IM2 Phase III
IP1: Integrated Multimodal Processing

IP Head: Aude Billard (EPFL)
Deputy IP Head: Stephane Marchand-Maillet (UniGE)

IM2 Summer Institute
September 2009

iM
INTERACTIVE
MULTIMODAL
INFORMATION
MANAGEMENT

IP1: Integrated Multimodal Processing

- Goals

- Integration within IM2

- Structure of the IP

- Embedded research

IM
INTERACTIVE
MULTIMODAL
INFORMATION
MANAGEMENT

Goals and Integration

- To continue supporting leading developments in IM2
  - Collaborative research where IM2 has shown strength and leadership
  - → Technology provider to IP3 "Social Signal Processing"

- To make a transition between research developments and integrated and validated tools
  - In order to create a long-term asset and IM2 showroom
  - → Relationship with IP2 "HCI and Evaluation" geared towards innovative interfaces

- To support the emergence of a long-lasting IM2 research community

Integration within IM2

IM2 Management

- Profs. H. Bourlard, T. Ebrahimi
- IP Heads
- IP Deputy Heads
- Steering Committee members

IP1. Integrated Multimodal Processing

- Core multimodal technologies geared towards integration into end-to-end applications
- Multimodal processing of meetings and other data
- Meeting browsers and assistants
- Evaluation of integrative applications

IP2. Human Centered Design and Evaluation

- New educational application(s) based on IM2 technologies
- Exploitation of multimodal research
- Ergonomics, usability, and user studies

IP3. Social Signal Processing

- Analysis of social behavior through non-verbal communication
- Vocal behavior, gestures, other cues
- *Relations* to SSPNet and PetaMedia EU networks of excellence, and to Affective Sciences NCCR

Knowledge and Technology Transfer, Education, Exchanges, Advancement of Women

iM
INTERACTIVE
MULTIMODAL
INFORMATION
MANAGEMENT

Research structure

From data to users

- **Unimodal processing** research
  - Considered as key to all (multimodal) developments
- **Multimodal processing, recognition and abstraction**
  - Still the main focus of our work
- **User-centric** developments
  - Showing convergence of past research
- **User-based evaluation** of our developments
  - Demonstrating the usability and utility of our findings

IM
INTERACTIVE
MULTIMODAL
INFORMATION
MANAGEMENT

Research plans (I)

- Audio/visual processing and recognition
  - Detection and tracking: speech, body, face, gaze
  - Recognition: diarization, face, object
- Multimodal processing, recognition and abstraction
  - Joint work on scene analysis and recognition for multimodal emotion characterization
  - Extension of the BIMODET project to label video data from multimodal cues
  - Joint work on abstraction and interaction for multimodal data indexing and navigation
  - Automatic content linking (ACLD) for spontaneous document retrieval

Research plans (II)

- Prototype integration and information display
  - ACLD and user-friendly JIT information presentation
  - Facilitating remote presence in meeting via integrated information presentation
- Innovative interface technologies
  - Tangible UI, new sensors for information browsing
- New evaluation methodologies and tools
  - Geared towards assessing both usefulness and usability

=> Clear synergies with IP2 "HCI and evaluation"

FNSNF
FONDS NATIONAL SUISSE
SCHWEIZERISCHER NATIONALFONDS
FONDO NAZIONALE SVIZZERO
SWISS NATIONAL SCIENCE FOUNDATION

IM
INTERACTIVE
MULTIMODAL
INFORMATION
MANAGEMENT

Audio Processing
Meeting speech recognition

Objective : Leveraging previous/current IM2 researches

Keep advancing the state-of-the-art in feature extraction, acoustic modeling and decoding algorithms

Phase III: New focus on application of meeting room speech recognition in integrated systems

- Optimization of ASR processing for multimodal and integrated systems (eg. for summarization, dialog act classification etc.)
- Development of real-time integrated technology demonstrators (for eg. online content linking, etc.)
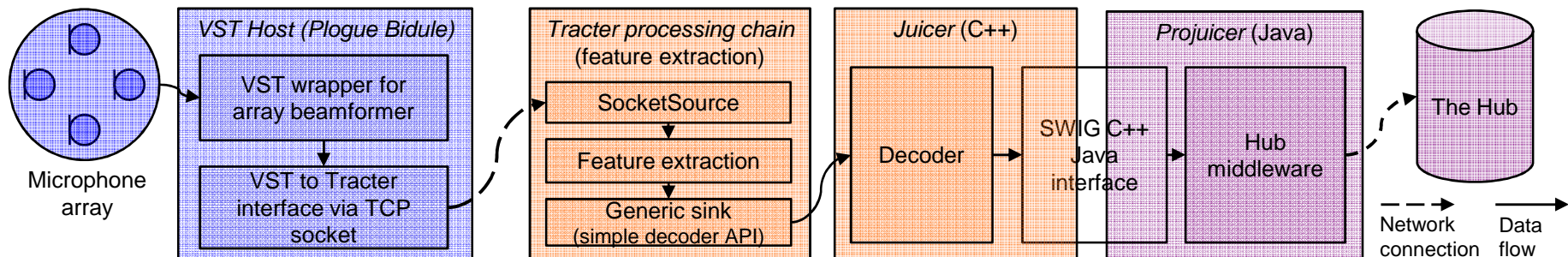
Audio/Visual processing
Speaker tracking, clustering and diarization

Objective:

- Integration of speaker tracking, clustering and diarization with visual modality – "audio visual association"

- Leveraging previous/current IM2 researches
  - Motion analysis
  - Visual Focus of Attention (VFOA)

Phase III:

- Study of online and closed loop systems (eg. interaction between diarization and VFOA)

- Analysis of higher level discourse content including vocal interaction, dominance and role detection

  - cf IP3 "social signal processing"
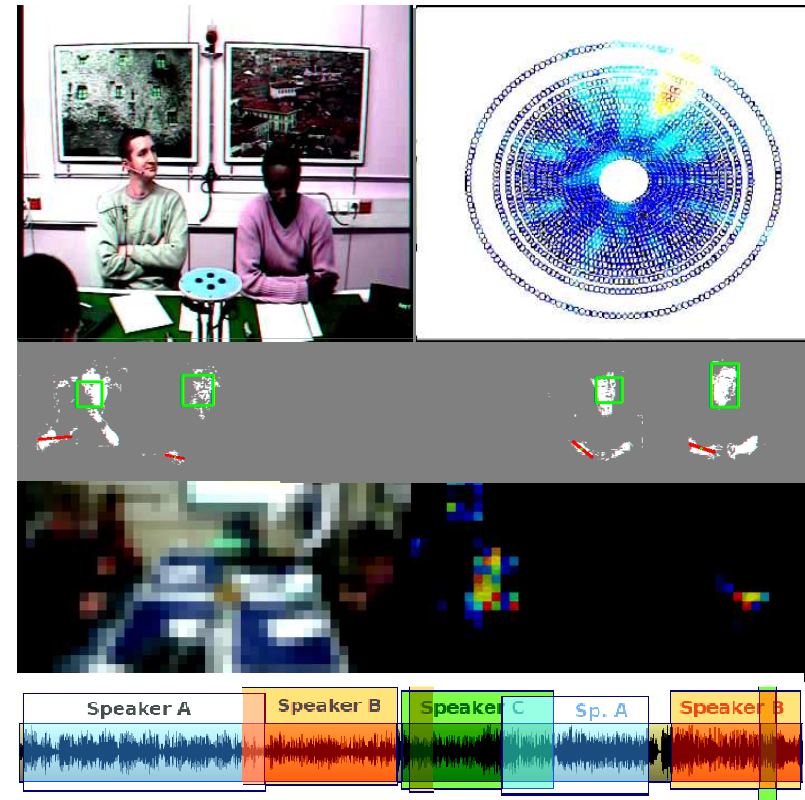
Audio/Visual processing
Multimodal emotion analysis in meeting data

Objective: Multimodal signal/image integration for emotion analysis in meeting data

- Leveraging previous/current IM2 researches
  - Face detection and tracking
  - Gesture analysis
  - Facial expression recognition
  - Audio-visual scene analysis
  - Multimodal integration

- Collaboration with the Affective Sciences NCCR (Prof. K. Scherrer)
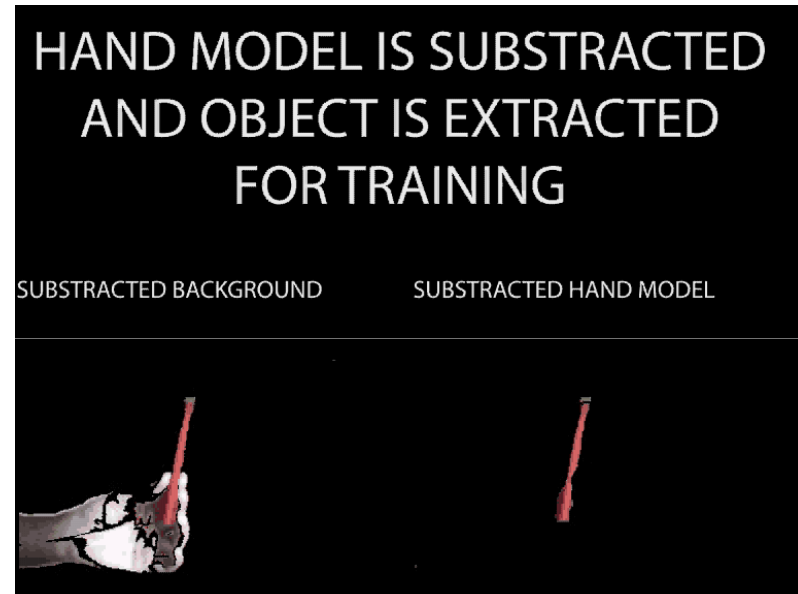 - Also relates to IP3
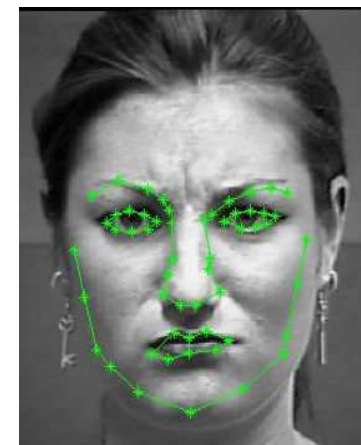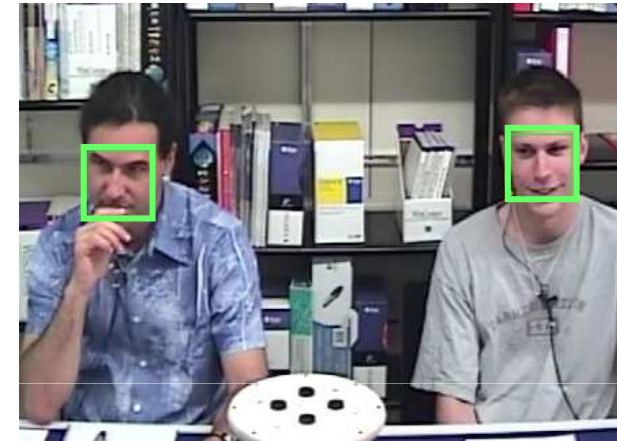
Multimodal Processing and Recognition
Bi-Modal Object Recognition

Objective: Automatic objects Labeling

- Leveraging previous/current IM2 researches
  - Object detection and recognition
  - Audio analysis for word spotting

Phase III:
- Evaluate efficiency of multimodal analysis on a real-world application with highly noisy data

- Collaboration EPFL and IDIAP

- Contribute to IP3 to analyze human-human interactions (focus of attention during dyadic social interaction)

HAND MODEL IS SUBSTRACTED AND OBJECT IS EXTRACTED FOR TRAINING

SUBSTRACTED BACKGROUND   SUBSTRACTED HAND MODEL

# Multimodal Abstraction and Indexing
Proactive information navigation

Objective: Leverage from our
multimodal data access models

- Query-based
- Browsing

PhaseIII:

- Combine focused interaction
  (search, browsing) with
  information propagation
  strategies to obtain semantic-
  level descriptions at low cost

- Automatic content linking to
  support meeting participants
  with novel information relevant
  to the ongoing task
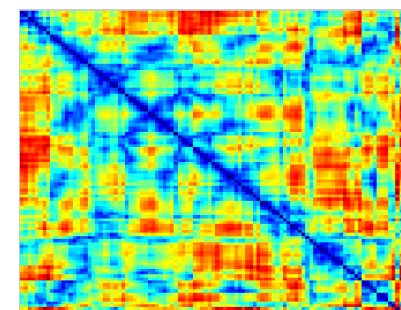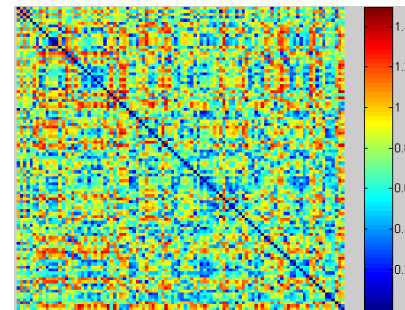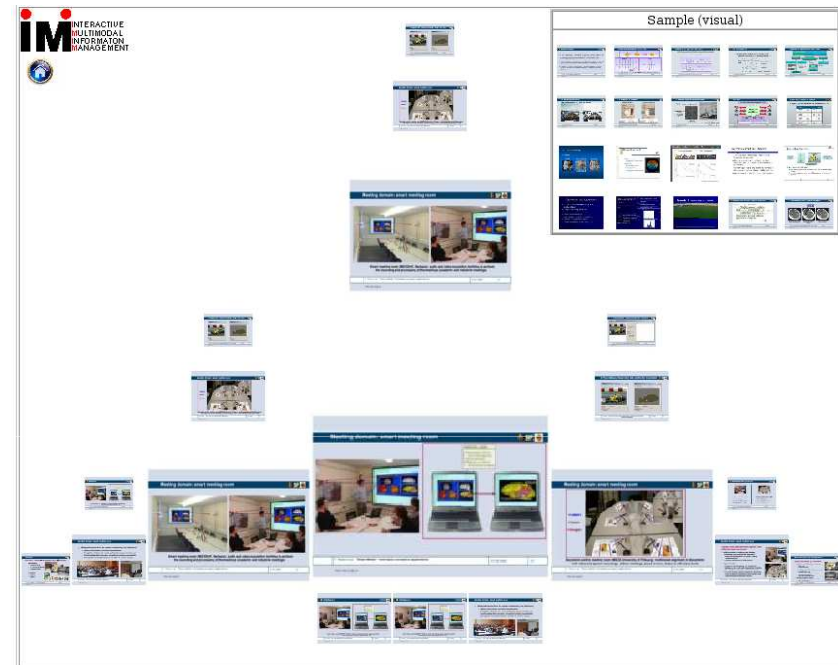
Human Computer Interaction
Interactive Meeting Assistants

Objective: Offer users online assistance

PhaseIII:
- Enhancing remote participation in meetings and conferences using unimodal and multimodal annotations

- User interface technologies for meeting assistance
    - integrating the most relevant IM2 research results
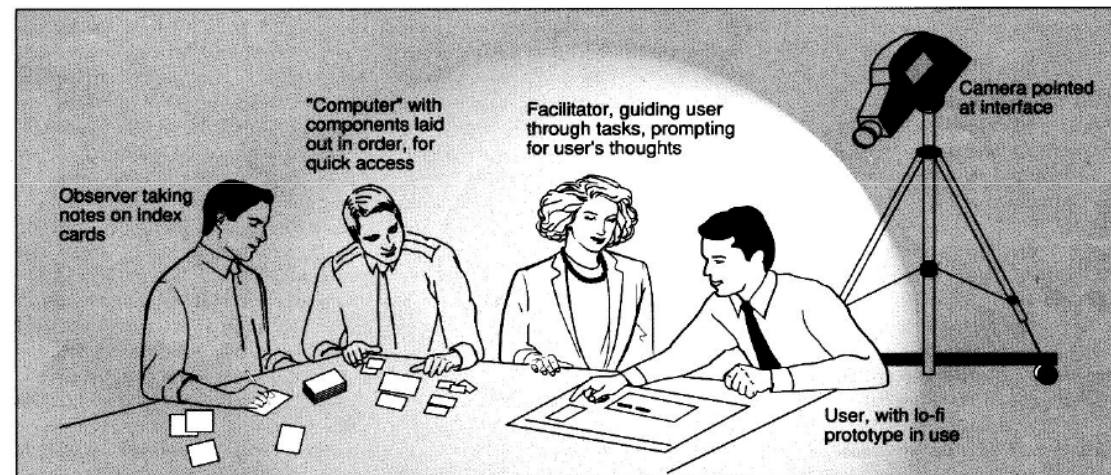   - → IP2: Integrating new input technologies, smart sensors & restitution techniques

iM INTERACTIVE MULTIMODAL INFORMATION MANAGEMENT

Evaluation

… and extension of the methodology and tools

<u>Objective:</u> User-based validation of IM2 assets

- Usability testing of meeting assistants
  - User performance, physiological responses, perceived usability, emotions, etc.

- Extend task-based evaluation methods
  - improve task-elicitation methods compared to the BET procedure, with more realistic tasks for each prototype
  - develop statistical methods for analyzing user interaction with integrated meeting processing applications

Composition

IM2 Management

- Idiap: **Prof. H. Bourlard**
- EPFL: **Prof. T. Ebrahimi**
- IP Heads
- IP Deputy Heads
- Steering Committee members

IP3. Social Signal Processing

- Idiap: **Dr. A. Vinciarelli** (IP Head)
- EPFL: **Dr. F. Dufaux** (Deputy IP Head, *new in IM2*), Prof. T. Ebrahimi
- ETHZ: Prof. L. Van Gool, Prof. V. Ferrari (*new in IM2*)
- Others

IP1. Integrated Multimodal Processing

- EPFL: **Prof. A. Billard** (IP Head), Prof. J.P. Thiran, Prof. T. Ebrahimi
- UniGe: Prof. T. Pun, **Dr. S. Marchand-Maillet** (Deputy IP Head)
- Idiap: Senior researchers in speech, vision, machine learning, interfaces and evaluation
- UniFr: Prof. R. Ingold, Dr. D. Lalanne

IP2. Human Centered Design and Evaluation

- UniFr: Prof. R. Ingold, **Dr. D. Lalanne** (IP Head), Prof. J. Sauer (*new in IM2*)
- Idiap: **Dr. A. Popescu-Belis** (Deputy IP Head)
- EPFL: Prof. P. Dillenbourg (*new in IM2*)

Knowledge and Technology Transfer, Education, Exchanges, Advancement of Women

- Idiap and all partners

IM
INTERACTIVE
MULTIMODAL
INFORMATION
MANAGEMENT

Timeline and milestones

Y1: Definition of final objectives

- M1: Consolidation of objectives for collaborative analysis of multimodal data. Cooperations and collaborations are in place towards clearly identified objectives.

Y2: Technology development and integration

- M2: Intermediate research reports available and definitions of academic prototype for demonstration of the two multimodal applications are available.

Y3: Prototype development

- M3: Basic research prototypes for multimodal applications are in place and ready for evaluation, including in cooperation with relevant partners from IP2 and IP3.

Y4: Evaluation

- M4: Release of documented final academic prototypes for potential technology transfer and user evaluation methods (e.g., field trials).