

# Recognition and Understanding of Meetings and Lectures

## EU AMI and AMIDA Projects

---

***Prof. Hervé Bourlard***  
***IDIAP Research Institute***



***CHORUS Final Conference***  
***Brussels, 26-27 May 1009***

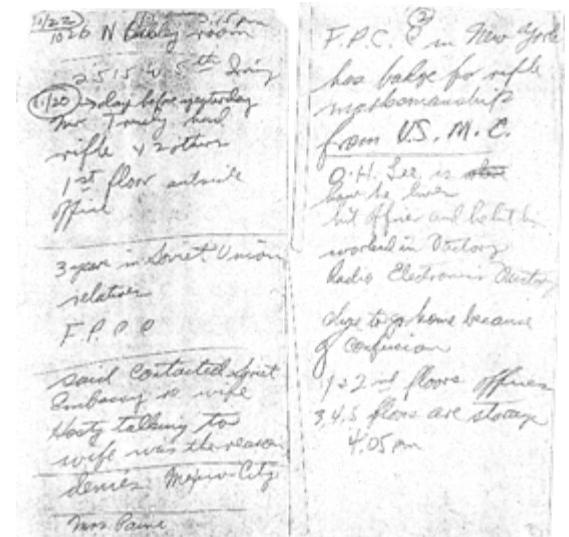


# Meetings, meetings, meetings,...

- All important decisions are still taken in (face-to-face) meetings
- Which results in many (ever increasing) meetings
  - >12 million business meetings daily (only in the US)!!!
- Expensive, typically:
  - A few person-weeks of meeting
  - 3,000 euros travel budget/person/meeting
  - 24 days away from family and friends
  - 24 days backlog of work
  - Evenings spent in transit at Schiphol
- Often (usually) not as efficient as expected!!!!



Former FBI Special Agent Hosty's contemporaneous handwritten notes from November 22, 1962 post assassination interrogation of Lee Harvey Oswald.



# AMI Consortium

- An 12-member **multi-disciplinary** consortium dedicated to research and the development of technologies that will **enhance (Augment) Multiparty Interactions**



The University  
Of  
Sheffield.

PHILIPS

Noldus  
Information Technology

# Areas of focus

---

- Real-time team meeting dynamics
- Automatic meeting content indexing and viewing
- Data collaboration and/or consensus building
- Content management (publishing, indexing and repurposing of pre-recorded meetings)
- Knowledge management (mining/extracting information about and from meetings)
- Consulting about improvements in meetings

# AMI/AMIDA funding

- The European Commission
  - Framework Programme 6
- Corporate sponsorships and contributions
- Augmented Multiparty Interaction (AMI) Project
  - Jan 2004-Dec 2006
- AMIDA project = AMI+Distance Access
  - Jan 2007-Dec 2009
- Total approx 30M Euros, 50 FTEs

# AMI and AMIDA

Area of focus	AMI Project	AMIDA Project
Real-time meetings	Human-human, face-to-face	Same but in remote meetings
Automatic meeting capture	Automatic indexing/tags	Same with remote meetings
Data collaboration	Low emphasis	Higher
Content mgt	High emphasis	Equal
Knowledge mgt	Low emphasis	Higher
Consulting	High emphasis	Equal

# Highly multi-disciplinary

## Multimodal Processing

- + Speech, audio and video processing
- + 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 signal processing
- + Social networks
- + Contextual environment (adaptation)
- + Cross-cultural factors



## Multiparty Collaboration

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



# Technologies required for improved (remote) meetings

- Technology to create “presence”
  - Audio and video systems, network bandwidth, room arrangement, shared workspaces (“whiteboards”), etc
  - Example: Instrumented meeting rooms, User Engagement and Floor Control
- Technology to create “archives”
  - Audio and video-taped recordings, archived handouts and slide presentations, automated meeting notes
  - Example: Meeting browser
- Technology to create “context”
  - Automatic meeting moderator to append related material from group input or other sources, including URLs and archive files
  - Example: Automatic Content Linking Device



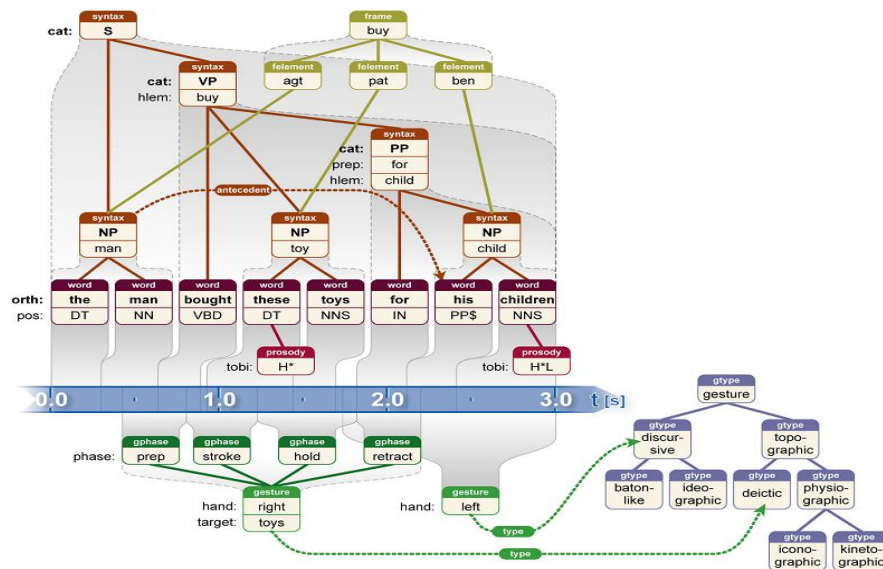
# Instrumented Meeting Rooms

- All media synchronized:
  - Close-up and wide angle cameras
  - Microphones (far-field, close-talking)
  - PC projector, I/O pens, white board
- Instrumented Meeting Rooms complemented *with audio and video conferencing*

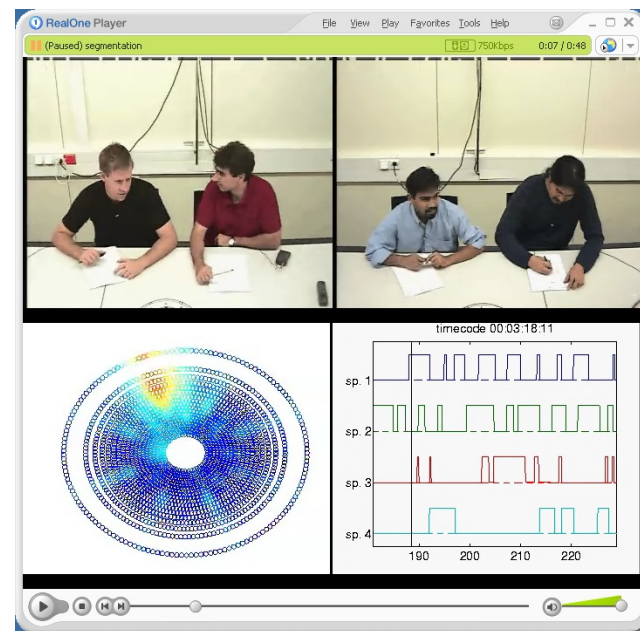


# AMI Multimodal Meeting Database

- 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> (DVD taster available on request)
- Enriched by many multimedia databases, including multimedia lecture recordings.

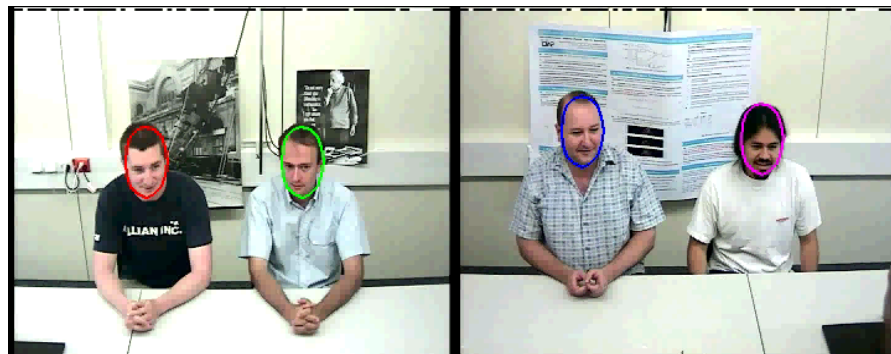


- Speech/non-speech detection
- Speaker turn detection:
  - Based on acoustic features
  - Based on sound source localization (mic array)
  - Based on both (fusion)
- Speaker segregation
- Speaker identification
- Conversational speech recognition
- Extraction of audio metadata, dialog acts, hesitations, etc



# Computer Vision in Meetings

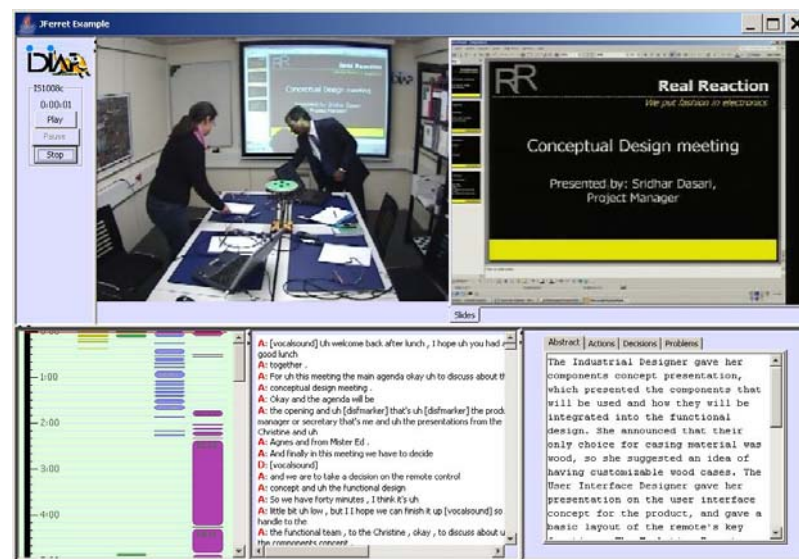
- Active audio-visual shape tracking (multi-camera, multi-person tracking)
- Face and body tracking
- Face identification
- Facial expression recognition
- Gesture and action recognition
- Video semantic indexing
- Visual Focus of Attention (VFOA)



- Defined according to application requirements
  - Segmentation of multimodal streams
  - Structuring by meeting events
  - Identification of group activity
  - Linguistic and discourse events
  - Indexing and retrieval
  - Summarization, and generation of textual and multimodal summaries

Abstract | Actions | Decisions | Problems

The Industrial Designer gave her components concept presentation, which presented the components that will be used and how they will be integrated into the functional design. She announced that their only choice for casing material was wood, so she suggested an idea of having customizable wood cases. The User Interface Designer gave her presentation on the user interface concept for the product, and gave a basic layout of the remote's key



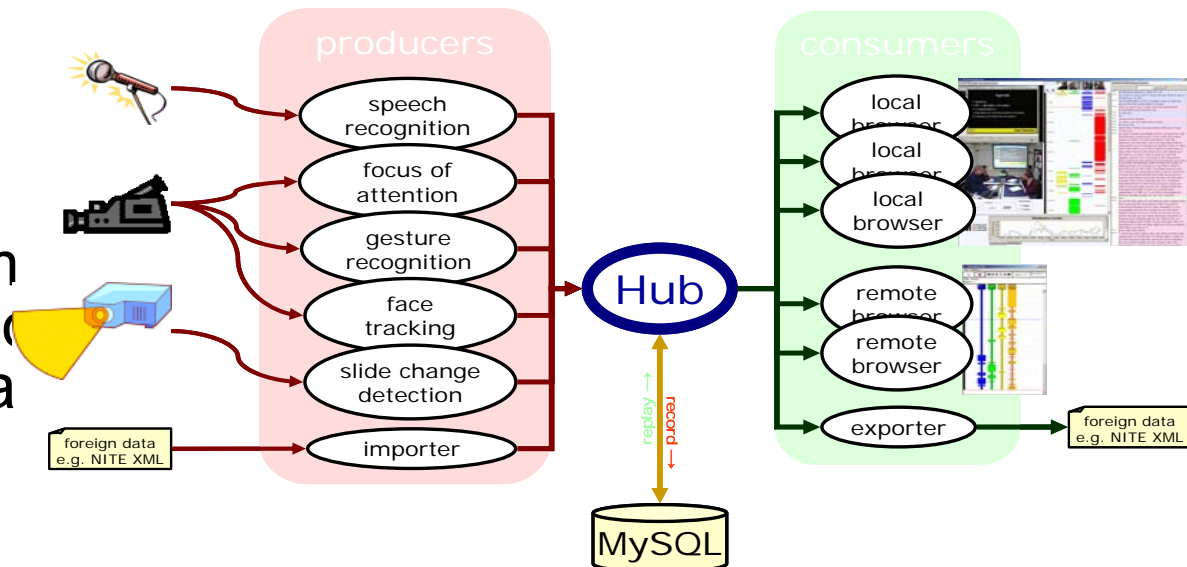
The screenshot displays a software interface for analyzing a meeting. It features a video window on the left showing two people in a meeting room. On the right, a window titled 'Real Reaction' displays a slide titled 'Conceptual Design meeting' presented by Sridhar Dasari, Project Manager. Below the video, a transcript window shows a list of speech segments with timestamps (1:00, 2:00, 3:00, 4:00) and corresponding text. To the right of the transcript is a summary window with a tabbed interface (Abstract, Actions, Decisions, Problems) containing a detailed textual summary of the meeting content, matching the text shown in the top-right window of the slide.



- Multi-page graphic novels
- API for metadata input
- Parameterizable presentation
- Future version embedded in

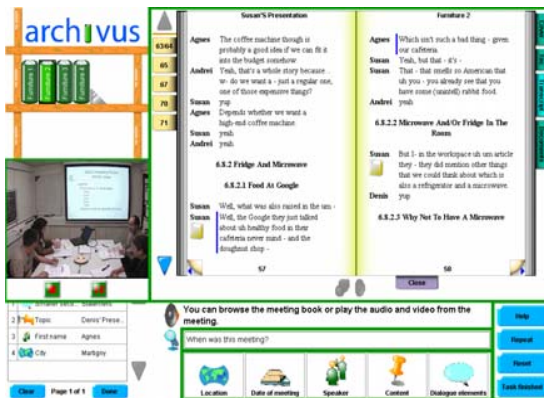
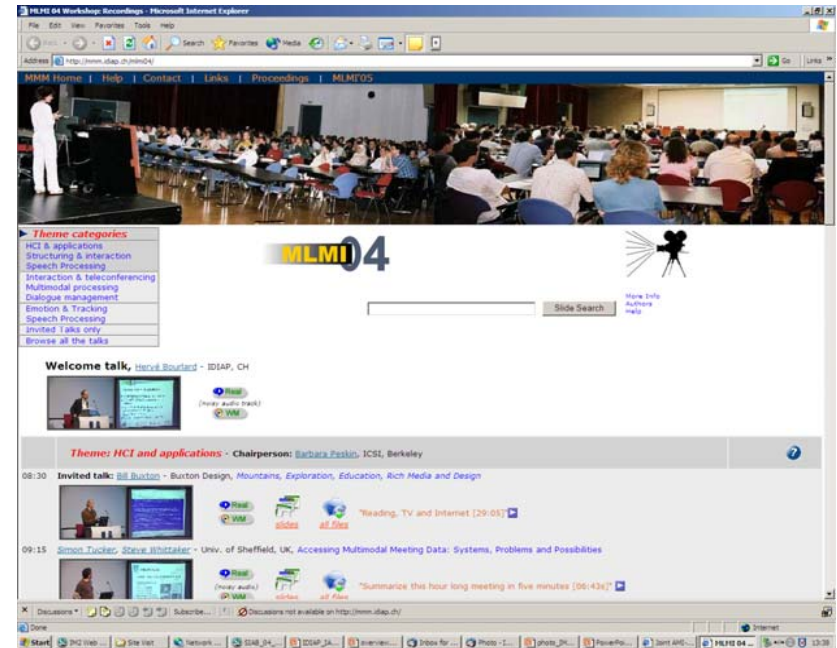
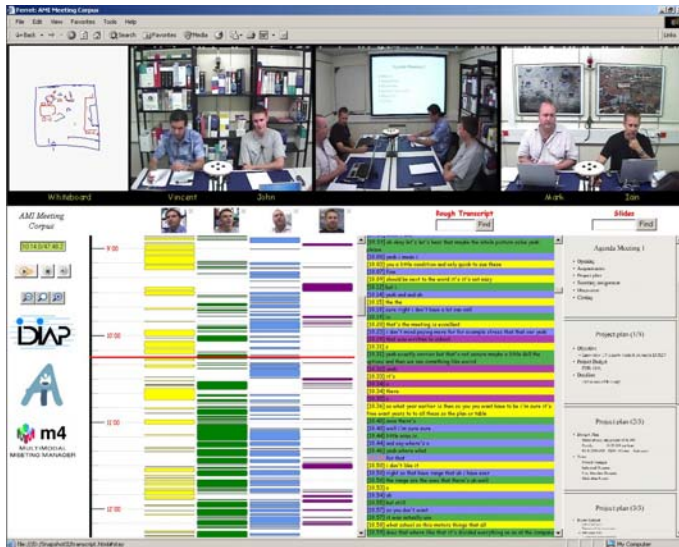
# Flexible and Inter-Operable Server (Hub)

- Innovative client-server architecture for live meeting annotation
- Java-based system for live distribution of data between “data producers” (e.g., recognizers) and “data consumers” (e.g., meeting browsers)
- Supports both past archived meetings as well as live meeting processing and browsing



# Browse, search, and navigate in meeting archives

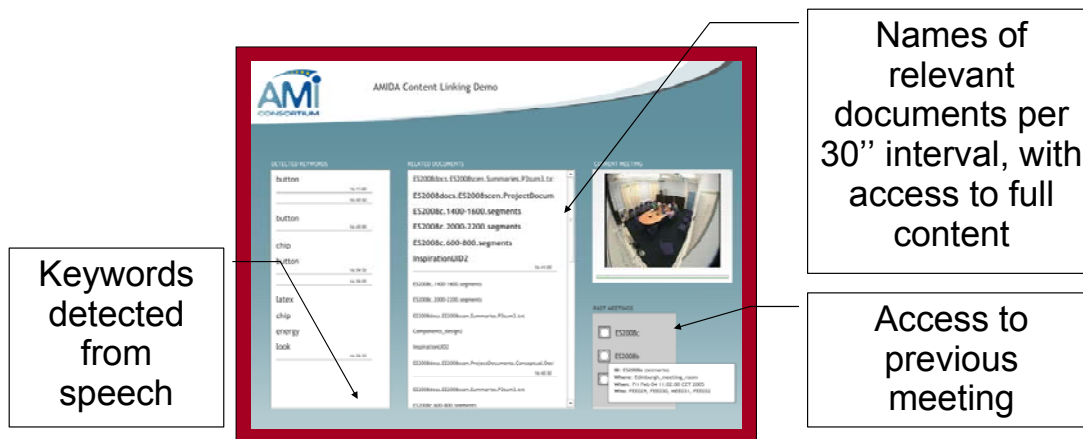
- JFerret plug-ins (browsing, search, navigation)
- Applications and interfaces easily customized





# Automatic Content Linking Device

- 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

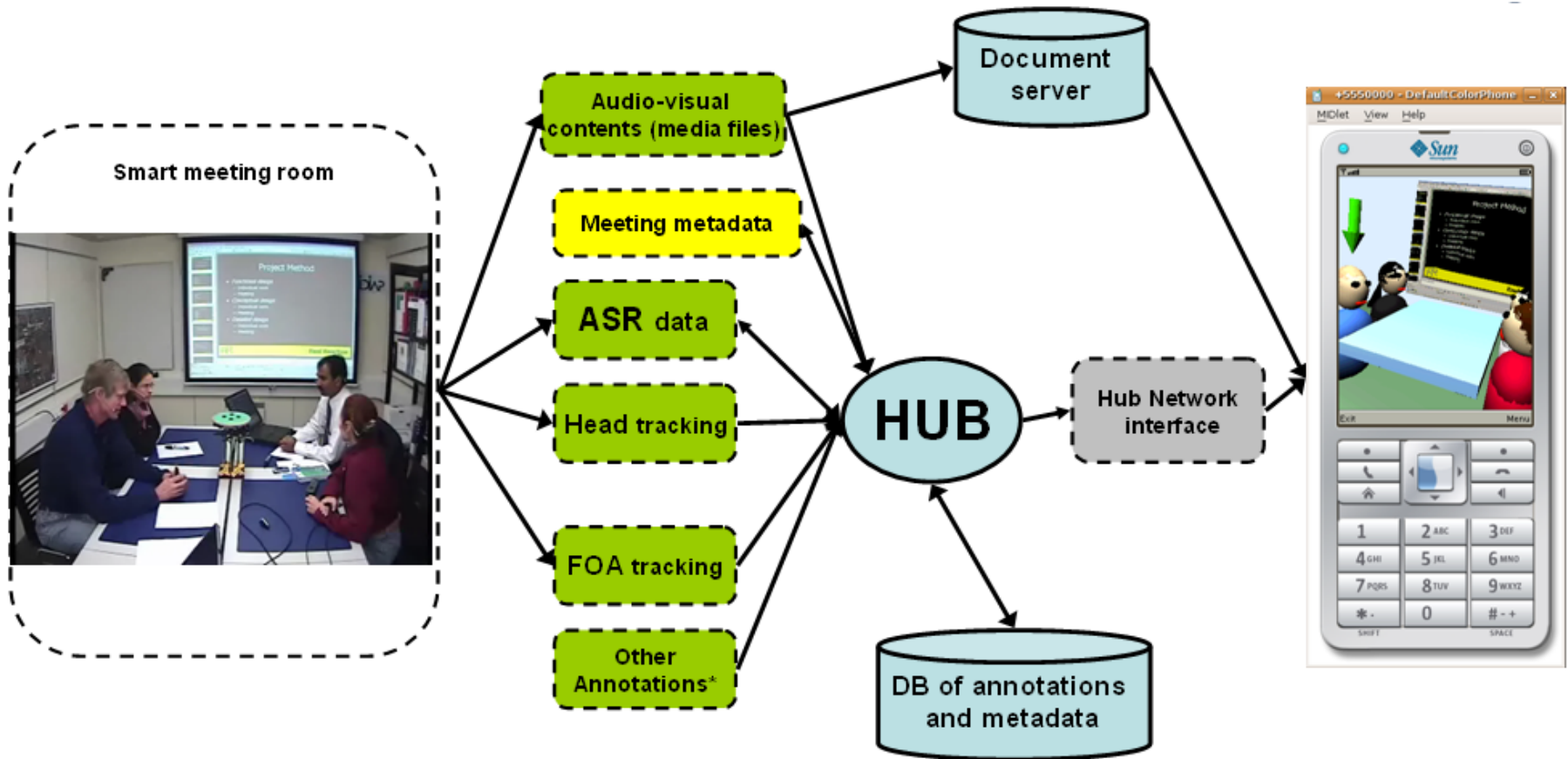


Keywords detected from speech

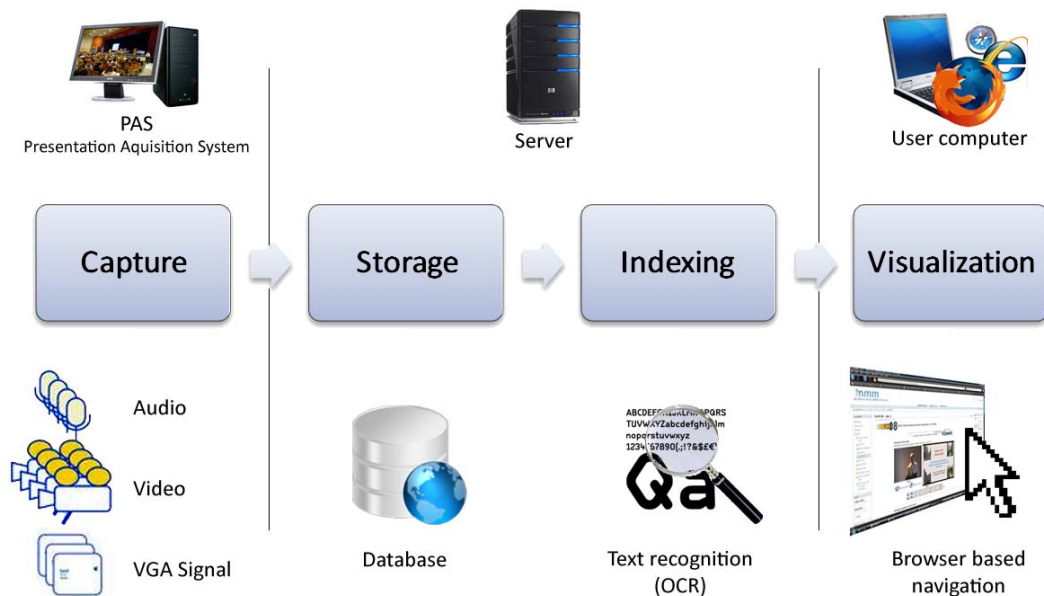
Names of relevant documents per 30'' interval, with access to full content

Access to previous meeting

# User Engagement and Floor Control (UEFC)



# Klewel: Lecture Captioning



<http://www.klewel.com>

# Community of Interest



# Key Takeaways

- Large multimodal meeting data available (100 hours annotated)
- Advances in Research
  - Audio and Video Signal processing
  - Multimodal content capture and analysis
  - Innovative representations of meeting metadata
- Can be used in new tools to automatically
  - Enhance people during meetings
  - Use/reuse recorded meetings better