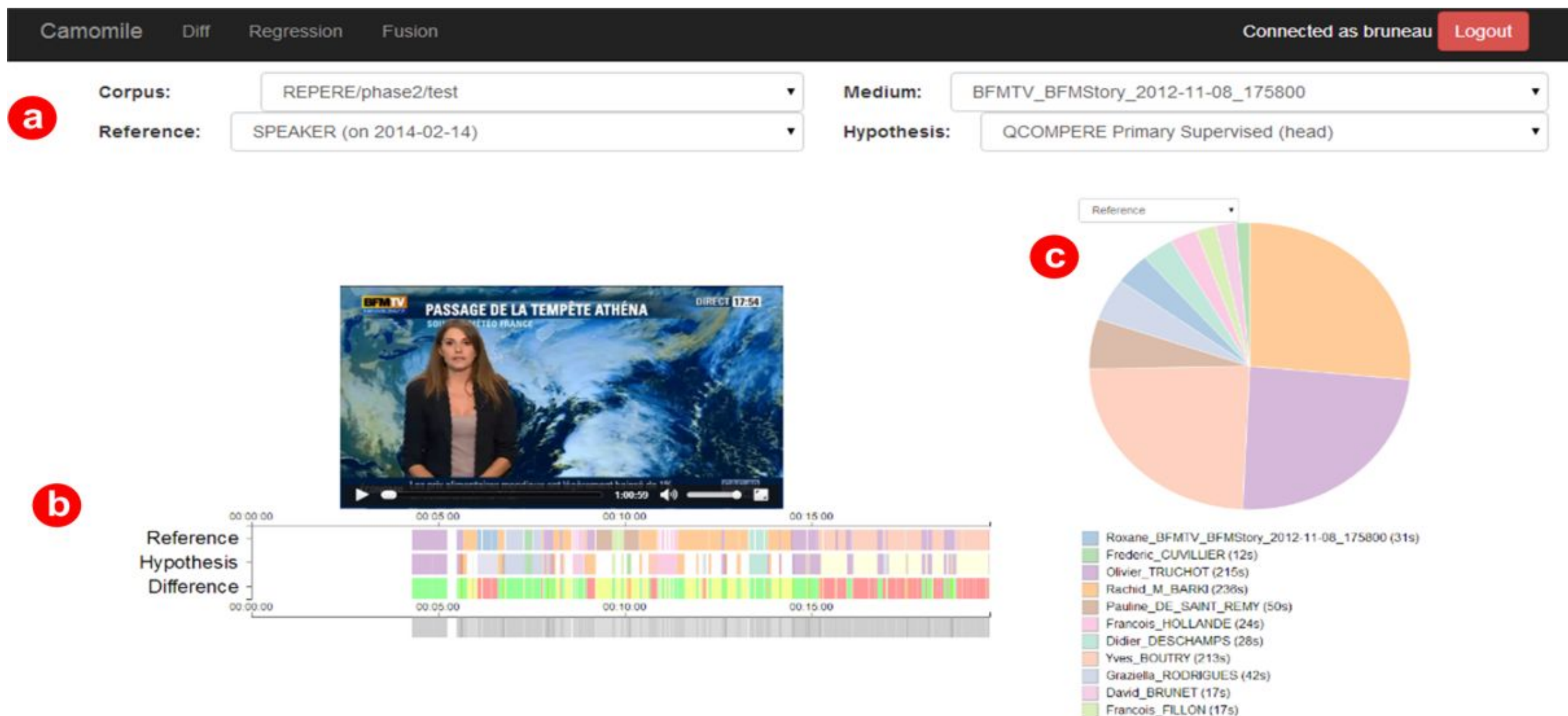
Merge branch 'hotfix/0.4.1'

hbredin authored 6 days ago

controllers	add filter for get
doc	add spec server 2.0
models	suppression var global + mv fichier
routes	fusion addannotations -> addannotation
.gitignore	ignore doc/build
CHANGELOG.md	fix: version in package.json
Dockerfile	docker: use official node image
LICENSE	chore: MIT license

Collaborative Annotation Prototype Interface

Development + incremental tests & dissemination

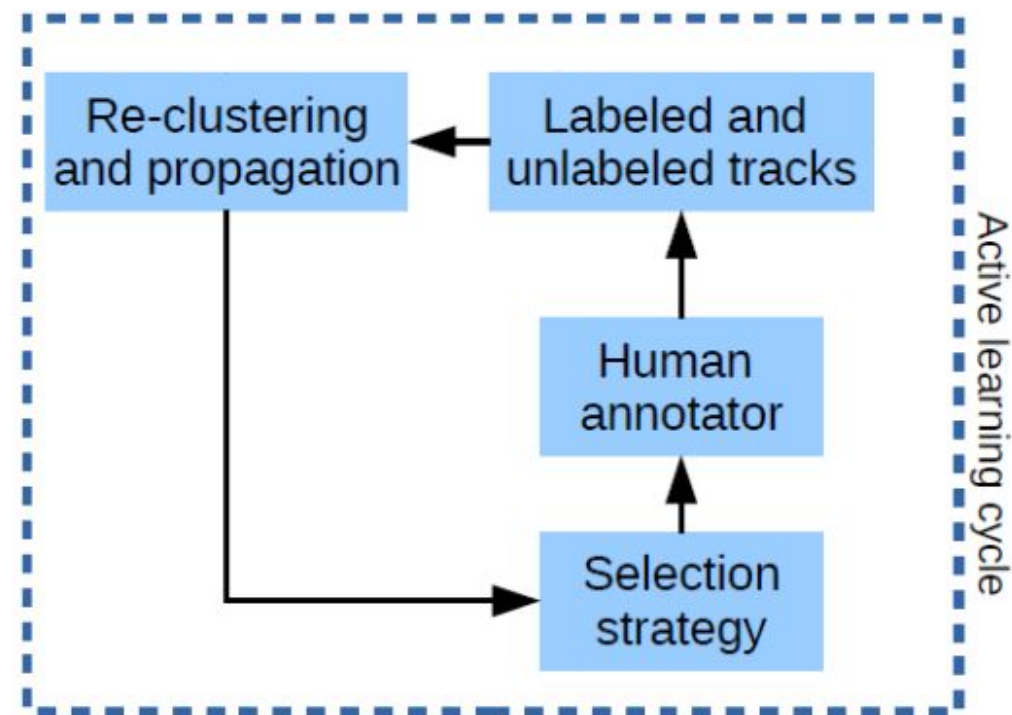


Multimodal Person Recognition

- Visual Person Identification
 - Face tracking performance improvement using context based cues
 - Perform alignment for face matching across pose
- Speaker Identification
 - Use of prosodic and voice quality features
 - Explore Deep Belief Networks (DBN) or Convolutional Networks (CNN) vs. Gaussian mixture models (GMM) and Total Variability Space (TVS)
- Fusion
 - Compare cascade or simultaneous fusion of modalities
 - Integer linear programming (ILP) resolution

Active learning for multimodal person annotation

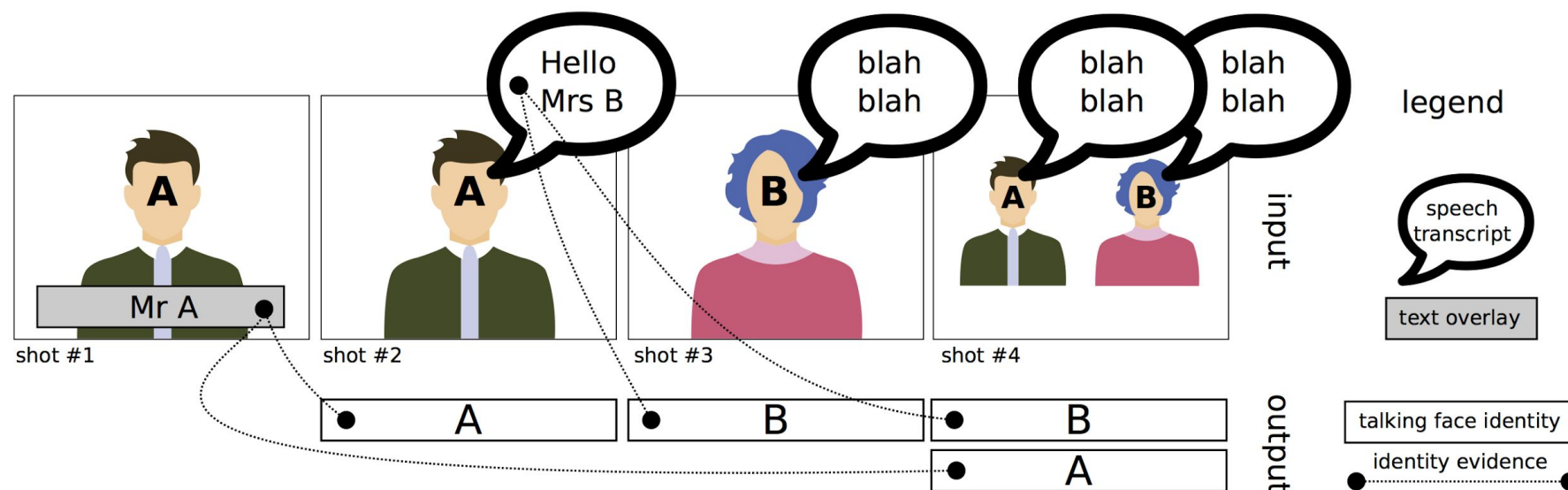
- Goal: efficient speaker model training using the least possible amount of annotated data
- Active learning (AL) approach with different selection strategies
- Based on different modalities: faces, speech, overlaid names



New MediaEval task

Multimodal person discovery

- **Multimodal**
 - Identify persons of interest (audible **AND** visible at the same time) within a shot
 - Combine speaker identification, face recognition and speech/overlaid text transcription
- **Person discovery**
 - Discover names in speech and overlaid text transcription
 - Unsupervised approaches only (no prior biometric models)
- **in TV Broadcast**
 - 2015: REPERE corpus + INA French TV Broadcast News corpus (100 hours)
 - 2016: new INA corpus + Catalan TV + Deutsche Welle (German and English)



- Distribution of a baseline system (monomodal and fusion components)
- A posteriori collaborative annotation using Camomile framework



Dissemination & Sustainability

Dissemination & Sustainability

- 12 multi-partner and 20 single-partner international communications
 - CBMI, CDVE, Eusipco, ICASSP, ICIP, Interspeech, IV, LREC, Odyssey, SLAM...
- Co-organization of ERRARE Workshops
 - “Errors by Humans and Machines in multimedia, multimodal and multilingual data processing”, 2013 & 2015
- International collaboration
 - Internship co-supervised by LIG and Ozyegin U. + joint MediaEval submission
 - 4-month stay of Pr. Tamura (Japan) at LIMSI/IMMI in 2013
- Spin-off company from Dr. Ekenel on computer vision in multimedia
- Collaborative Annotation Framework in Open Source on GitHub
 - Presentation and demo at Interspeech 2015 and LREC 2016
- MediaEval task on person discovery
 - contacts with content providers (INA)
 - 9 participating teams in 2015



Questions?
