

# Joint Affective Sciences-IM2 Summer Institute

Riederalp/Aletsch, Sep.1-2, 2008

*Prof. Hervé Bourlard  
Idiap Research Institute*



# Welcome to Riederal



- All necessary information and technical program available from  
<http://www.im2.ch/summer-institute-2008>
  - Talks
  - Parallel GEKO (Get to Know each Other) session
  - Posters
  - Social activities
- **One request:** GEKO chairmen should provide me with a couple of slides summarizing their session

- **Goals:**

- Intra-NCCR training and collaboration
- Inter-NCCR training and collaboration
  - GEKO...
  - Better understand possible synergies
  - Develop joint projects

- **Thank you to:**

- Sarah Favre
- Dr. François Foglia
- Ms. Valérie Devanthéry

# Interactive Multimodal Information Management (IM2)

Swiss National Center of Competence in Research (NCCR)

*Prof. Hervé Bourlard*

*IM2, Director*

*Idiap Research Institute  
(Leading House)*

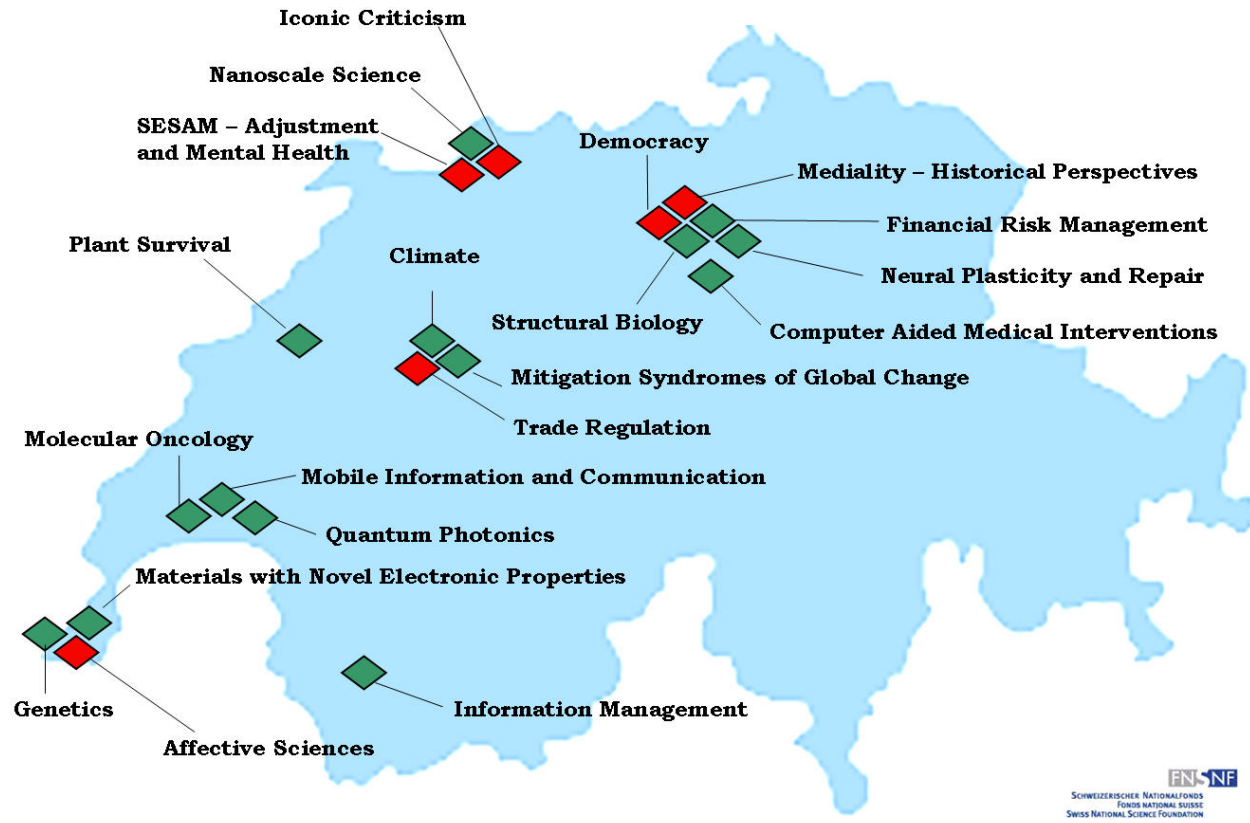


The **National Centers of Competence in Research** are a research instrument of the National Science Foundation.



# 20 NCCRs (National Centres of Competence in Research)

## National Centres of Competence in Research (NCCR)



# IM2 R&D Vision

- General Vision:
  - Structuring of **multimedia information systems**
  - Advanced **multimodal human computer interfaces**
- Specific Vision:
  - Analysis, understanding, modeling and enhancement of human-human communication scenes

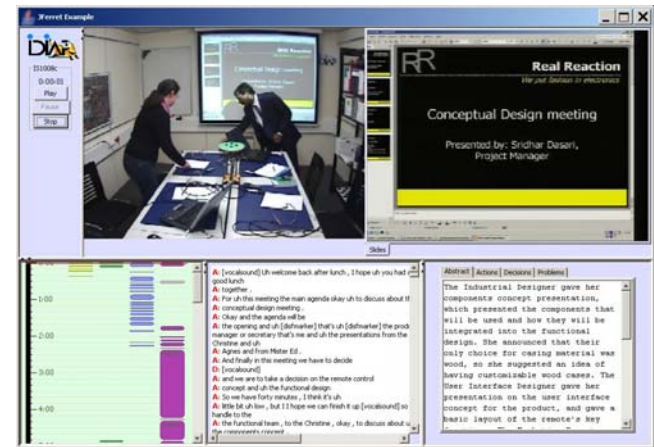
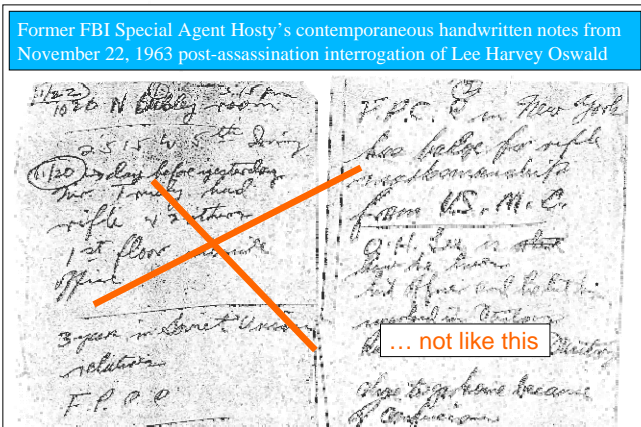
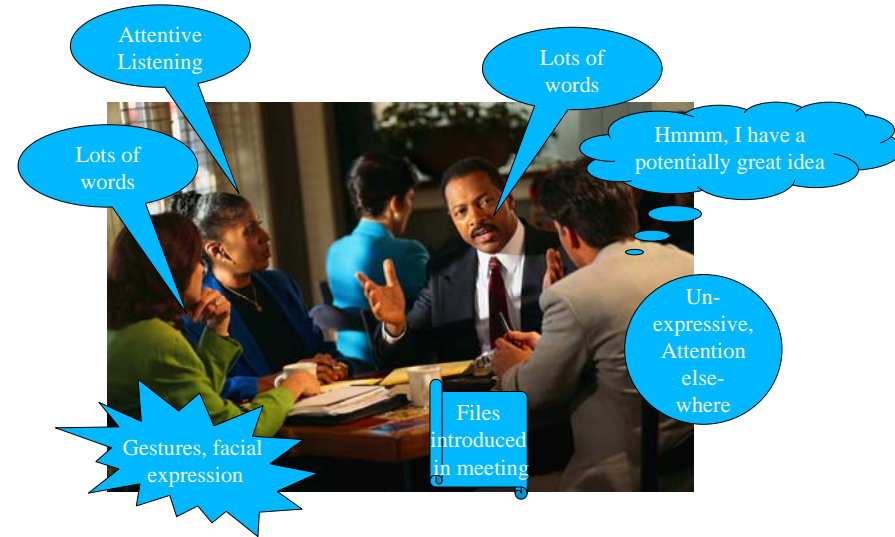




# One specific focus:

## Human-Human Communication in Meetings

- All important decisions are taken in (face-to-face) meetings.
- However, this results in many meetings
- >12 million business meetings daily!!!!
- Expensive and often not as efficient as expected.





# IM2 Goals

---

- **Research**: Highest quality, and most competitive research at the world level
- **Integration**: Integration in national and international projects
- **Training**: Enhancing teaching, training, and knowledge transfer
- **Technology transfer**

# General Figures

---

- Home Institution: Idiap Research Institute, Martigny
  - Partners: EPFL, UniGe, Univ. Fribourg, ETHZ, and Univ. Bern
  - Started on January 1, 2002
  - Expected duration: 3 x 4 years (3 phases)
  - Funding/budgets:
    - Phase 1:
      - SNFS funding: 15'349'000.- CHF
      - Self & third-party funding: 19'655'000.- CHF
    - Phase 2:
      - Budget: 12'000'000.- CHF (SNSF) + 14'000'000.- (self & third party)
    - Phase 3 (2010-2012): pre-proposal end of this year
-

# YOU are part of a group of...

---

## About:

- 8 Project Leaders
  - 38 Senior reseachers
  - 21 Postdocs
  - 81 PhD students
  - 2 MSc students
  - 2 Diploma students
  - 36 Management, development, others
-

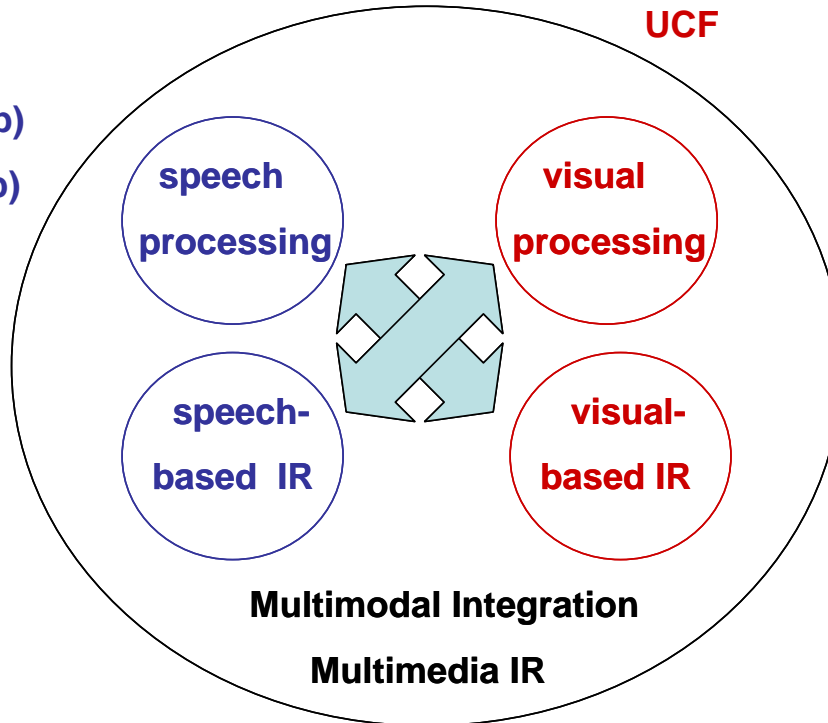
# Strong EU and US leveraging

---

- EU-FP6 and FP7 projects:
  - STREP (FP6): M4 (Multimodal Meeting Management)
  - IP (FP6): AMI (Augmented Multi-party Integration)
  - IP (FP6): AMIDA (Augmented Multi-party Integration with Distance Access)
  - IP (FP7): TA2 (Together Anytime, Together Anywhere)
  - NoE: PetaMedia
  - SSPNet: Social Signal processing, NoE coordinated by Idiap
- US projects:
  - DARPA GALE
  - DTO VACE: RoadMap

# A new, very active, research domain

Meeting Recorder (3p)  
Mapping Meetings (4p)  
NIST



**FX-Japan**  
**UCF**

EU-AMI/AMIDA (13p)  
EU-CHIL (15p)  
CH-IM2 (11 p)

US-VACE (5p)  
VACE-III (7p)  
US-CALO (CAMEO)  
- CMU

MIT  
UCI (Rescue)

Microsoft  
FXPAL  
Ricoh  
NTT  
ATR

# IM2 Multi-disciplinary

---

Meetings provide a realistic, yet circumscribed arena to address multiple multi-disciplinary research, including:

- Scene analysis
  - Unconstrained speech recognition
  - Model of individual and group dynamics
  - Sociology and social-psychology
  - Structure, index, summarize communication scenes
  - User interfaces
-



# Exploiting the nature of group interaction

## Multimodal Processing

- + Speech and audio
- + Non-verbal cues from video and audio
- + Cognitive psychology (emotions)
- + Attention focus, postures, expressions
- + Subjective content in conversations
- + Complementary multimodal

## Signal Processing

- + Computer vision
- + Audio processing
- + A/V fusion



## Social Psychology

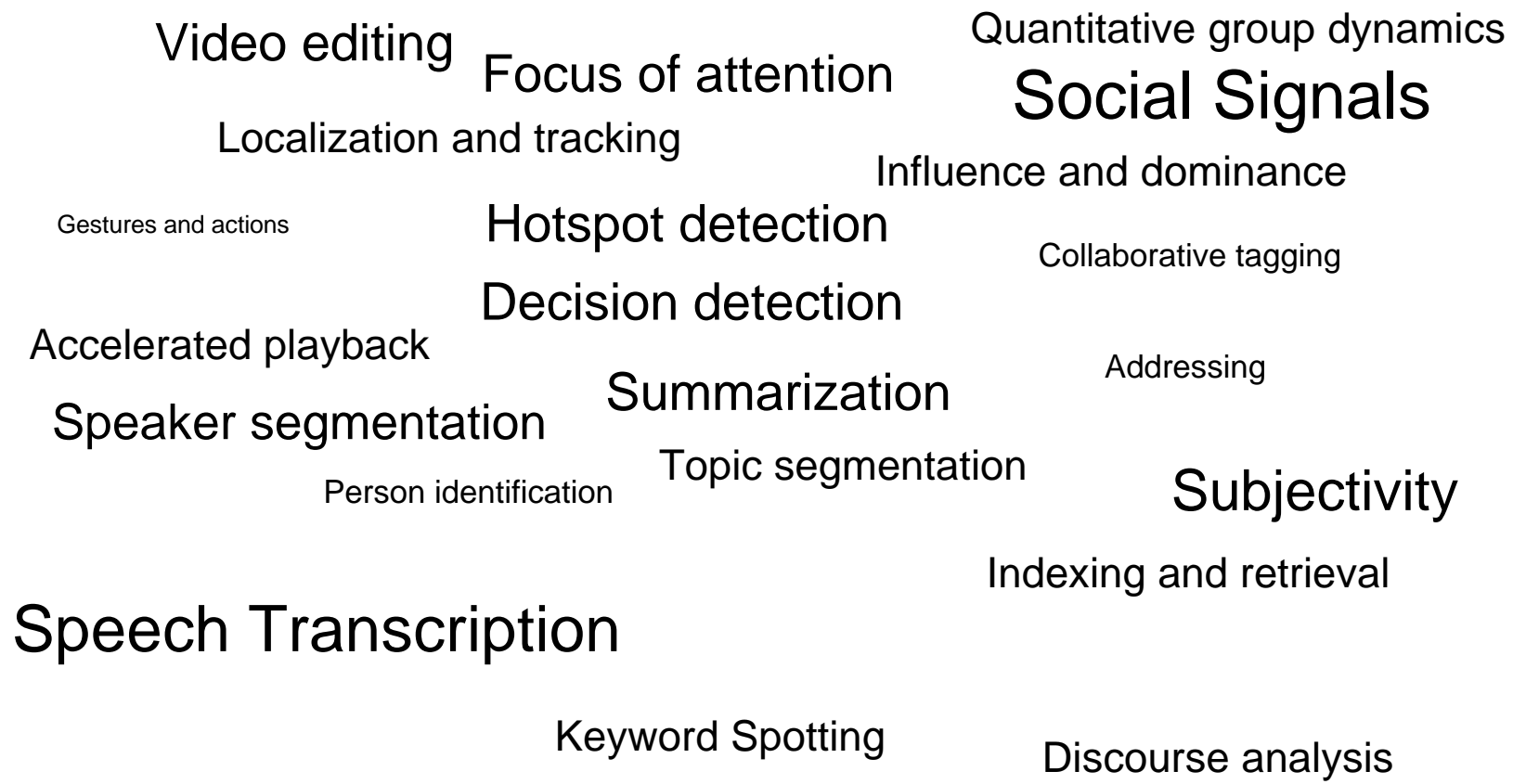
- + Human behaviour
- + Social signaling
- + Social networks
- + Contextual environment (adaptation)
- + Cross-cultural factors

## Multiparty Collaboration

- + Dialog understanding
- + Behaviour constrained
- + Role constrained by group
- + Group size matters
- + Complementary multiparty cues

# IM2 Core Technologies

---

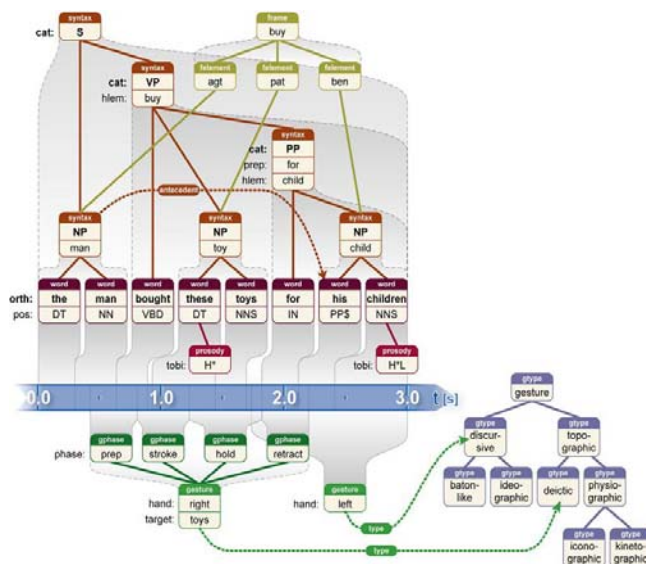


# IM2 Common Framework Instrumented Meeting Room



# Multimodal meeting database

- Large multimodal database of more than 100 hours of meetings
- Annotated in terms of:
  - Audio (checked) transcription
  - Named entities
  - Dialogue acts
  - Topic segmentation
  - Extractive and abstractive summaries
  - Hand gestures
  - (limited) Head gestures
  - Location of person on video
  - (limited) gaze direction
  - Movement around room
- Available to the community through multimedia file server  
<http://mmm.idiap.ch>
- Enriched by many other (IM2 and other) multimedia databases, including multimedia lecture recordings.



# IM2 Key Achievements (1)

---

- 1. Meetings rooms (IDIAP, IniFri and ICSI) and lecture rooms (IDIAP and UniFri)**
  - 2. Collection and full annotation** of large amounts (100 hours) of multimodal meeting recordings (<http://mmm.idiap.ch>)
  - 3. Mono-modal signal processing:** Major progresses in scene analysis, speaker segmentation and tracking, and large vocabulary speech recognition.
  - 4. Multimodal data processing:** Development and large scale evaluation of original approaches including multi-channel processing, multimodal group activity segmentation/recognition, etc.
-



# IM2 Key Achievements (2)

---

## 5. Multimodal information search and retrieval:

- Acquisition and automatic structuring of multimodal lecture data (<http://mmm.idiap.ch>)
- Generic multimodal fusion mechanism for search and retrieval of multimedia data (<http://viper.unige.ch/vicode>)

## 6. Multimedia document content abstraction:

- A number of new techniques for **automatic discourse processing** (e.g., decision-tree based method for detecting discourse markers based on lexical and prosodic information).
- Lectures acquisition and structuring (ACM)
- Leading edge in the extraction and exploitation of audio-video social signaling (e.g., in the processing of broadcast debates).



# IM2 Key Achievements (3)

---

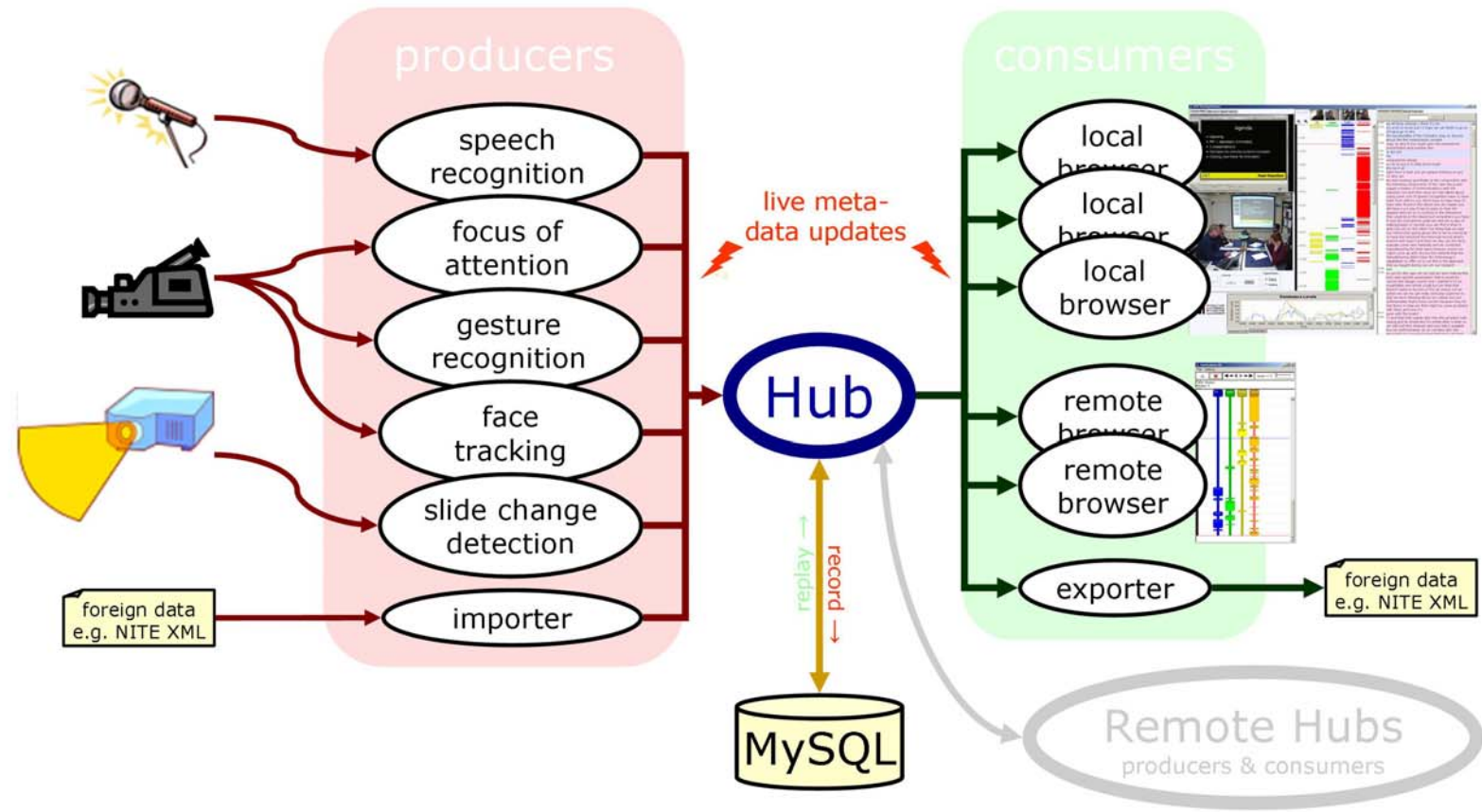
## 7. HMI formal usability and utility testing:

- User requirements
- Multiple meeting browsers
- BET (Browser Evaluation Test)

## 8. Multiple (integrated) demonstration systems

- Many based on common software infrastructure (JFerret), allowing for easy integration of multiple plugins, interface design, etc.
- Online content linking system
- Remote user engagement
- Virtual and physical showroom.... **Much more in progress!!!!**

# Integration: The “Hub”



→ benefits: real-time, distributed, subscription-based, extensible

# Online Content Linking

- Motivation
  - participants in a meeting often mention documents containing facts that are currently discussed
  - **but** they do not have the time to search for the facts
- Objective
  - display automatically the documents (from an archive including past meetings) that are potentially relevant to a discussion
- Main components
  - Query Aggregator: builds queries from ASR, searches over document archive, filters results
  - User Interface
  - the Hub: subscription-based data exchange architecture

The screenshot shows the AMIDA Content Linking Demo interface. It features a search bar on the left, a list of search results in the center, and a video thumbnail on the right. A red box highlights the interface, with arrows pointing to specific elements: 'Keywords detected from speech' points to the search bar, 'Names of relevant documents per 30'' interval, with access to full content' points to the search results list, and 'Access to previous meeting' points to the video thumbnail.

# Engaging Remotely














- Online aid for remote meetings
- Indicate who is looking at whom
- Foci of attention
- Presented slide

# IM2 Intl. Evaluation Activities

In addition to participating in multiple international evaluation campaigns, e.g., in biometric authentication, EEG processing, information retrieval, etc:

- *ASR: Automatic speech recognition*
- *KWS: keyword spotting*
- *SEG: speaker segmentation*
- *ID/LOC: identification and localization/tracking*
- *FOA: focus of attention*
- *GAA: gesture and action recognition*

• *IM2 was also among the first ones to propose methods to objectively evaluate usability and utility of (multimodal) user interfaces: BET*

	Internal Evaluation	External Evaluation	Contributing Data
ASR	✓	NIST  	✓
KWS	✓	NIST  	
SEG	✓	NIST  	✓
ID/LOC	✓	VACE (CLEAR)  	✓
FOA	✓	VACE-III (RoadMap)  	✓
GAA	✓		

- **More recently: IDIAP Ranked 1st on the *ImageCLEF 2007 medical annotation task***
- **Again in 2008!!!!**

<http://www-i6.informatik.rwth-aachen.de/~deselaers/imageclef07/medaat-results.html>

**Enjoy and take full benefit of  
this unique workshop!**