

Visualizing Multimodal Person Recognition Errors

a **REPERE** use case of
the **can****mile** project

Hervé BREDIN, Pierrick BRUNEAU, Anh-Phuong TA, Claude BARRAS

ERRARE workshop | Nov. 2013 | Ermenonville

Visualizing Multimodal Person Recognition Errors

- the **can****mile** project
 - partners
 - objectives
- the **REPERE** challenge
 - multimodal person recognition
 - fusion
- the **REPERE** use case of the **can****mile** project
 - collaborative annotation framework
 - demo: error analysis

the **cam** **mile** project

camomile.limsi.fr

the **cam** **mile** project

-  **chist-era** project

France	LIMSI , IMMI, LIG
Luxembourg	CRP-GL
Turkey	ITU
Spain	UPC

- **collaborative annotation**
of multi-**modal**, **multi-lingual** and multi-**media**
documents

the **cam** **mile** project

- WP2 / **multimodal person annotation components**
 - speaker identification
 - face recognition
 - multimodal fusion
- WP3 / automatic processing for annotation
 - machine-assisted annotations
 - lightly supervised learning
 - active learning
- WP4 / collaborative annotation
 - annotation guidelines
 - **web-based annotation framework**
 - web-based annotation monitoring

the **REPERE** challenge

www.defi-repere.fr

the **REPERE** challenge

- multimodal person recognition in TV shows

- three  projects
co-funded by 

QCOMPERE

SODA

PERCOL

LIMSI et al.

LIUM et al.

LIF et al.

- two evaluation campaigns in 2013 and 2014

- data collection & annotation by 

- evaluation by 
Le progrès, une passion à partager

the **REPERE** challenge



multimodal person recognition

- Multiple sources of information
 - **Audio stream**
Speaker diarization & identification
Speech transcription (ASR)
 - **Visual stream**
Face clustering & recognition
Optical character recognition (OCR)
 - **Text stream** (from ASR and OCR)
Named entity detection
Name normalization



multimodal person recognition

- Multiple sources of information **errors**

- **Audio stream**

- Speaker diarization & identification

IER \approx 30%

- Speech transcription (ASR)

WER \approx 20%

- **Visual stream**

- Face clustering & recognition

IER \approx 50%

- Optical character recognition (OCR)

WER \approx 10%

- How do these errors impact the overall performance?

Does adding a **cam^omile** use case always help?

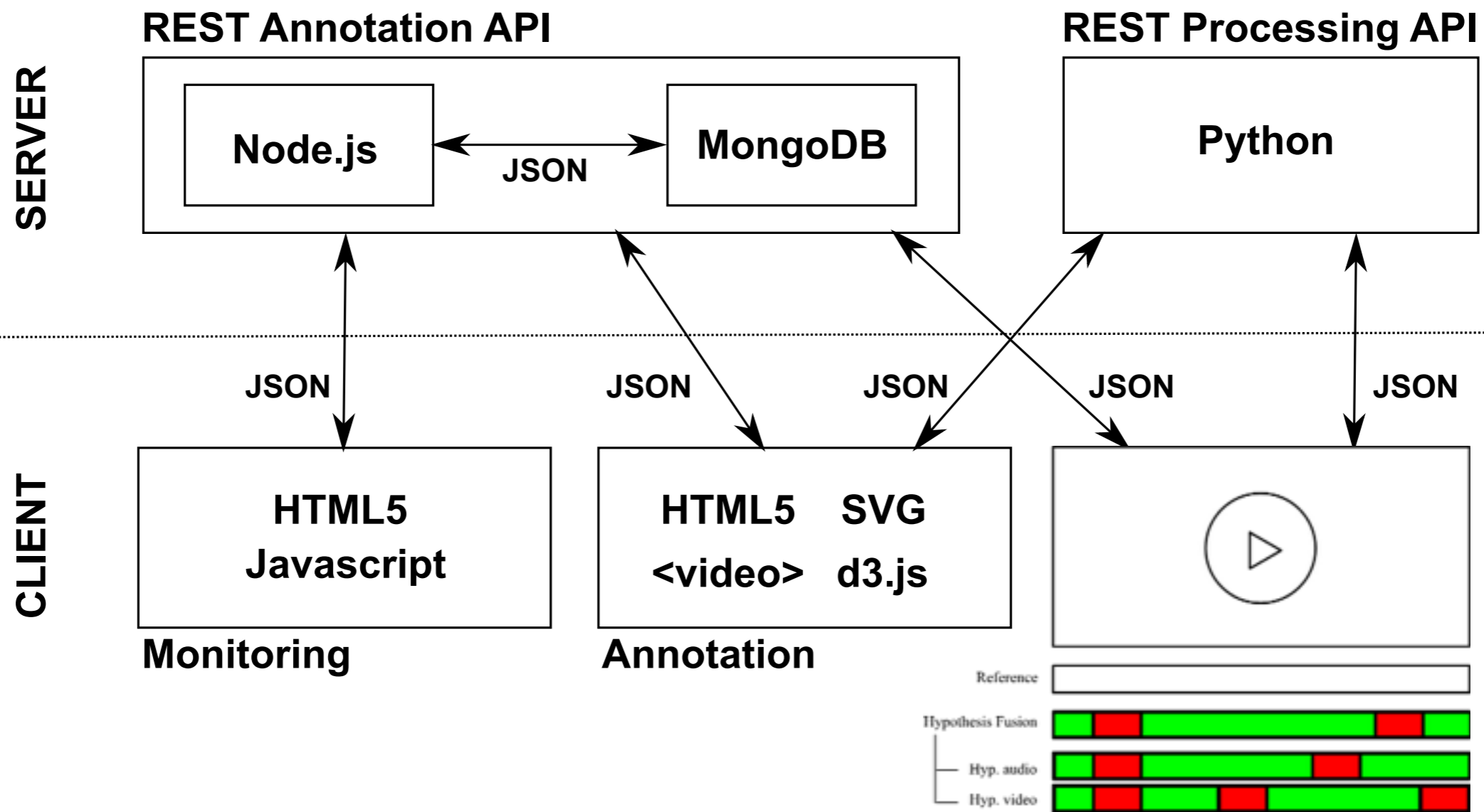
Why did the fusion system commit a particular error?

the **REPERE** use case
of the **cam****mile** project

error analysis for the researcher

- **/diff** use case
"where does my system make mistakes?"
visualize differences between reference and hypothesis
- **/regression** use case
"does [MyRevolutionaryIdea] consistently improve the system?"
visualize improvements brought by one system over a baseline
- **/fusion** use case
"why does my multimodal system make a particular error?"
jointly visualize multimodal system and its monomodal components

the framework



REST annotation API

corpus \supset **medium** \supset **layer** \supset **annotation**

/corpus/	POST	Create a new corpus and return its identifier :cid
	GET	Get list of corpora
/corpus/:cid/media/	POST	Create a new media in corpus :cid and get back its identifier :mid
	GET	Get list of media in corpus :cid
/corpus/:cid/media/:mid/layer/:lid		
	POST	Create a new layer for media :mid of corpus :cid
/corpus/:cid/media/:mid/layer/:lid/annotation/		
	POST	Create a new annotation in layer :lid of medium :mid in corpus :cid
	GET	Get all annotations contained in layer :lid
/corpus/:cid/media/:mid/layer/:lid/annotation/:aid		
	GET	Get annotation :aid in layer :lid, media :mid, corpus :cid
	PUT	Update annotation :aid in layer :lid, media :mid, corpus :cid
	DELETE	Remove annotation :aid in layer :lid, media :mid, corpus :cid

Table 1: A few HTTP requests supported by the REST Annotation API

JSON format

```
"modality": "speaker",  
"layer_type": "reference",
```

```
"fragment_type": "segment",
```

```
"data_type": "label",
```

```
"annotations": [  
  {  
    "fragment": {  
      "start": 2530.0,  
      "end": 2530.698  
    },  
    "data": "Olivier_TRUCHOT"  
  },  
]  
]
```

REST processing API

<code>/parser/:pid/</code>	POST	Upload file, apply parser <code>:pid</code> , and get file content in JSON format
<code>/metric/:mid/</code>	POST	Send reference and hypothesis, apply evaluation metric <code>:mid</code> , and get results in JSON format
<code>/error/diff/</code>	POST	Send reference and hypothesis, compute their differences, and get them back in JSON format
<code>/error/regression/</code>	POST	Send reference and two hypotheses, compute regression between them, and get them back in JSON format

Table 2: A few HTTP requests supported by the REST Processing API

dem 

conclusion n

- **a work in progress...**
 - integration with audio-video player
 - user management (access control list)
 - versioning of annotations
 - new visualization paradigms
- a **collaborative** framework for **annotation** of **multimodal**, **multilingual** and **multimedia** documents
- **open-source** by the end of the project

Visualizing Multimodal Person Recognition Errors

a **REPERE** use case of
the **cam^{smile}mile** project



cam^{smile}mile

Hervé BREDIN
Pierrick BRUNEAU
Anh-Phuong TA
Claude BARRAS