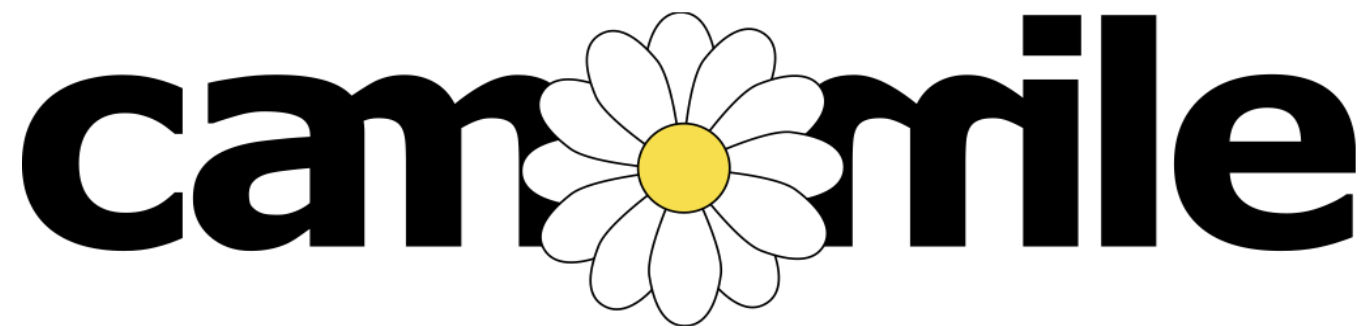


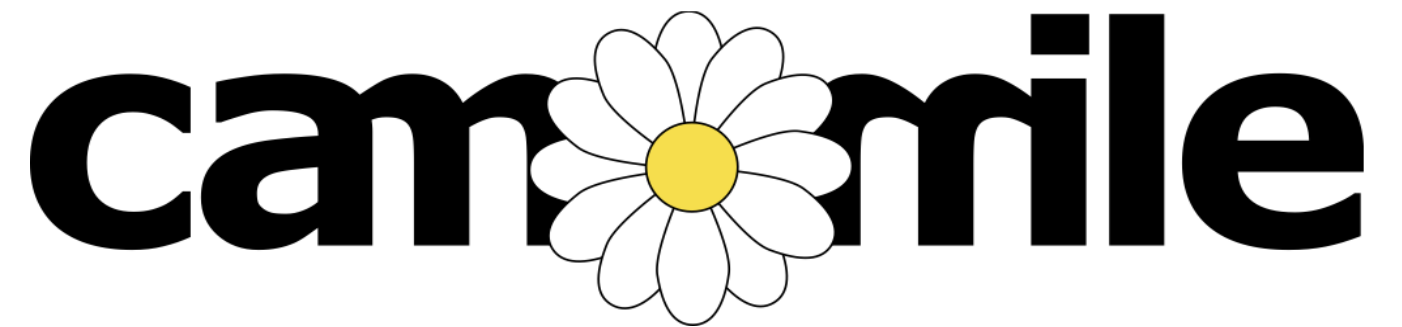


CHIST-ERA Project Seminar

Istanbul, March 5, 2014

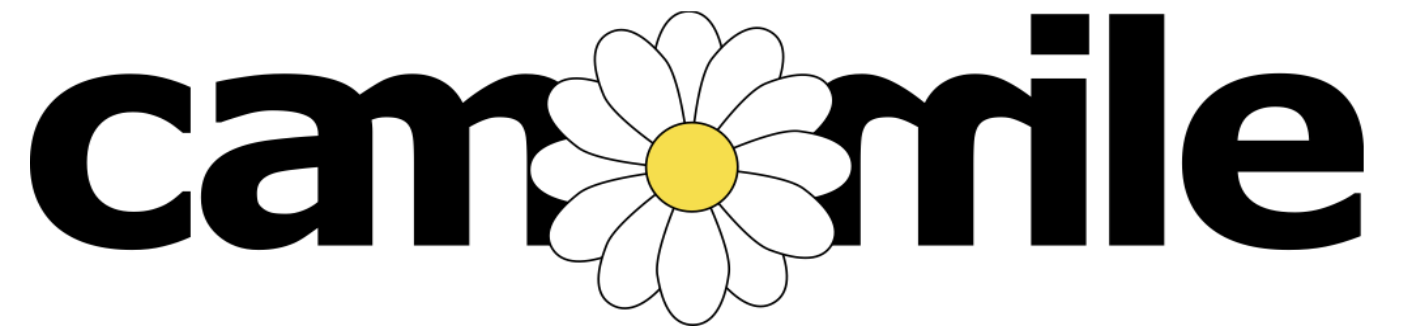


Centre de Recherche Public
Gabriel Lippmann



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 March 31, 2016 (42 months)
- 6 participants from 4 countries
 - France, Luxembourg, Spain, Turkey

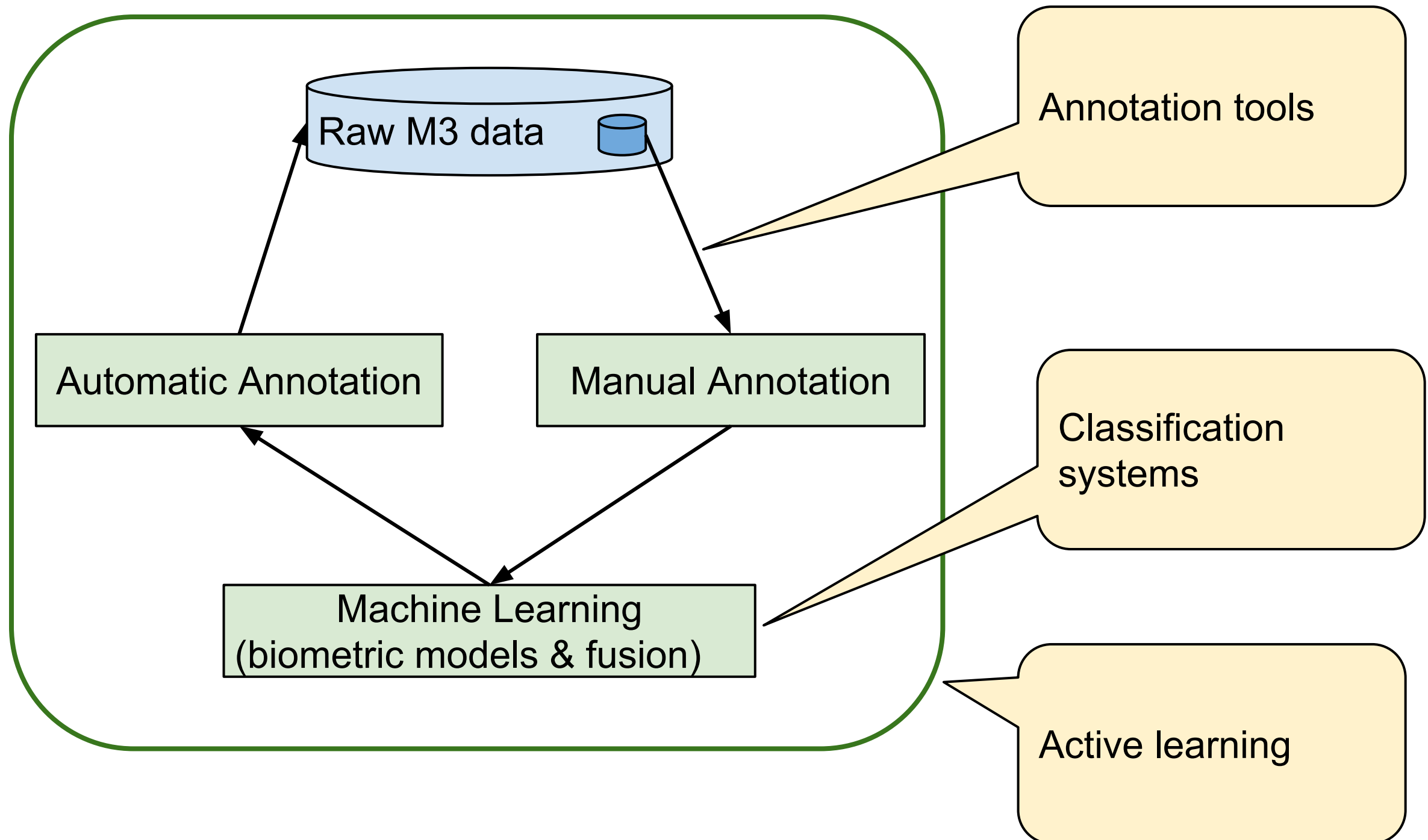


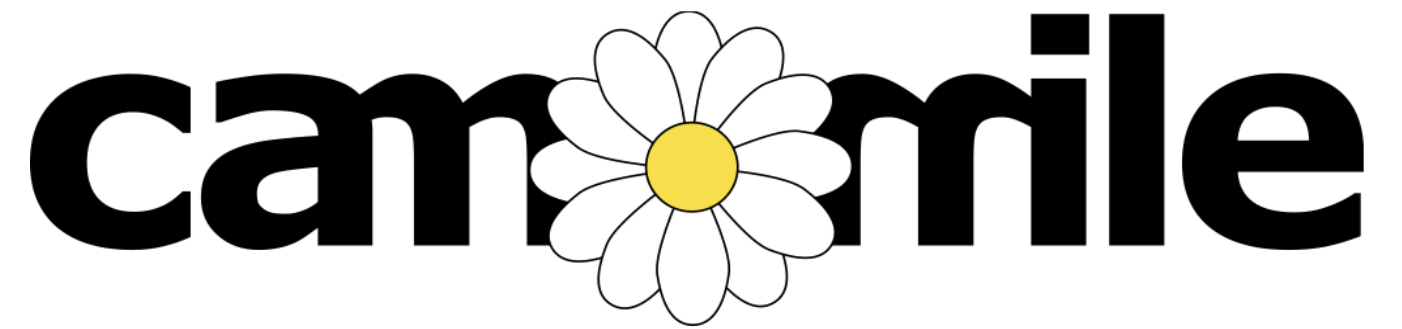
Scientific Background

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-the-art automatic systems
 - integrated framework for the different modalities
- Combine automatic and manual annotations
 - output automatic treatments as help for manual annotation

- Objectives

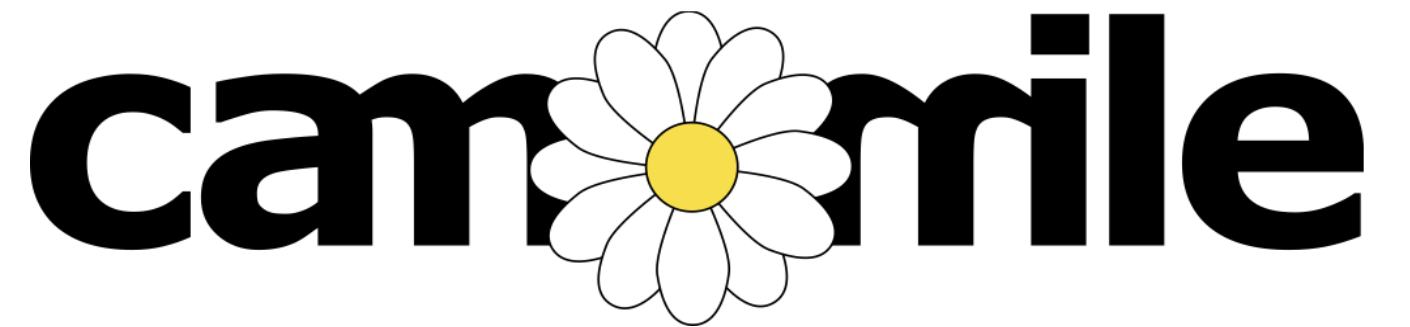
- Prototype of an experimental platform
- Proof of concept on different use cases

Use cases

- Multimodal person annotation in audiovisual broadcasts
 - REPERE challenge (2010-2014) organized by French ANR and DGA, evaluation by LNE, 3 consortiums
 - Fusion from modalities: face and speaker recognition + overlay text and pronounced names extraction
- Semantic concept detection from videos
 - TRECVID evaluation campaign
- Extension to other use cases
 - New domains (TV fictions, documentaries)
 - New languages (Catalan, Turkish)
 - Connection with Digiturk as data provider and potential technology user

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



Consortium



Centre de Recherche Public
Gabriel Lippmann



Consortium CNRS - LIMSI (Orsay)



- Spoken Language Processing group
 - 18 academic + support staff, 8 contractuels and post-doct, 12 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
- Participation to QUAERO program and REPERE challenge

- 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:
 - Asst. Prof. Hazım Kemal Ekenel (Principal Investigator –Computer Vision)
 - Assoc. 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

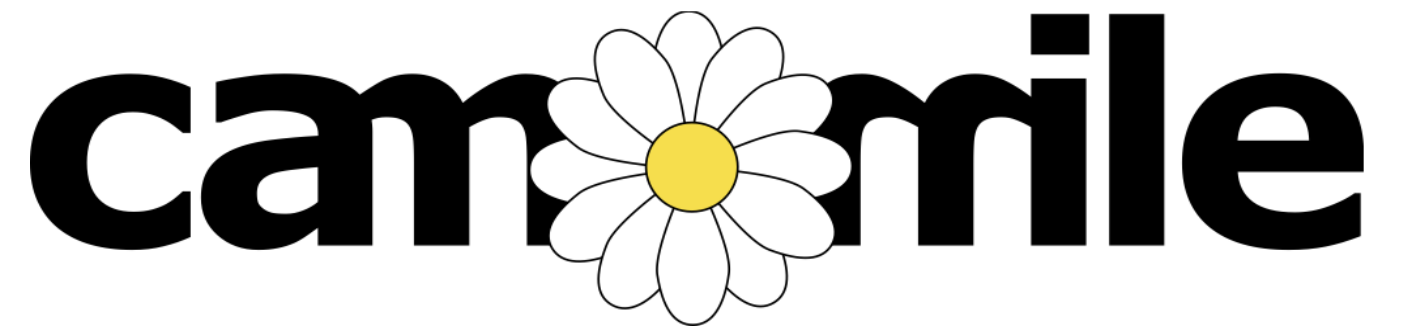


Consortium CNRS - IMMI (Orsay)



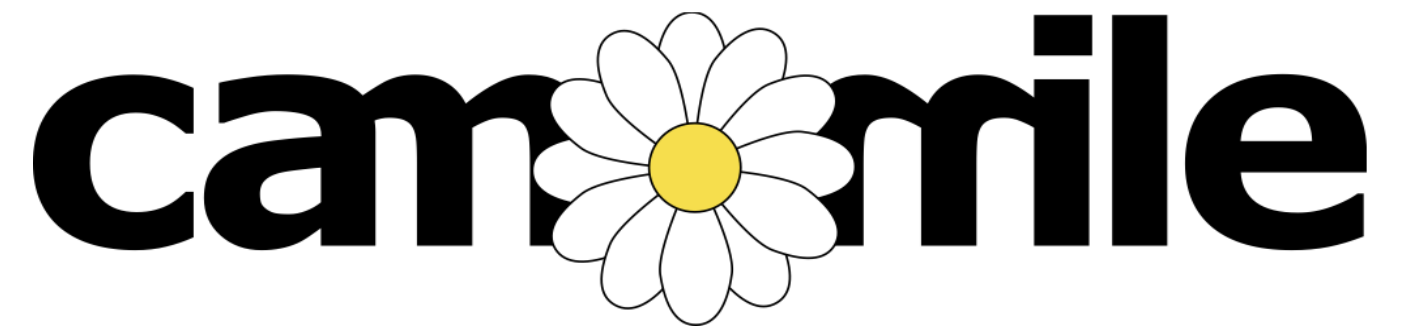
- Institute for Multilingual and Multimedia Information
- French-German Institute
 - Joint venture between LIMSI-CNRS (Paris) + RWTH (Aachen) + KIT (Karlsruhe) (IMMI-Labs)
 - Director : G. Adda, Vice Director : J.J. Mariani
 - Executive Board: A. Waibel, J.L. Gauvain, H. Ney
 - Launched in the framework of QUAERO program
- 10 people at IMMI / 160 people in IMMI-Labs on MMI processing
- Participation in EC projects FLaReNet and META-NET

- Department Informatics, Systems & Collaboration (ISC)
 - Applied Research and Technology Transfer: funding from research programs (international, national) & contract research (private, public)
 - Specialized in Visual Analytics: Data visualization, Data processing, Data provisioning, Software tools delivery
- 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 ISC: Calluna (advanced visualization) & Cadral (data mining)



Workplan

- WP1 - Management
- WP2 - Multimodal Person Annotation Components
- WP3 - Automatic Processing for Annotation
- WP4 - Collaborative Annotation



WP2 - Multimodal Person Annotation Components

Multimodal Person Annotation

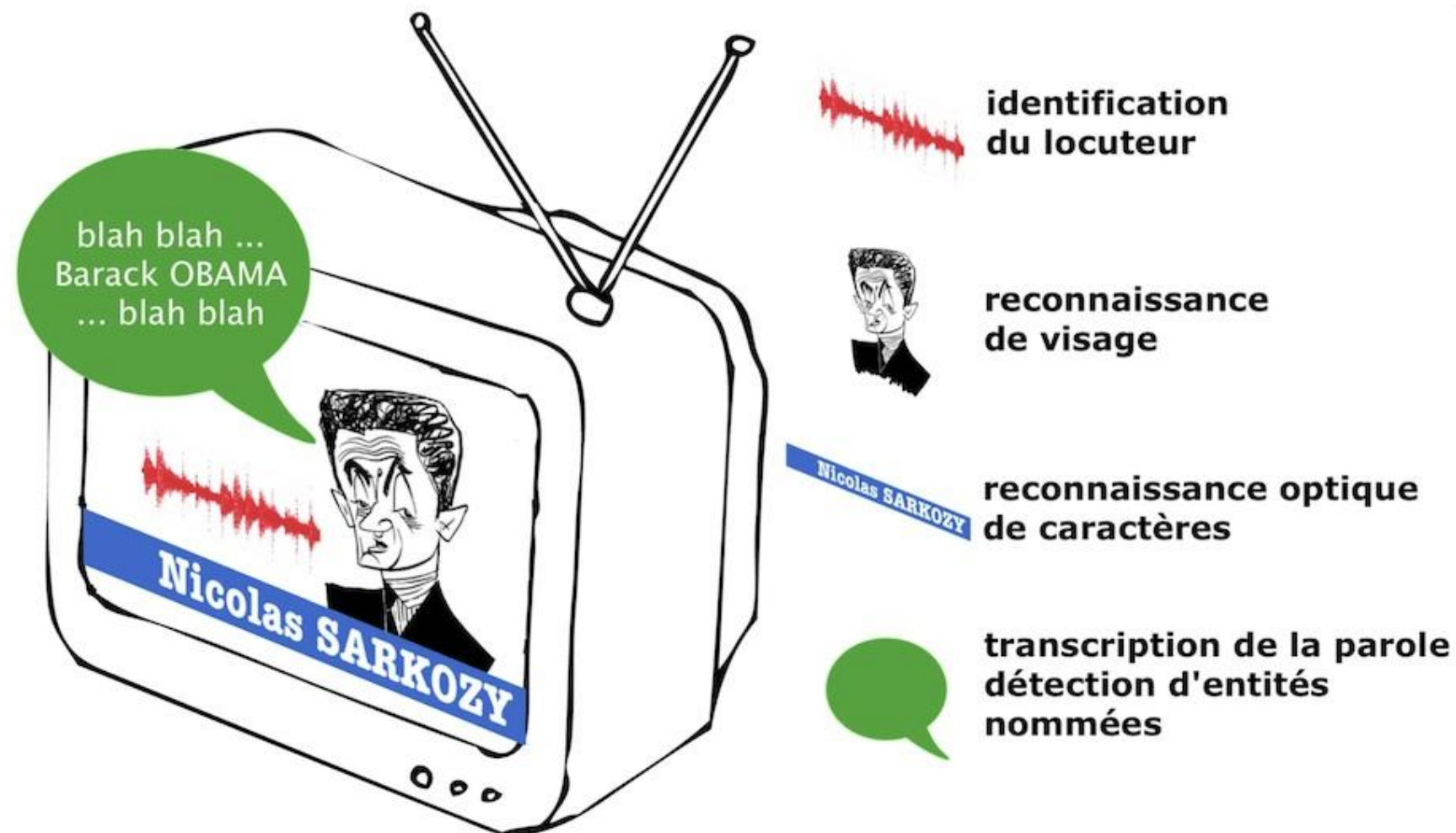
- Development of advanced identification components
 - Visual Person Annotation
 - locate, track and identify people in videos
 - Speaker Annotation in Audio
 - speaker segmentation, clustering and tracking system with acoustic-based speaker recognition
 - Fusion (early vs. late)
 - quite challenging, has additional issues: the person, who is visible may not be the one who is speaking, there might be many visible faces but only one is speaking -how to do the face-speech assignment, etc.
- The outputs of the technologies that have been developed within this WP will be used as input to the technologies that will be developed within WP3 -Automatic Processing for Annotation.

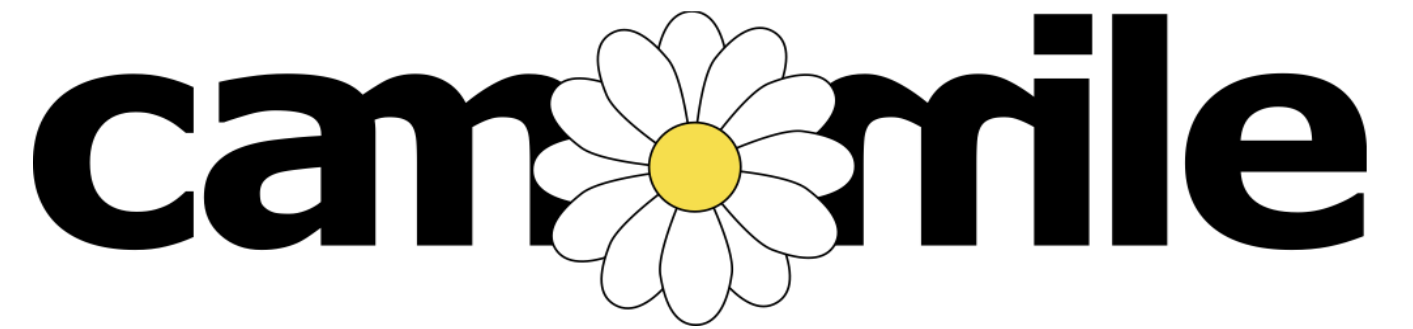
Work in Progress

- Developing a robust face detection and tracking system for videos
 - Current focus is to reduce the amount of false positives
- Improved facial feature localization for better face alignment and face matching across pose
- Investigating new face representation & classification schemes for better face matching (available system is based on DCT/SVM)
- Developed an i-vector based speaker verification system
- Address overlapping speech, cross-show diarization; integration of long-term prosodic features
- Classifier combination with kernelized eigenclassifiers

Assessment

- Participation to REPERE Challenge January 2014 evaluation
 - 10 hours of French TV broadcasts (news, talkshows) annotated by ELDA
 - Collaboration between QCOMPERE consortium (incl. LIG, LIMSI) and ITU



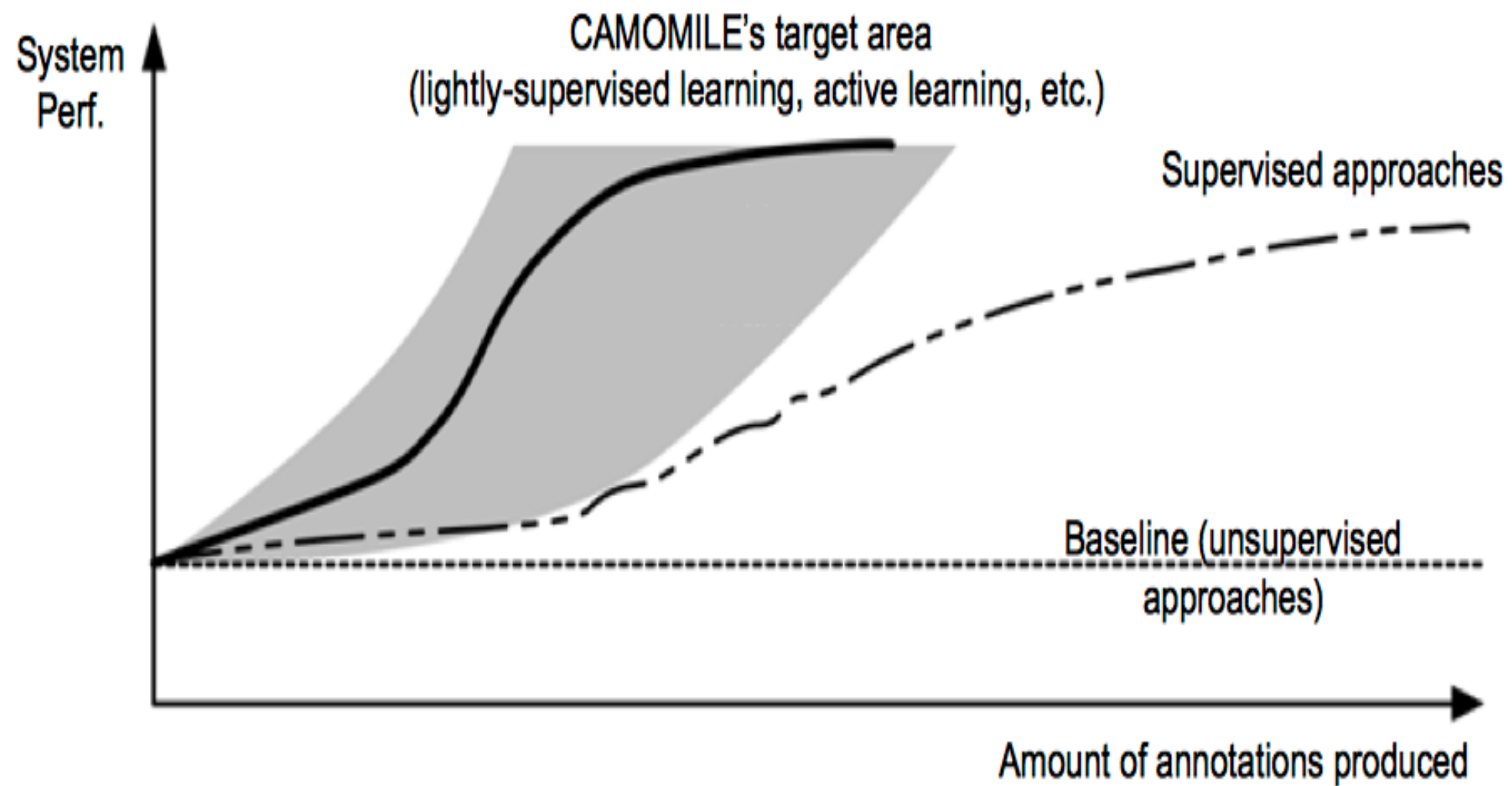


WP3 - Automatic Processing for Annotation

WP3 : Automatic Processing for Annotation

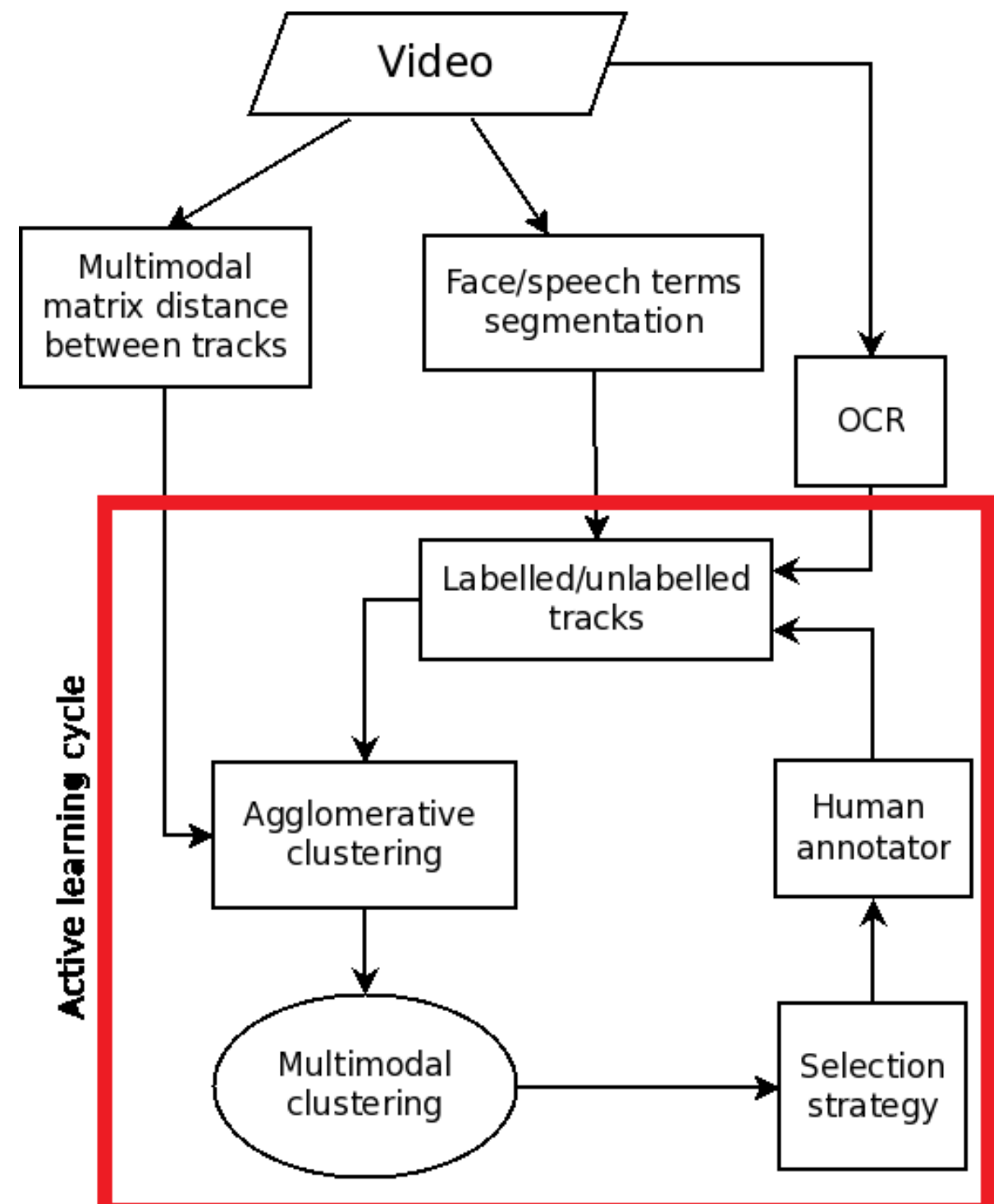
- **Machine-assisted annotation of multilingual / multimodal data...**
 - Initialisation of annotations, lightly-supervised learning, active learning, etc.
- **... on real scenarios where annotations are partial and noisy**
- **If you cannot measure it, then you cannot improve it !**
 - Metrics / tools to evaluate annotation quality (task-based) and productivity gains

WP3 : Big picture



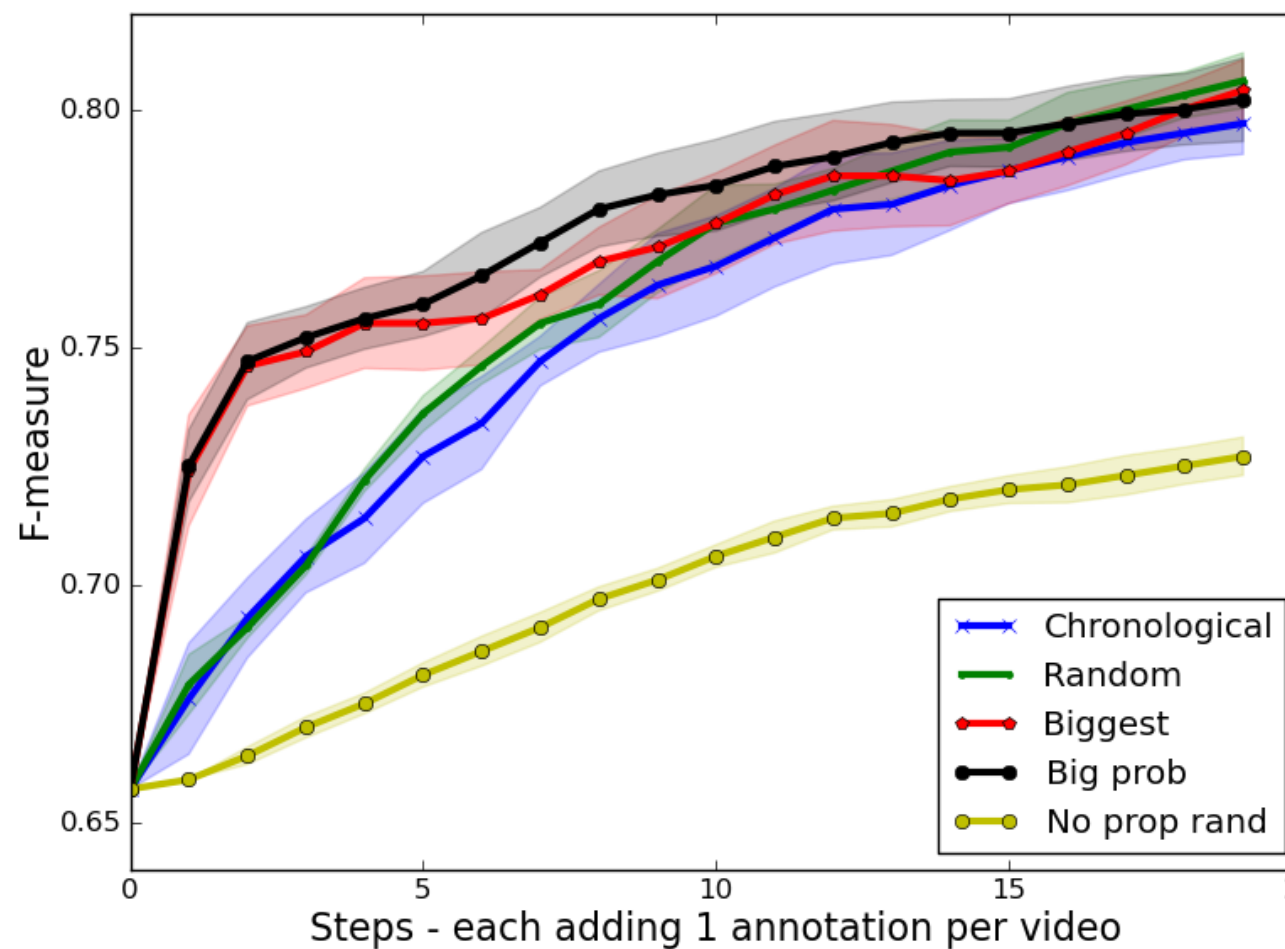
Ex. 1: Multimodal speaker recognition (REPERE)

- **Propagation** of manual annotations
- **Multimodal clusters** of speech and face tracks
- The **written name recognition** system
- **Active learning cycle** involving the human annotator

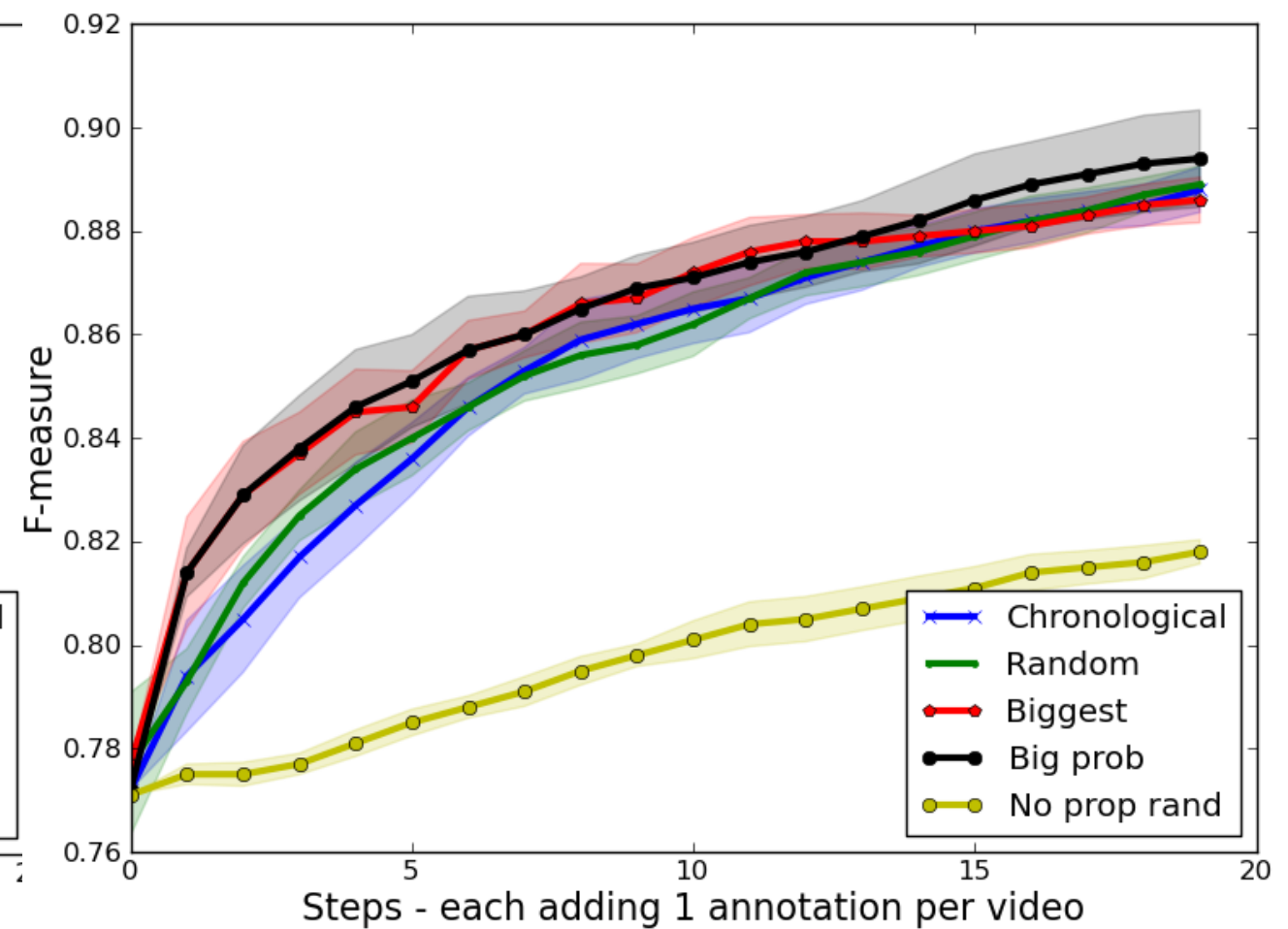


Ex. 1: Multimodal speaker recognition (REPERE)

Face annotation – face naming
score

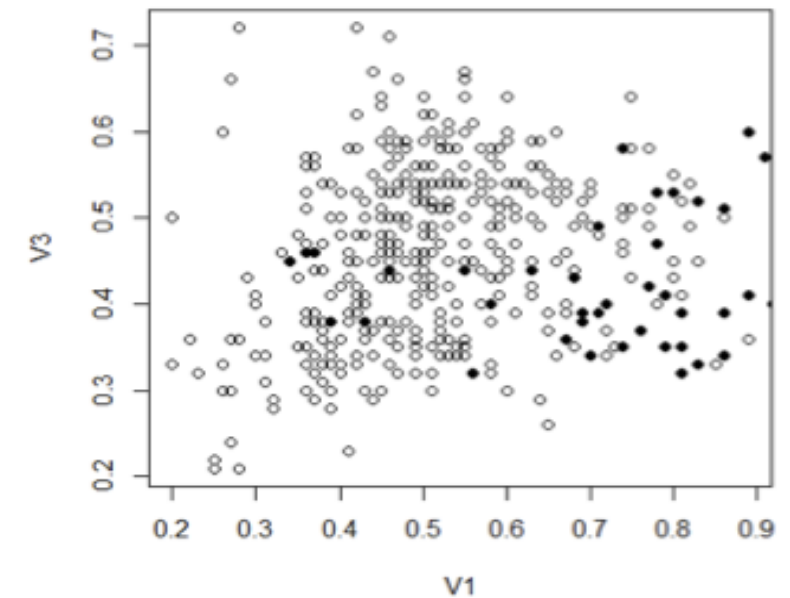


Speaker annotation – speaker
naming score

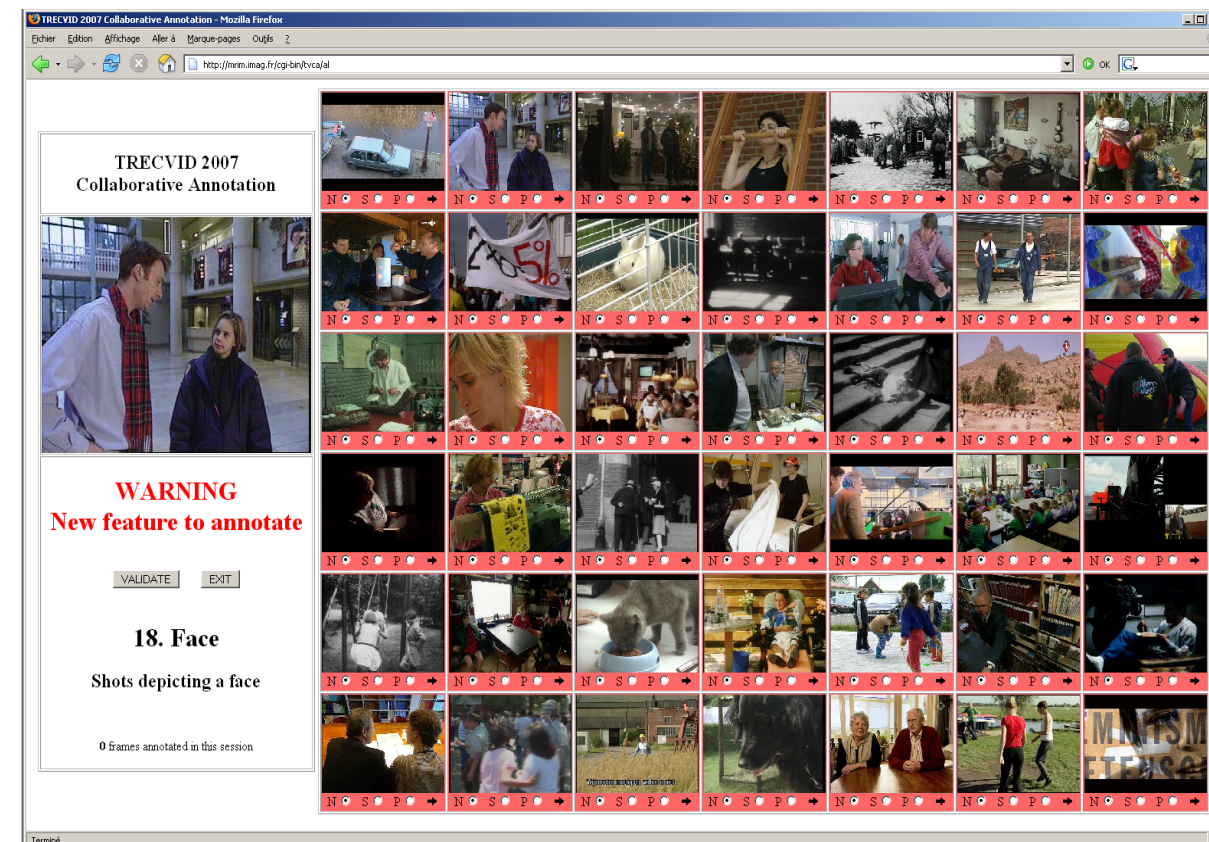


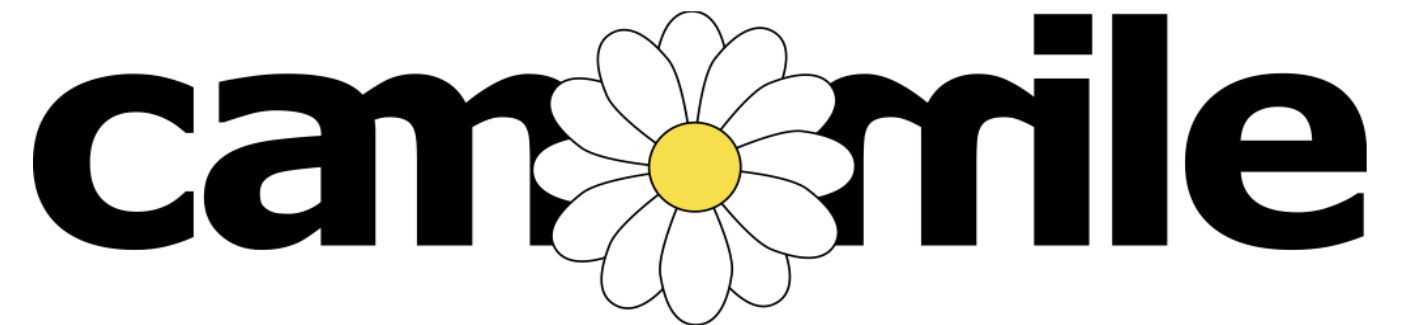
Ex. 2 : Semantic concept detection from videos (TRECVID)

- **Initialisation of binary classifiers**
 - Cold start
 - Un-balanced data sets
- **Annotation platform**
 - on-line, collaborative, measuring annotation time
- **Concept models retraining and evaluation**
 - Compare fully-supervised, un-supervised and lightly supervised approaches



(b) Imbalance ratio equal to 8.11

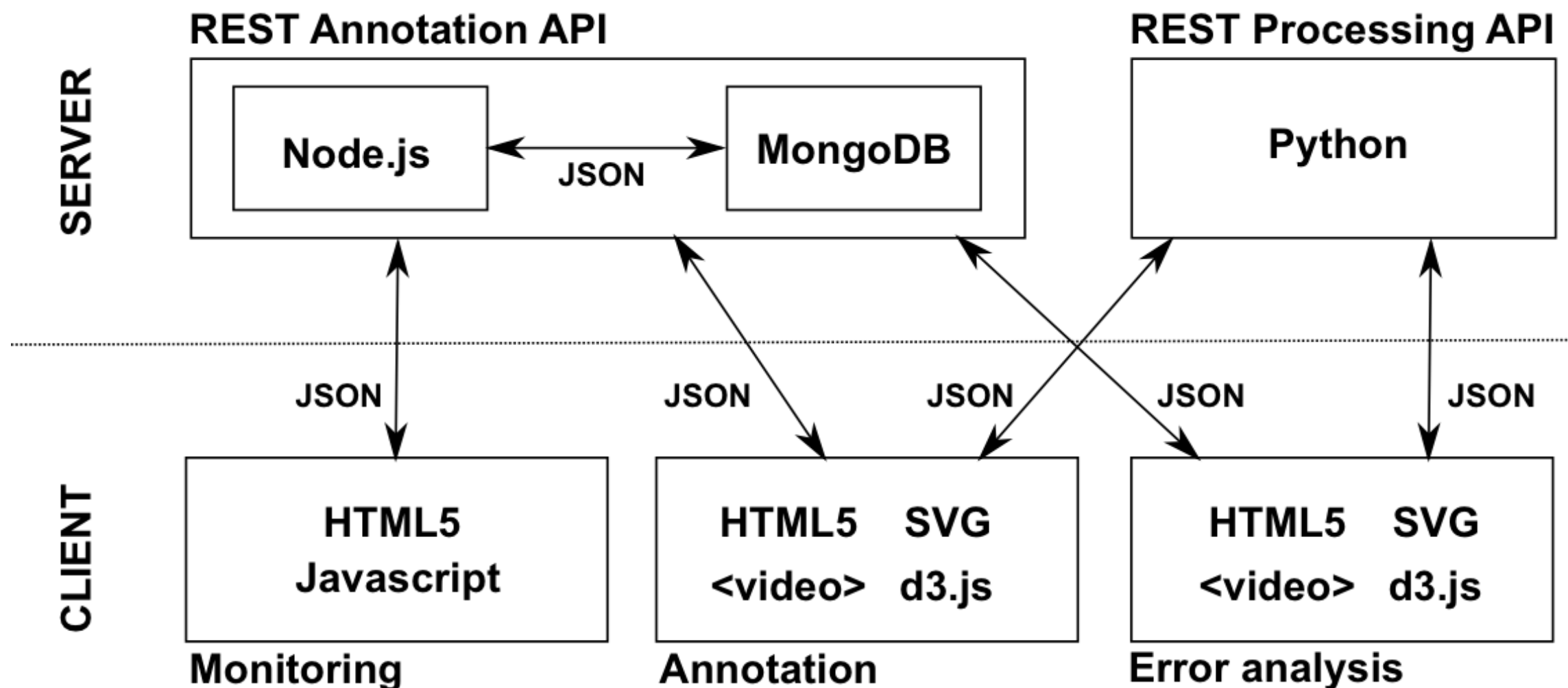




WP4 - Collaborative Annotation

Collaborative Annotation

- Framework for collaborative annotation
 - Interoperability through web services with controlled access to share and perform collaborative annotations

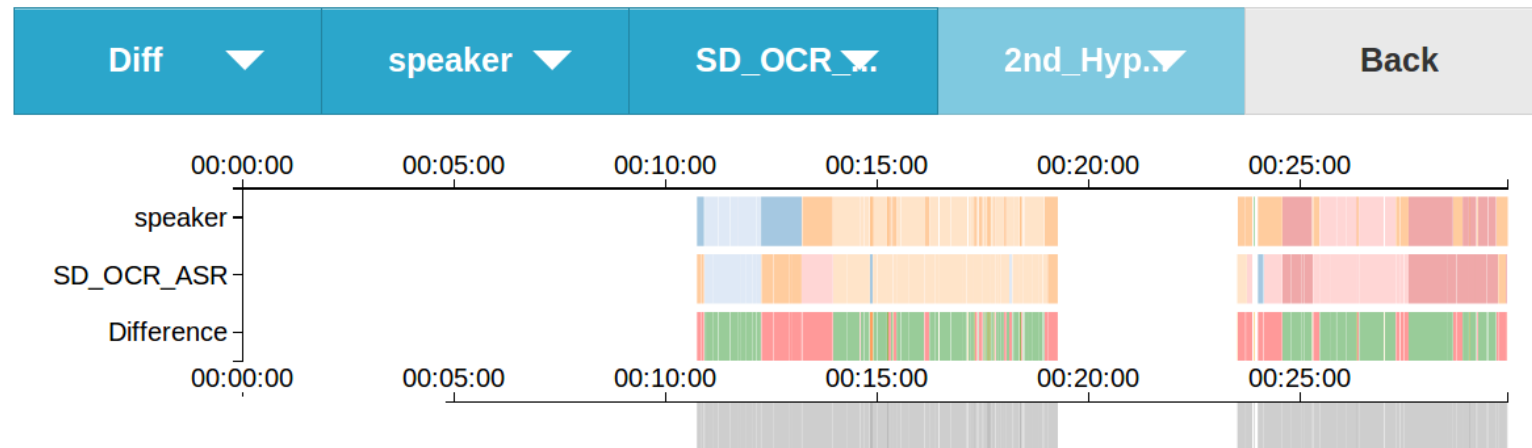


Collaborative Annotation

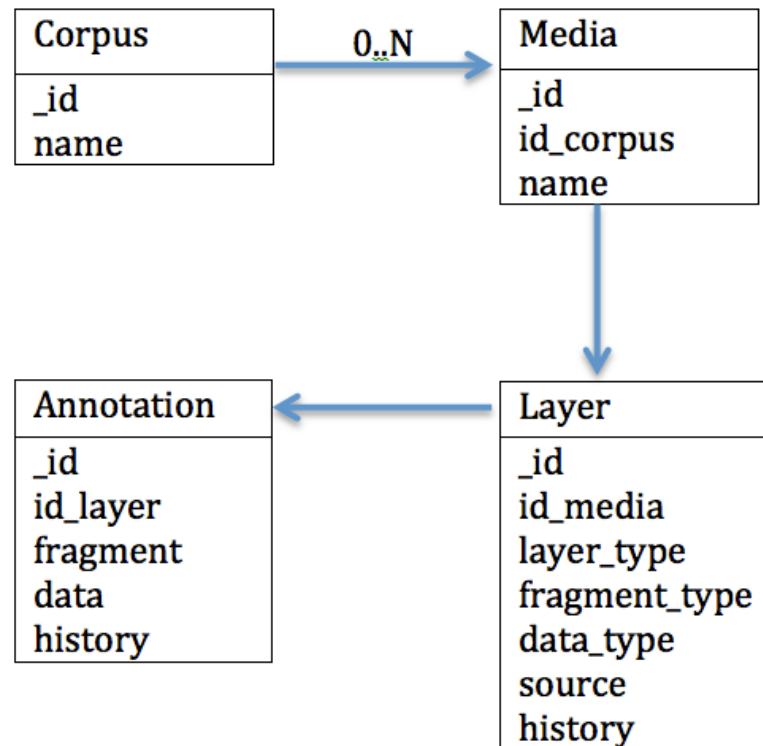
- Cooperative Annotation Monitoring
 - Design & implementation of web architecture + visual components
 - Integrates cooperative visualization research
 - Overviews and details on demand about history of contributions to annotation of MM/ML data
 - Various filtering features based on visual query
- Next Steps
 - Integrate error handling + develop collaborative aspects
 - Review and implementation of various scenarii related to other use cases

Annotation Platform Architecture

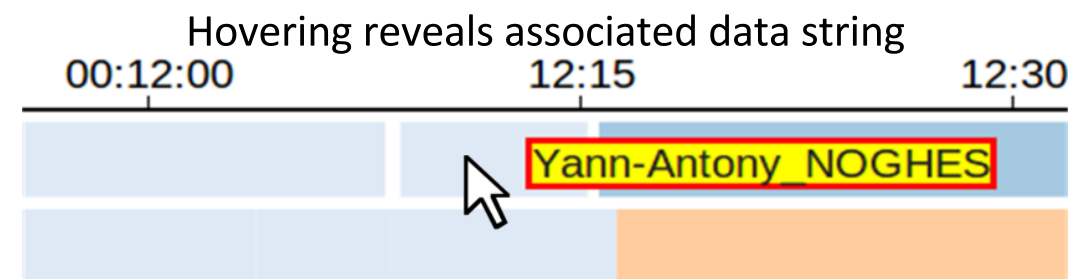
Full Timeline Component



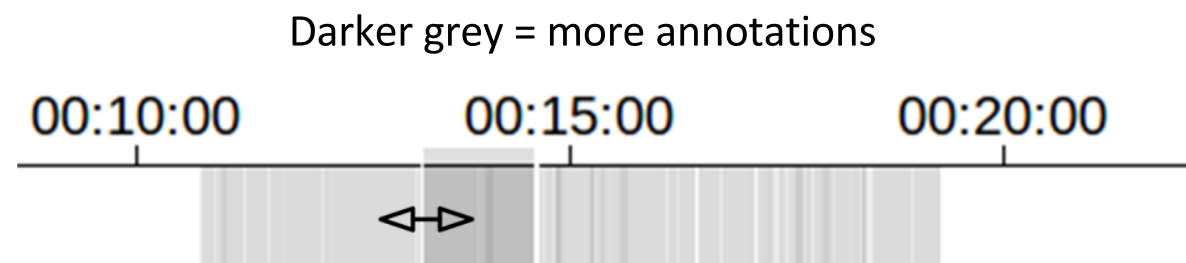
Application Data Model



Temporal Representation of Annotations



Aggregated View of Annotations

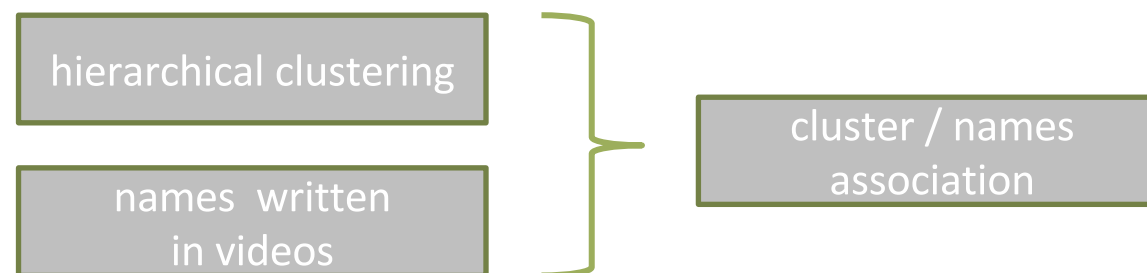


Annotation Platform

Typical Use Case

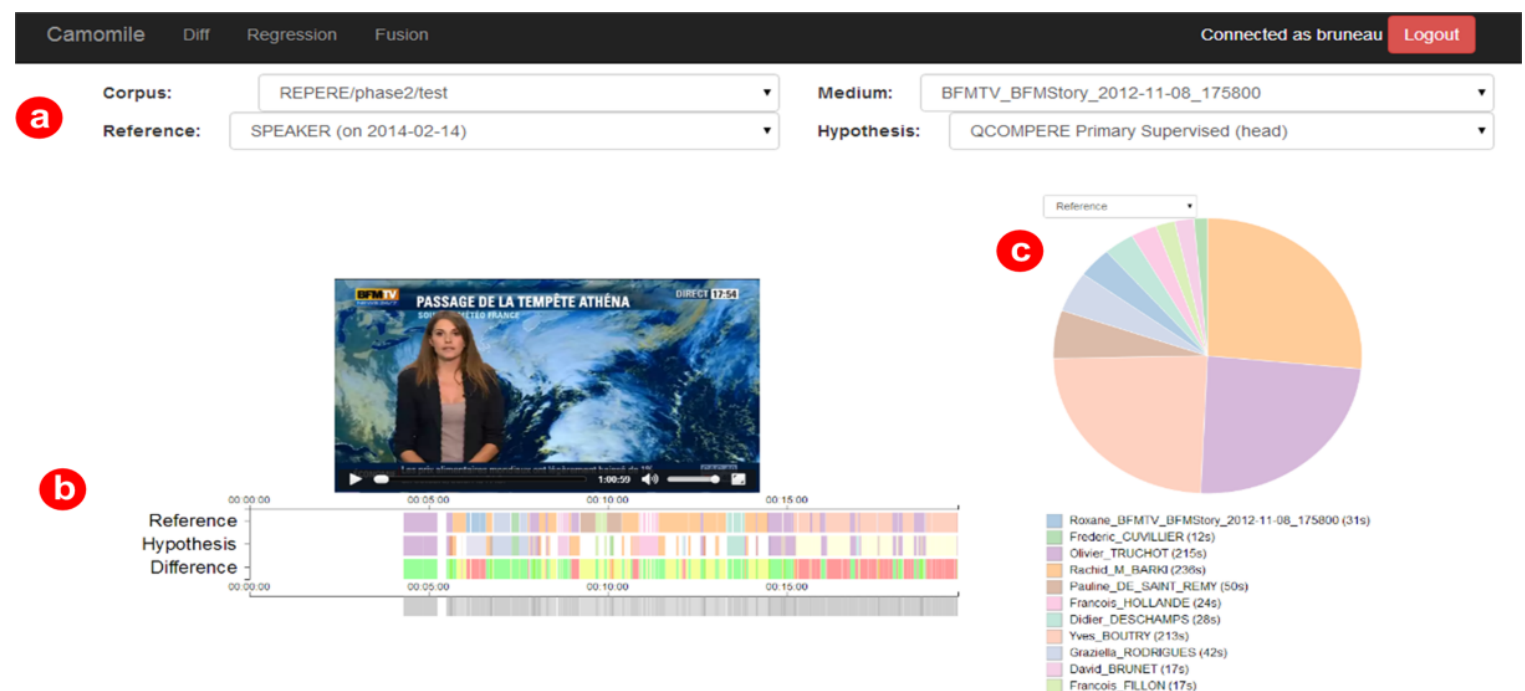
Step 1: Multimodal speaker identification in TV Broadcast

Speakers are identified in TV videos by propagation of written name in acoustic diarization



Step 2: Annotation platform

on-line,
collaborative,
with visual interaction,
temporal presentation,
data measurement,
...



Collaborative Annotation

- Annotation guidelines
 - Definition of annotation guides within a collaborative system
 - Specification of various layers of annotation, in terms of semantic content and formal structure
 - Conformity automatically checked by program
 - Metadata based on META-SHARE
 - Named Entities in a multilingual/multicultural environment
 - Attention to the granularity of annotation in order to lock/unlock small annotation chunks for people working in parallel

- Institutional guidance

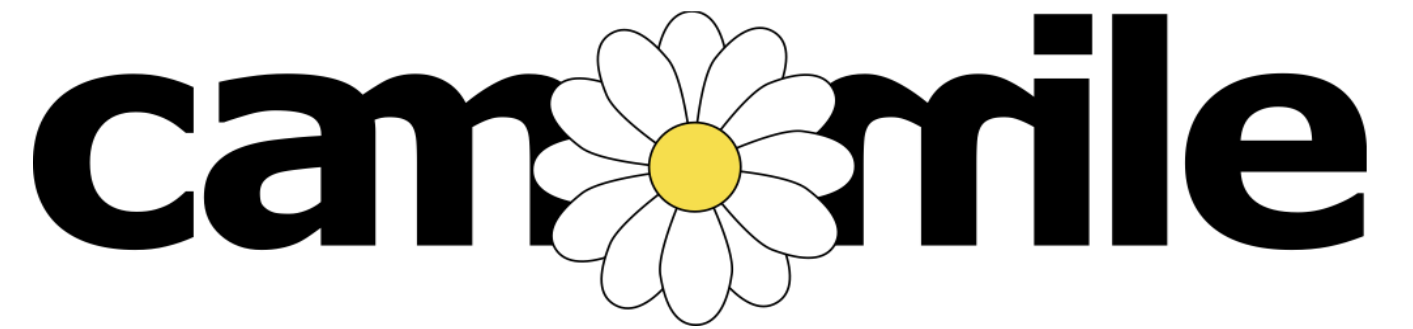
- Establish links with ELRA, LNE, LDC, META-SHARE, ISO,... and other projects of interest for CAMOMILE
- Tools and corpora should match needs and requirements of scientific community
- Presentation of annotation guidelines to potential institutional users
- Metadata presented to META-SHARE (META-DATA)
- Framework presented to META-SHARE for distribution / collaborative resource production / validation / evaluation
- Ethical dimension of LR production, including crowdsourcing
- **Set up International Advisory Committee**

International Advisory Committee



- CAMOMILE AdCom members
 - Khalid Choukri (ELDA/ELRA, France) (+META-SHARE, ISO)
 - Juliette Kahn (LNE, France)
 - Yves Raimond (BBC, UK)
 - Marcel Worring (Univ. of Amsterdam, The Netherlands)
 - Julia Hirschberg (Columbia Univ., USA)
 - Mark Liberman (UPenn and LDC, USA)
 - Liz Shriberg (SRI, USA)
- Observer and guest
 - Nicoletta Calzolari (CNR-ILC, Italy) (CHIST-ERA Observer)
 - Christopher Cieri (LDC, USA) (Guest)
- AdCom met at Ermenonville on November 20, 2013
 - Errare Workshop (21-22 November 2013)

- Agenda
 - Introduction and Presentation of the CAMOMILE project (CAMOMILE partners)
 - Presentations by AdCom members of actions (domestic or international) on M3 resources (collaborative) production, annotation, distribution etc. which may be of interest for Camomile
 - Feedback from AdCom members on CAMOMILE and general discussion
- Main Comments and Recommendations
 - CAMOMILE should establish and maintain contact with potentially interested (end-)users (such as BBC) and one way is to organize events / special sessions related to CAMOMILE in the future
 - CAMOMILE may aim at providing multiple tools (that have a maximum in common) widely adaptable for “multiple use cases” ?
 - “Propagation of annotations through the whole database is an important challenge / subject”



Management

Management

- Meetings

- Kick-off meeting (1 Feb. 2013, Paris)
- CHIST-ERA Project Seminar (27 March 2013, Brussels)
- Framework architecture meeting (5 May 2013, Luxembourg)
- Framework architecture meeting (17 Oct. 2013, Orsay)
- AdCom meeting (20 Nov. 2013, Ermenonville)
- Annual workshop (3-4 March 2014, Istanbul)
- Upcoming WP3-WP4 meeting (18 March 2014, Paris)

- Regular Visio/audio conferences

- Wiki and web site + mailing list + server for development

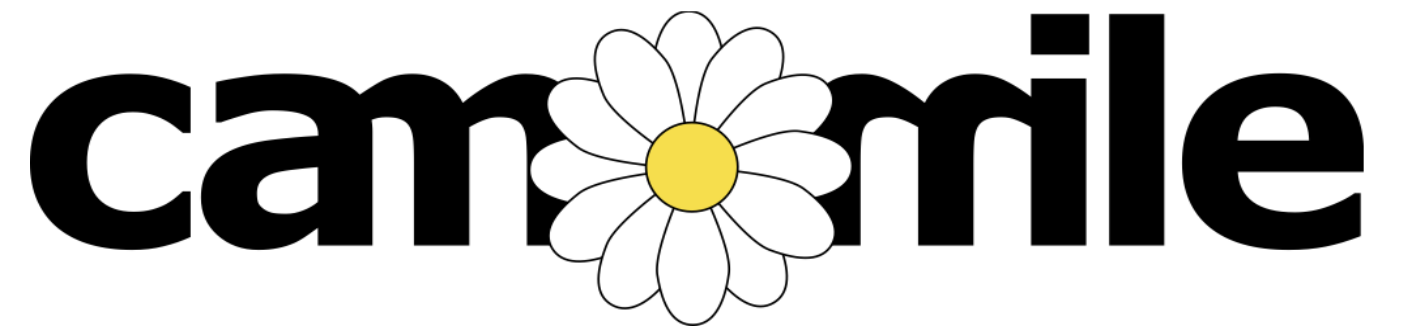
Planning issues

- Spanish funding started only in 2014
 - with partial cut in budget and duration
- Impact on the project not critical but significant
 - UPC mainly involved in WP2, but whole project slightly delayed (+ 6M)
 - WP2 activity was reduced in 1st year compared to plans
 - UPC could not participate to REPERE challenge in Jan. 2014
- Goal is to normalize the situation in 2014
 - UPC actively working in the project and hiring people
 - First version of WP2.1 and WP2.2 by end of March
and updated version incl. UPC contribution 6 months later

Management

- Financial Reporting

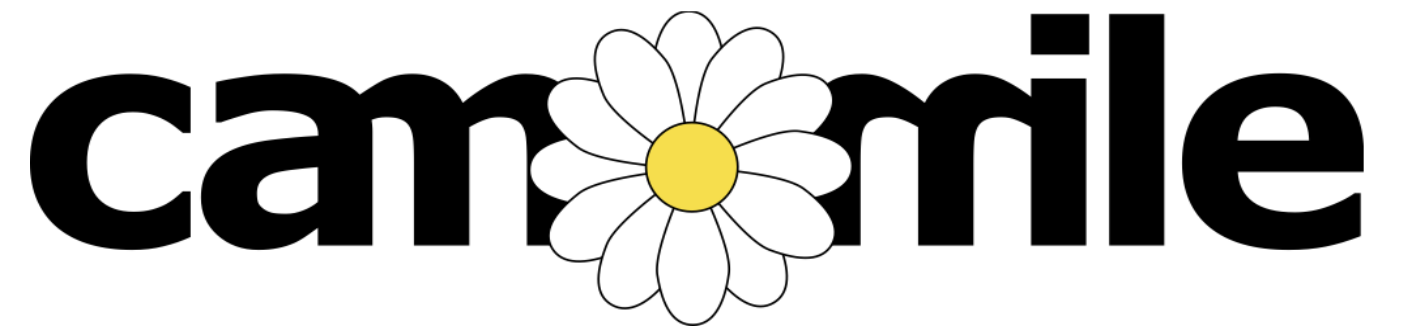
Partner	P.M	Cost	% budget
LIMSI	13	66k€	28%
ITU	47	22k€	37%
UPC	0	0	0%
LIG	17	70k€	28%
IMMI	6	34k€	28%
CRP-GL	6	46k€	19%



Scientific Results

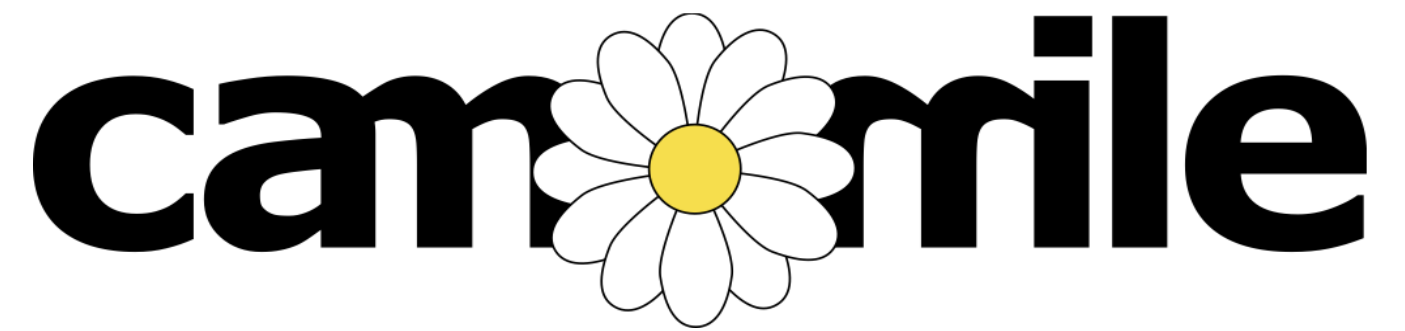
Dissemination

- **Conferences and workshops**
 - M. Turan, H.K. Ekenel, SIU 2013
 - U. Ekmekci, Z. Cataltepe, FUSION 2013
 - H. Bredin, P. Bruneau, A-P. Ta, C. Barras, Workshop Errare 2013
 - C. Barras, G. Adda, DGA workshop Multimedia Information Proc. 2013
 - M. Richter, H. Gao, H.K. Ekenel, IEEE WACV 2014
 - M. Budnik, J. Poignant, L. Besacier, G. Quenot, subm. to CBMI 2014
 - M. Budnik, internal project report
 - P. Bruneau, H. Bredin & al, subm. to IV2014
- **Co-organization of 1st Workshop ERRARE 2013**
 - on “Errors by Humans and Machines in multimedia, multimodal and multilingual data processing”
- **International collaboration**
 - 4-month stay of Ass. Prof. Satoshi Tamura (Gifu University, Japan) at LIMSI/IMMI on Audiovisual speech processing (09.2013-12.2013)



Sustainability

- Spin-off company from Dr. Ekenel on computer vision in multimedia
- Contacts with content providers: Digiturk, BBC...
- Possible participation to a REPERE follow-up
- Collaborative Annotation Framework to be released in Open Source



Questions?
