IM2 Phase III
Executive Summary

Hervé Bourlard, IM2 Director
Idiap Research Institute
Touradj Ebrahimi, IM2 Deputy Director
EPFL

IM2 Advisory Board,
September 2, 2011

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

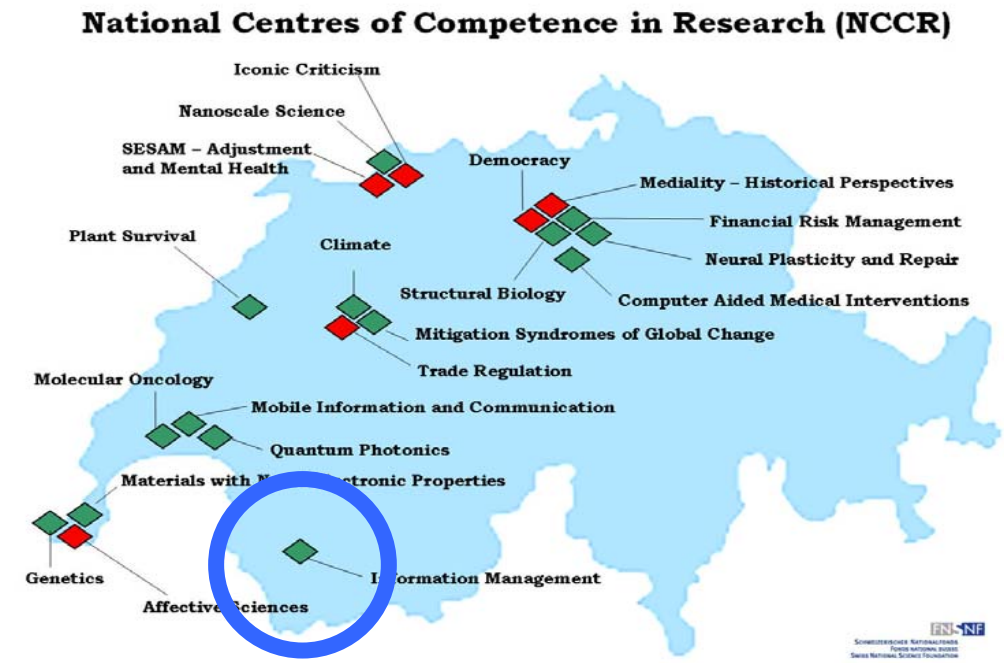
INTERACTIVE
MULTIMODAL
INFORMATION
MANAGEMENT

Outline

- Background and context
- Summary of IM2 Phase I & II (2002-2009)
- Summary of IM2 Phase III (2010-2013)

Background and context

20 National Centres of Competence in Research



Interactive Multimodal Information Management (IM2)

- Phase I (2002-2005):
  - SNFS funding: 15’400’000.- CHF
  - Self & third-party funding: 16’220’000.- CHF
- Phase II (2006-2009):
  - SNFS funding (Y3 20%, Y4 40% reduction): 11’900’000.- CHF
  - Self & third-party funding: 11’900’000.- CHF
- Phase III (2010-2013)
  - SNFS funding (40% of Phase II): 4’760’000.- CHF
  - Self & third-party funding: 4’765’000.- CHF

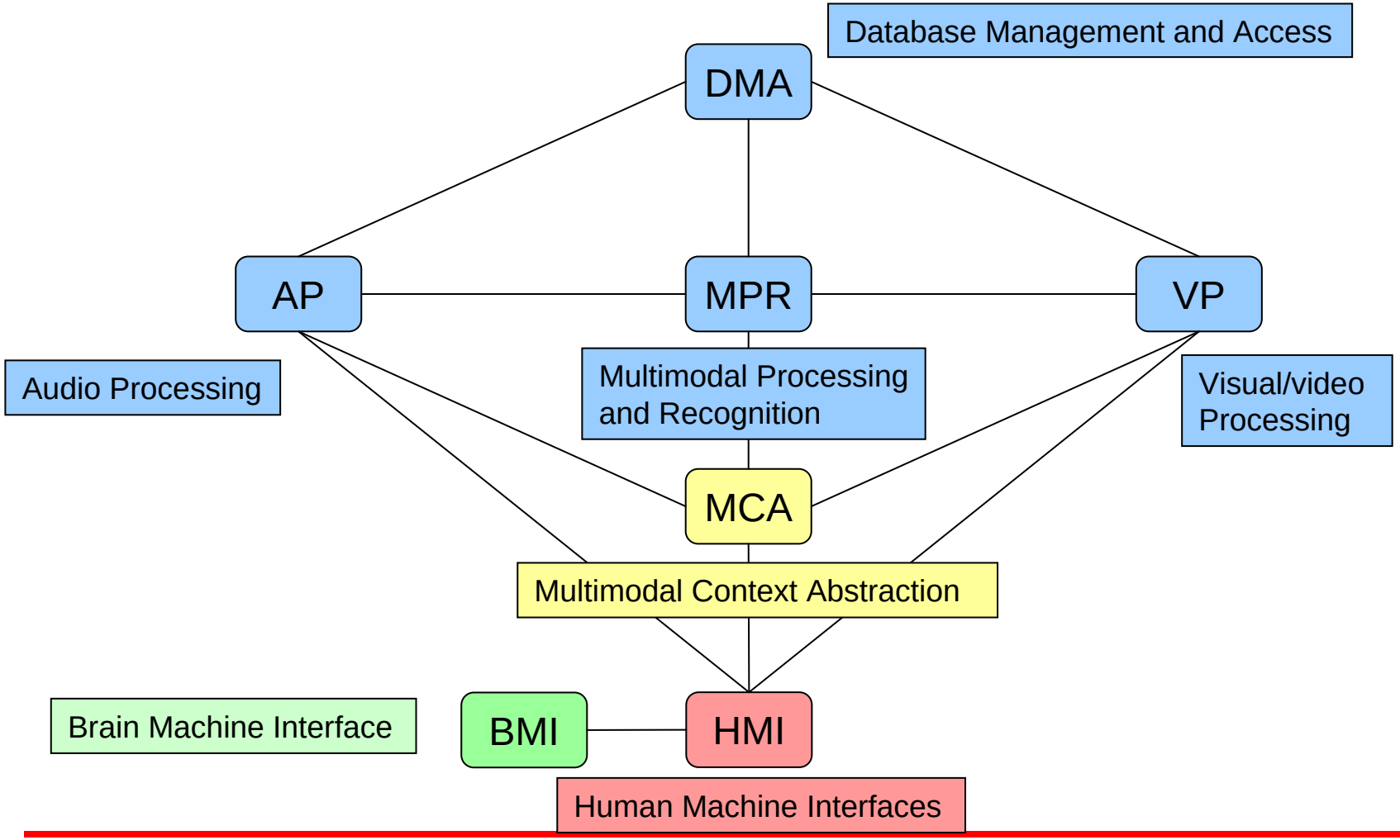
IM2 Phase I (2002-2005)

- Initiated the IM2 common vision:
  - Human-to-human interaction in multimodal meeting recordings, and indexing and retrieval of relevant multimedia information
- Developed the required hardware:
  - Instrumented meeting rooms at Idiap and the University of Fribourg), software infrastructure, etc.
  - Now available and used in multiple site in EU and US
- Collected and annotated a large common multimodal corpus:
  - Quite "unique" in size and (hierarchical) levels of annotation
  - Now used by multiple EU and US research institutions, include NIST as part of international evaluation campaigns
- Fundament research:
  - On this basis, we initiated and performed fundamental research in mono-modal and multimodal processing, encouraging the IM2 partners to test their research on the IM2 corpus, while adapting their software to the IM2 vision.

IM2 Phase II (2006-2009)

- Consolidation the IM2 research and development efforts on our common vision
  - IM2 involved in many EU and US projects related to IM2
  - IM2 involved in numerous evaluation campaigns
  - Resulting in large amounts of high quality publications
- Increased emphasis on multimodal processing, system integration, HCI, and system evaluation
- Development of the "Hub", allowing easy integration of multiple modalities, software tools, and applications
- IM2 was enriched with a flexible user interface development tool (JFerret)
- Significant structural impact

IM2 Structure before Phase III



IM2 Phase III (2010-2013)

- Consolidation of most of the IM2 activities in one single IP to further foster collaboration and focus **(IP1)**
- More emphasis on human-centered design and evaluation **(IP2)**:
  - Testing generalization properties of IM2 results (IP1) on different application domains
  - Developing more user-centric applications and evaluation protocols (required new partners)
- Extend fundamental research in one of the newest, but most promising, (longer-term) outcomes of IM2-II **(IP3)**:
  - Social signal processing

IM2 Phase III – Structure

Management, TT, Community Building, and other activities

- Knowledge and Technology Transfer
- Education
- Exchange programs
- Advancement of Women
- Community Building

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 EU networks of excellence, and to Affective Sciences NCCR

Key applications in Phase III

- Augmented Teams

- The Cboard

Thank you for your attention!

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

INTERACTIVE
MULTIMODAL
INFORMATION
MANAGEMENT

IM2.IP1: Integrated Multimodal Processing

Aude Billard (EPFL)
Stephane Marchand-Maillet (UniGE)

IM2 Advisory Board,
September 2, 2011

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

INTERACTIVE
MULTIMODAL
INFORMATION
MANAGEMENT

IP1: Integrated Multimodal Processing

IP1 covers primarily multimodal research

Goals:

- To investigate promising research directions initiated in IM2-Phase II for multimodal research
- To pursue uni-modal research (audio – visual)
- Especially collaborative research (intra- and inter-IP)

IM2.IP1

- 8 Reporting subprojects
  - Idiap (Gatica-Perez, Popescu-Belis, Garner), EPFL (LASA, LTS5, MMSPG), UniFri (DIVA), UniGE (Viper)
- One IP1 meeting per term

  →1 or 2 long subproject presentations (30 mins)
- Heads of IP2 participate to IP1 meetings
  - Ensure a good communication
- Collaborative writing of IP1.QReport

Outline

1. Interactive tools in support of human information sharing

2. Analysis of human behavior

3. Content analysis and processing

Outline

1. Interactive tools in support of human information sharing

2. Analysis of human behavior

3. Content analysis and processing

Multimodal meeting assistants

- Semantic search + software development for the ACLD
  - A. Popescu-Belis, M. Yazdani (PhD), A. Nanchen (Eng.)

- Semantic relatedness between words or documents
  1. map the words/documents to the English Wikipedia
  2. graph-based distances over the network
     - computed using visiting probability (VP) of random walk

- Theoretical advancements
  - learning a distance metric to make computations tractable
    - uses samples of closest and farthest articles in terms of VP
  - document enrichment with new related words from the network
  - justifications of truncations of the walk (theoretical & empirical)
    - leading to a motivated choice of system parameters

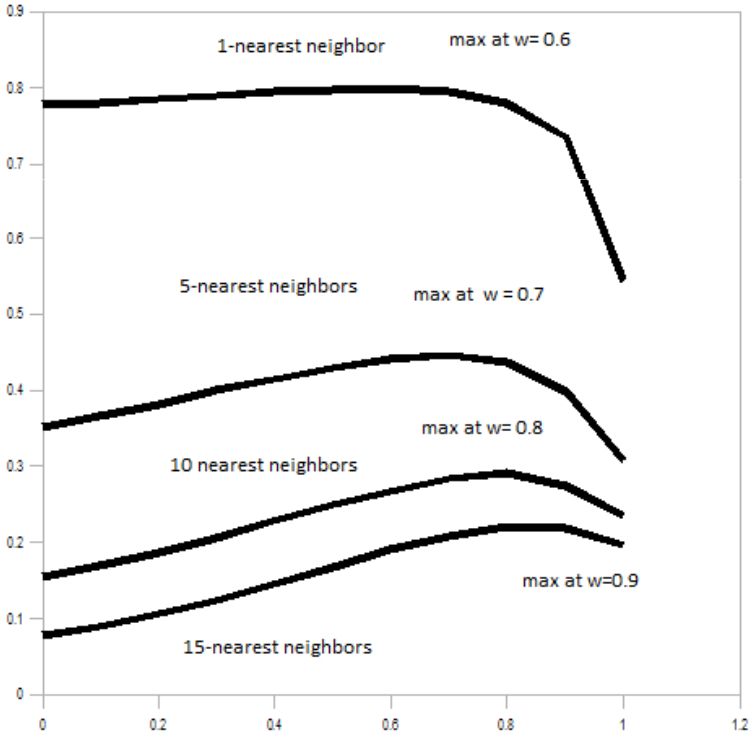
Semantic search: tasks, data, results

- Information retrieval (on Trec8)
  - doc. indexing: sort Wikipedia concepts by average VP to/from the document
  - best results for a combination of keyword-based and concept-based scores (with pseudo-relevance feedback) & improve state-of-the-art

- Document clustering (20 Newsgroups)
  - using Wikipedia hyperlinks increases recall
  - using lexical similarity increases precision

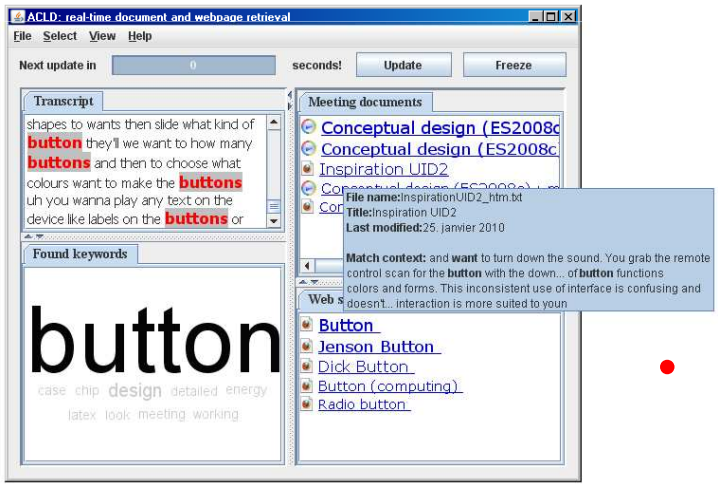
- Illustration
  - purity of clusters of k-closest neighbors (k = 1, 5, 10, 15) depending on the weight of visiting probability (w) vs. cosine lexical similarity (1-w) used for constructing them

  ➔ a mix of the two is optimal; VP is more important for better larger neighborhoods



Semantic search: application to ACLD

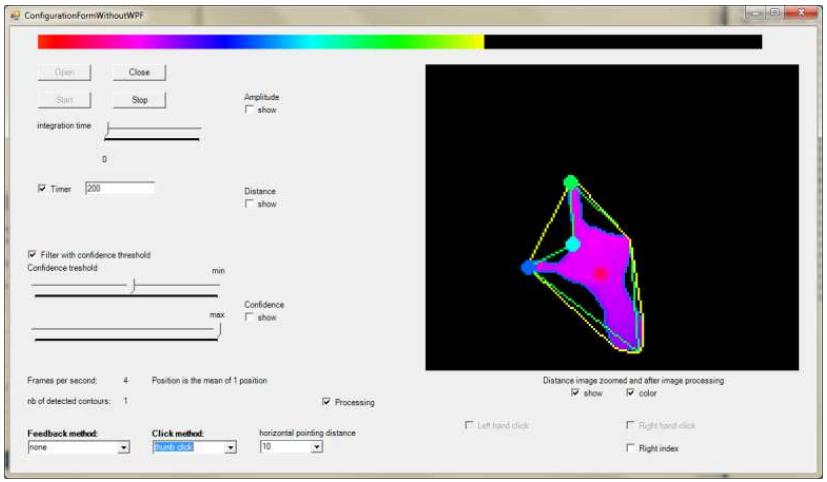
- Semantic search for the ACLD
  - compared Wikipedia articles suggested using semantic search vs. using key-word based search, for each 15'' snippet of a meeting
  - human evaluators found semantic search results more relevant



- Software development for the ACLD
  - installation at EPFL/CRAFT for IM2.IP2
  - streamlined code to reduce number of modules communicating through the Hub
    - ➔ solved performance issues, simplified code
  - ported to a Mac laptop with 8 GB RAM
    - ➔ solved issues related to 64-bit
- Demos
  - ACM Multimedia workshop, ACL-HLT 2011, SIGdial 2011

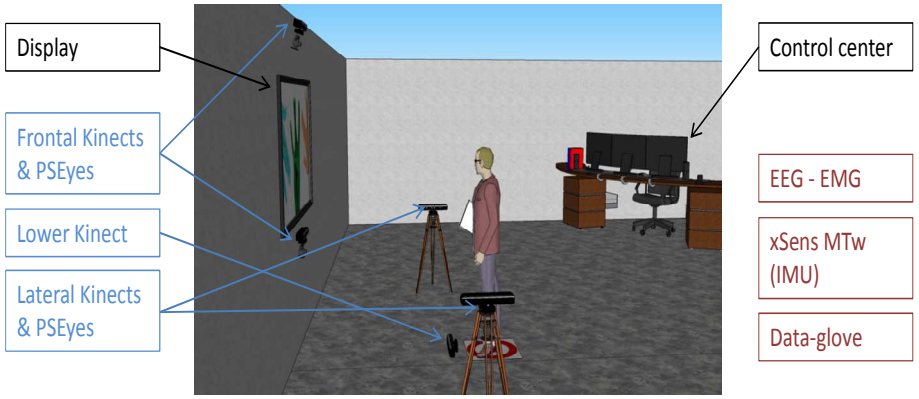
Economic Gestures for HCI (1)

- 1 PhD work (2 other PhD students on matching funds)
- Develop of a new way to interact with large vertical surfaces using hand gestures
  - Economic gestures (not tiring for users)
  - Current focus on pointing and selection
- Integration in Communication Board (C-Board - IP2)



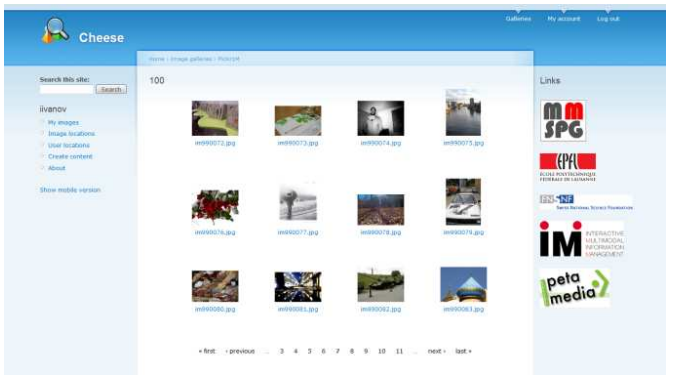
Economic Gestures for HCI (2)

- Comparison of different selection/pointing strategies performance using a Fitt's Law experiment (using SR4000 time of flight, Kinect)
- Best result so far with absolute pointing and thumb selection, as well as "dart" strategy (Index of Performance over 2 bps)
- Acquisition of a ground-truth with multiple sensors: high resolution (e.g. inertial motion units, gloves) and low resolution (e.g. cameras)
- Future work on two hands interaction: 1 hand for gestural command, other one for pointing/selection

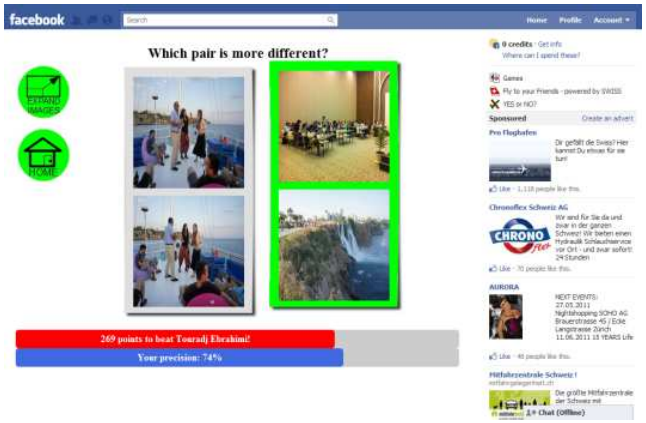


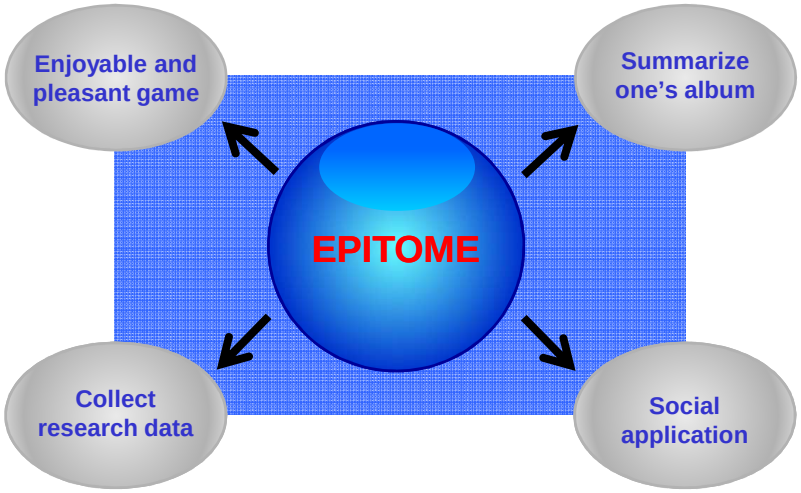
Multimodal Content Annotation

1. Swiss Cheese – An advanced image management platform
   - http://cheese.epfl.ch/
   - Visual similarity based search, semi-automatic tag propagation
   - Interoperability between different image repositories – JPSearch - Part 4 compliant



2. Epitome – A social game for photo album summarization
   - http://apps.facebook.com/epitome/





Distributed Multimedia Indexing

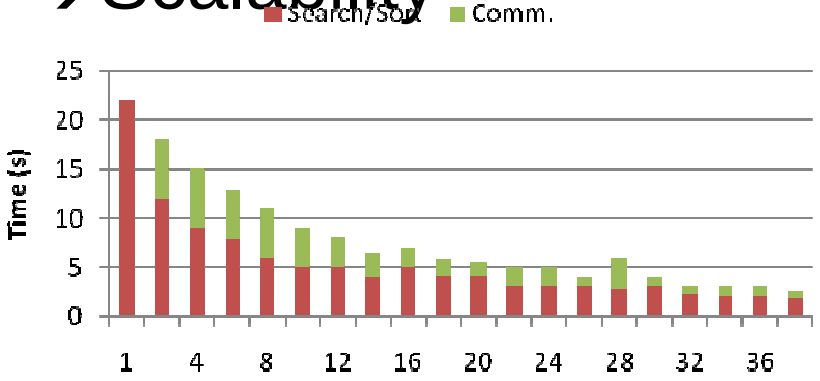
Indexing of large-scale multimedia collection via the distribution of our search engine over a computer cluster

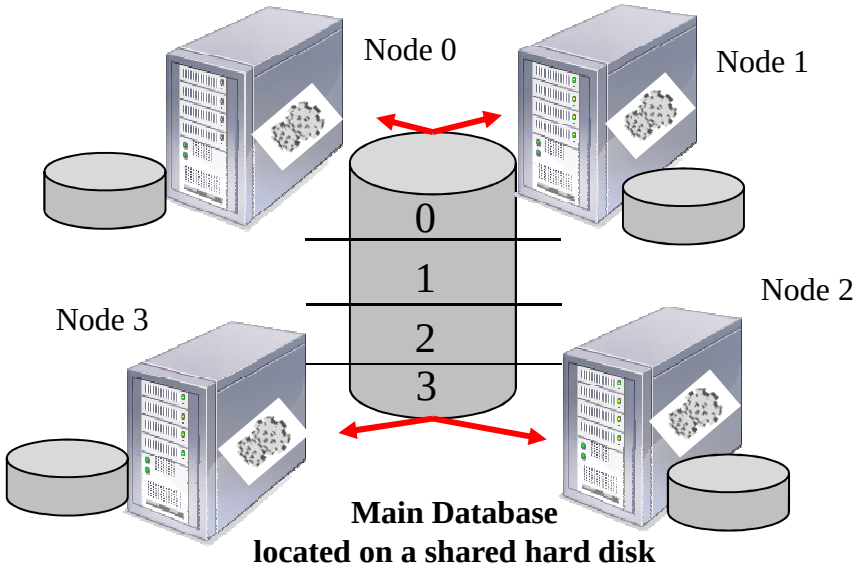
→1 million image collection

→40 available cores

→Performance gain

→Scalability





Outline

1. Interactive tools in support of human information sharing

2. Analysis of human behavior

3. Content analysis and processing

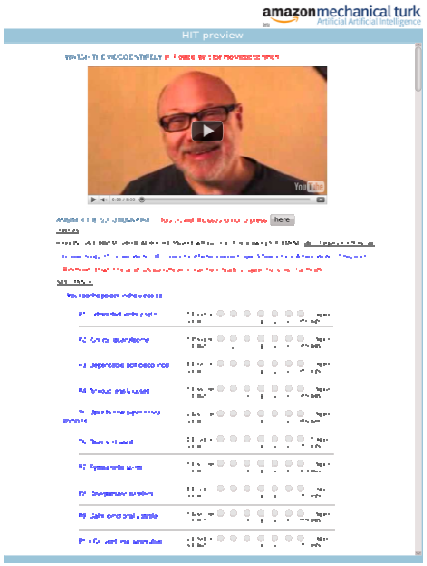
Recognition of group conversational context with sociometers

Supervised learning approach to discriminate two types of conversational context (brainstorming vs. decision-making) from nonverbal audio cues extracted from wearable sociometers (collaboration with A. Pentland, MIT)

Hypothesis: differences in conversational dynamics between the two types of contexts are significant and measurable

The method can obtain classification accuracy of 85%

Nonverbal behavior and personality impressions in YouTube



Study of the use of audio-visual nonverbal cues as video blogger (vlogger) behavioral descriptors

We found associations between automatically extracted cues and Big-Five personality judgments obtained via crowdsourcing (Mechanical Turk)

We also found associations between personality scores and the level of social attention vloggers receive in YouTube

Initial work with IP2: conversational analysis of CBoard data

photo: dev-audio.com

The CBoard team collected an interaction data corpus using a Microcone

The interactions involve three participants solving different tasks in a couple of contexts

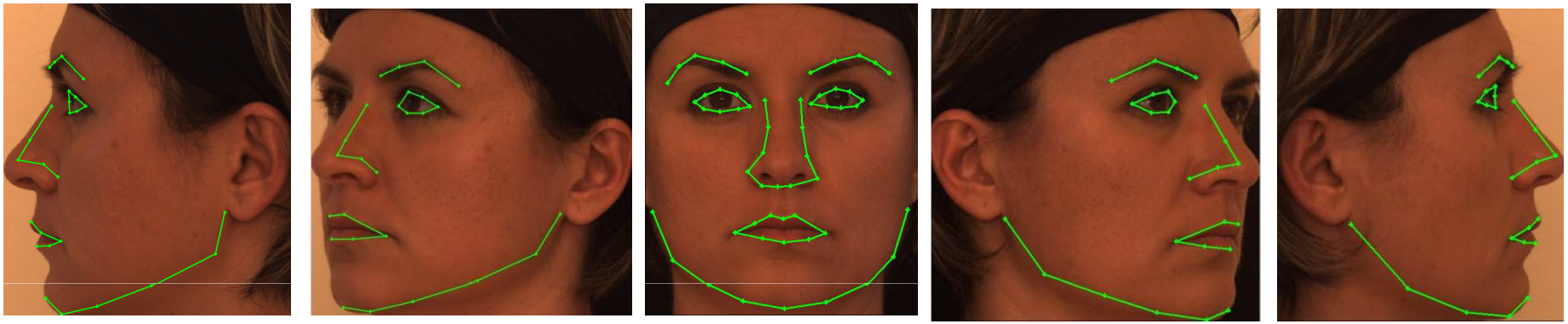
We automatically extracted basic turn-taking and prosodic features from the Microcone data, and made a first attempt to automatically characterize the interaction context. Initial recognition results show the task to be challenging

Dynamic Facial Expression Analysis

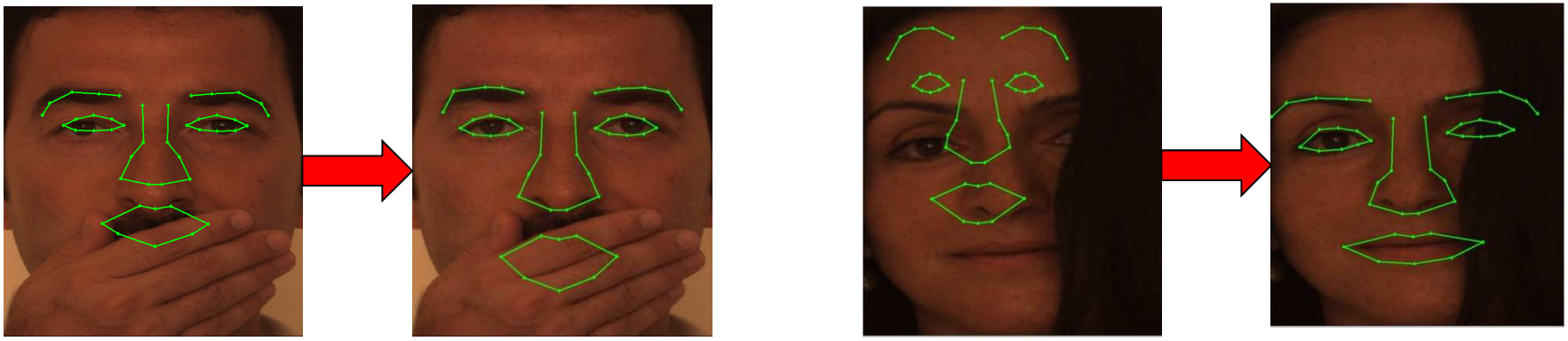
- Integration of temporal features from facial expressions for a detailed analysis of emotion
- Adopting the Component Process Model of Emotion instead of the basic emotion model
- Working on diagnosis of emotion deficits (flat / inappropriate affect) using the dynamic approach and the CPM
- Close collaboration with the Affective Sciences NCCR: joint PhD student supervised by Prof. J.-Ph. Thiran (EPFL – IM2) and Prof. David Sander (Univ. Geneva – Affective Sciences NCCR)

Pose – Occlusion Robust Face Tracking

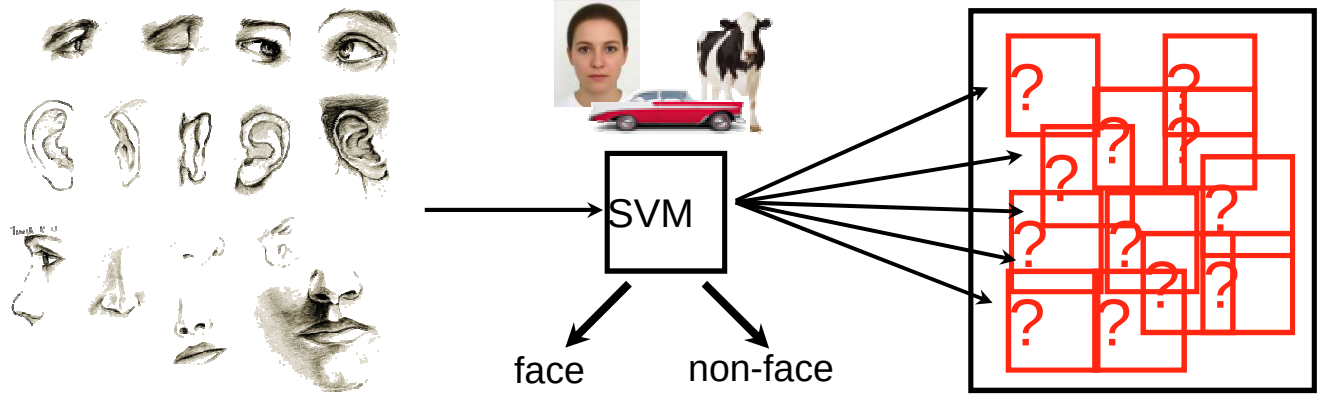
- Stack of models approach for handling extreme poses



- Occlusion robust pose detection & face detection using Fast Robust PCA



Robust Face/Object Tracking



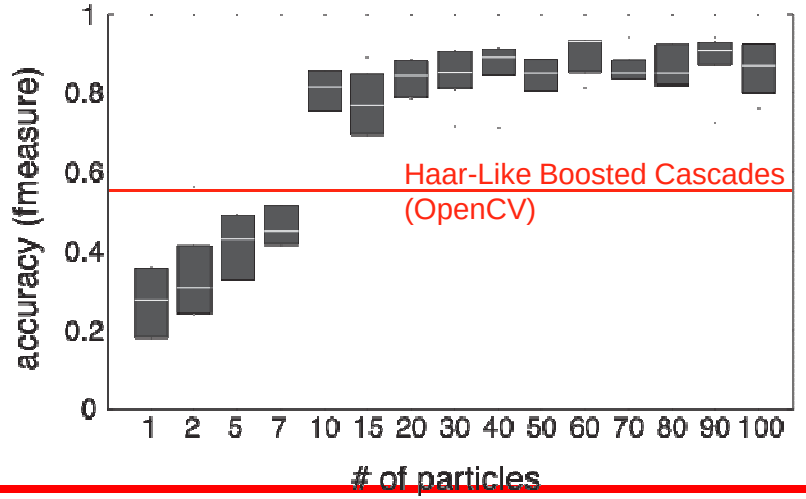
Feature Selection by
Iterative Kernel Polarization

SVM Classification
on IKP features

Tracking by
Particle Filtering

Protocol

Dataset:
10 WearCam Videos
6 min per video
Training:
1-100 particles
10-fold Cross-Valid.
Testing:
Hit = overlap > 75%



[EPFL]

Aggregation of asynchronous gaze-tracking streams

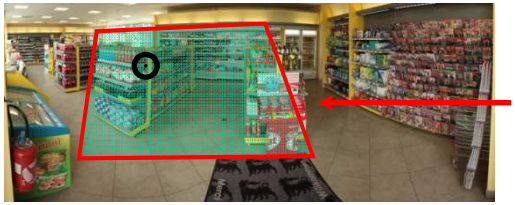
Step 1:
Localizing Wearcam Image

gaze direction

Wearcam Image(s)

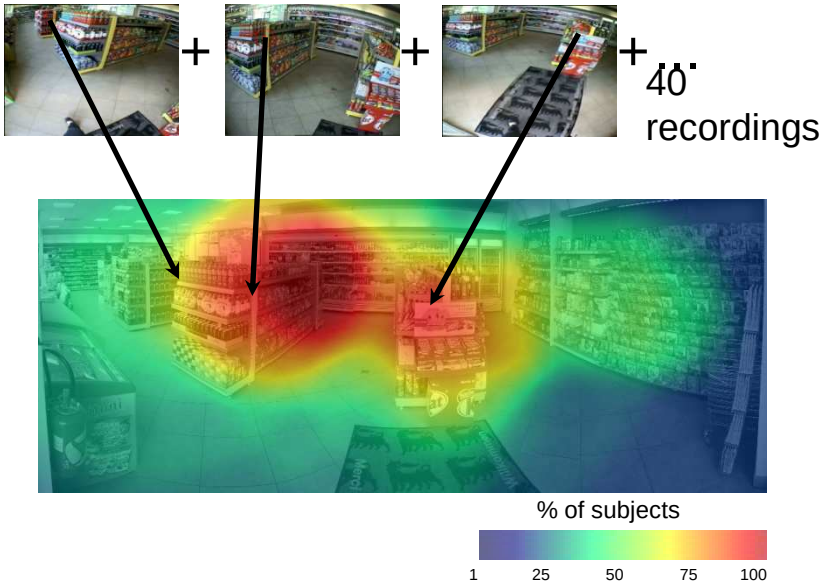
Reference Image

Matching SURF Descriptors



MMA-based Homography for Uncalibrated Images

Step 2:
Combining multiple recordings



Outline

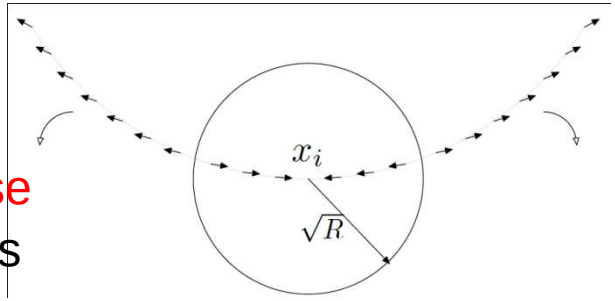
1. Interactive tools in support of human information sharing

2. Analysis of human behavior

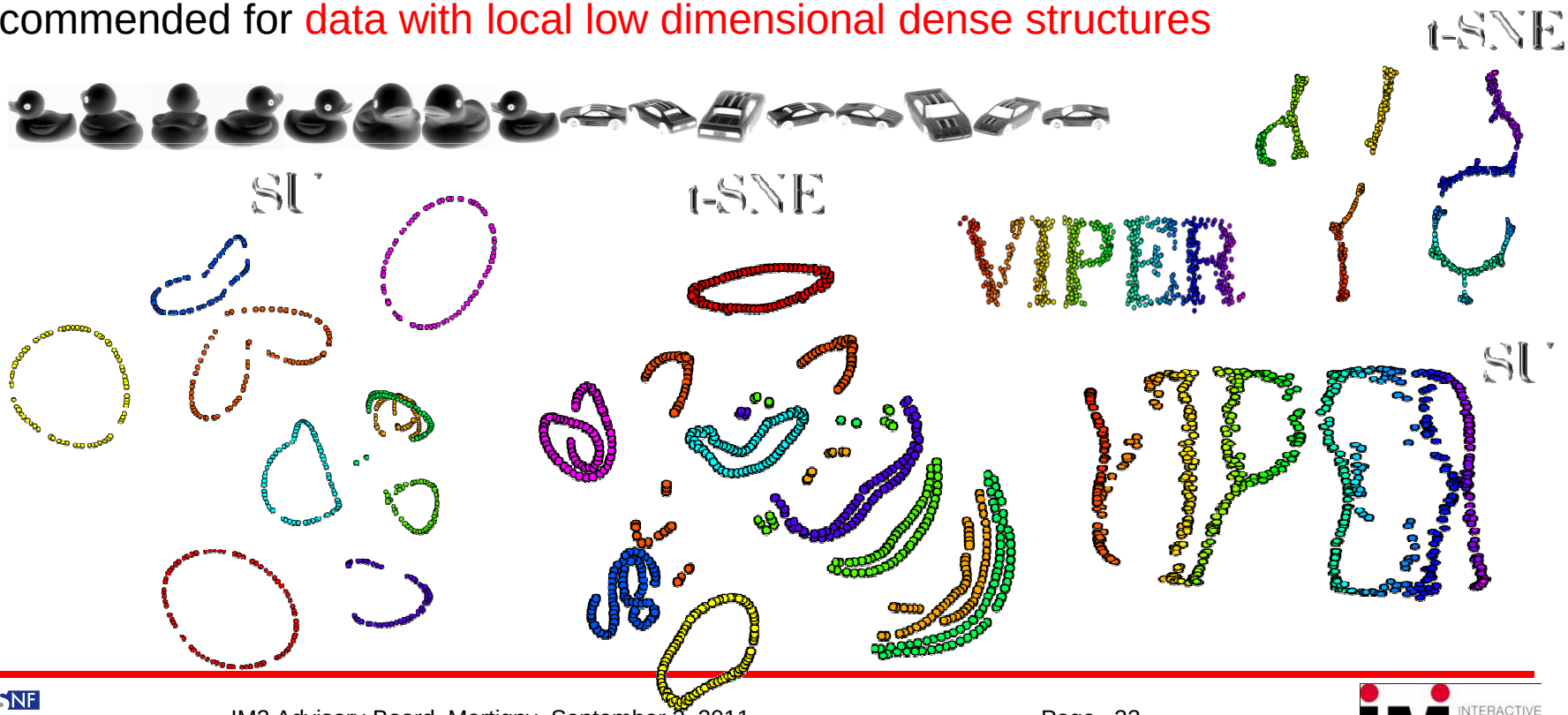
3. Content analysis and processing

Stochastic Unfolding

$$\min_{\mathbf{y}\in\Re^m} \sum_{i,j} q_{j|i}(\mathbf{y})||x_i - x_j||_2^2, \quad s.t. \sum_{i,j} p_{j|i}||y_i - y_j||_2^2 = 1$$



- A novel push-pull strategy to unfold a manifold and de-noise
- More robust on noisy real world data than spectral methods
- Better at preserving global topology compared with t-SNE
- Recommended for data with local low dimensional dense structures



Multimodal Quality Metrics for Multimedia Content

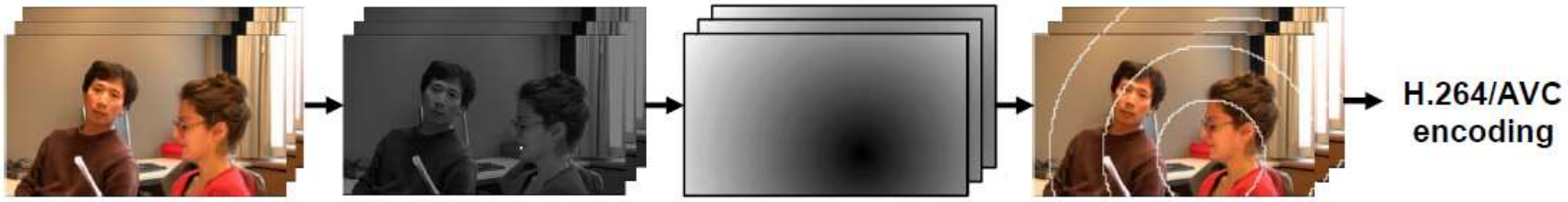
1. In the context of quality assessment for codec performance evaluation and comparison:
   - Pair comparison subjective methodology for the analysis of Scalable Video Coding
   - Comparison of VP8 image and video compression to JPEG, JPEG 2000, JPEG XR, H.264/AVC and HEVC

2. In the context of quality of experience in mobile scenario:
   - Study of user attention profiles during mobile video consumption (HAPQNET EuroNF Network of Excellence)

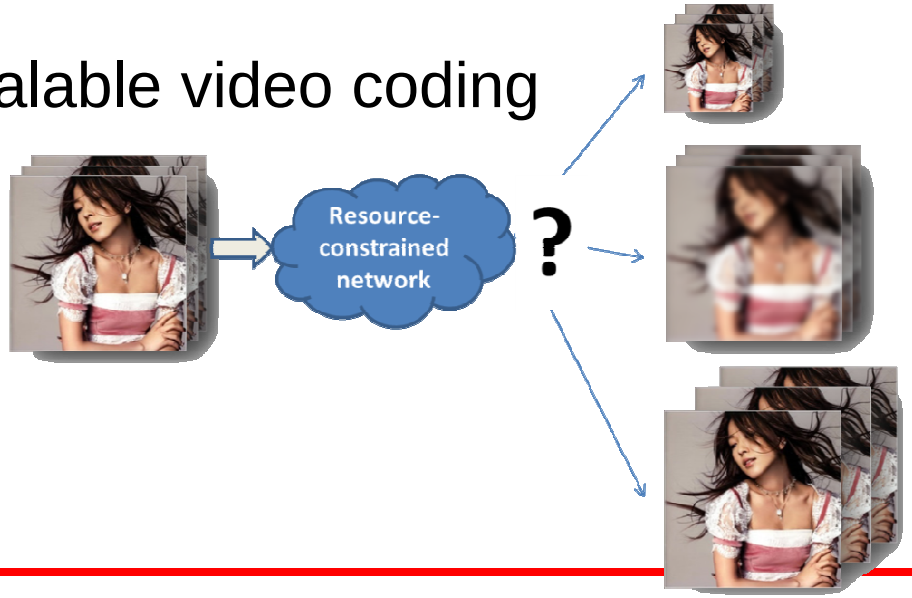
3. In the context of video transcoding:
   - Study of content and quality adaptive dynamic bit rate allocation for video transcoding in streaming application (iCelero industrial project)

Video Quality Assessment

1. Region-of-interest coding using audio-visual focus of attention
   - Study on effect of audio-visual focus of attention on quality perception



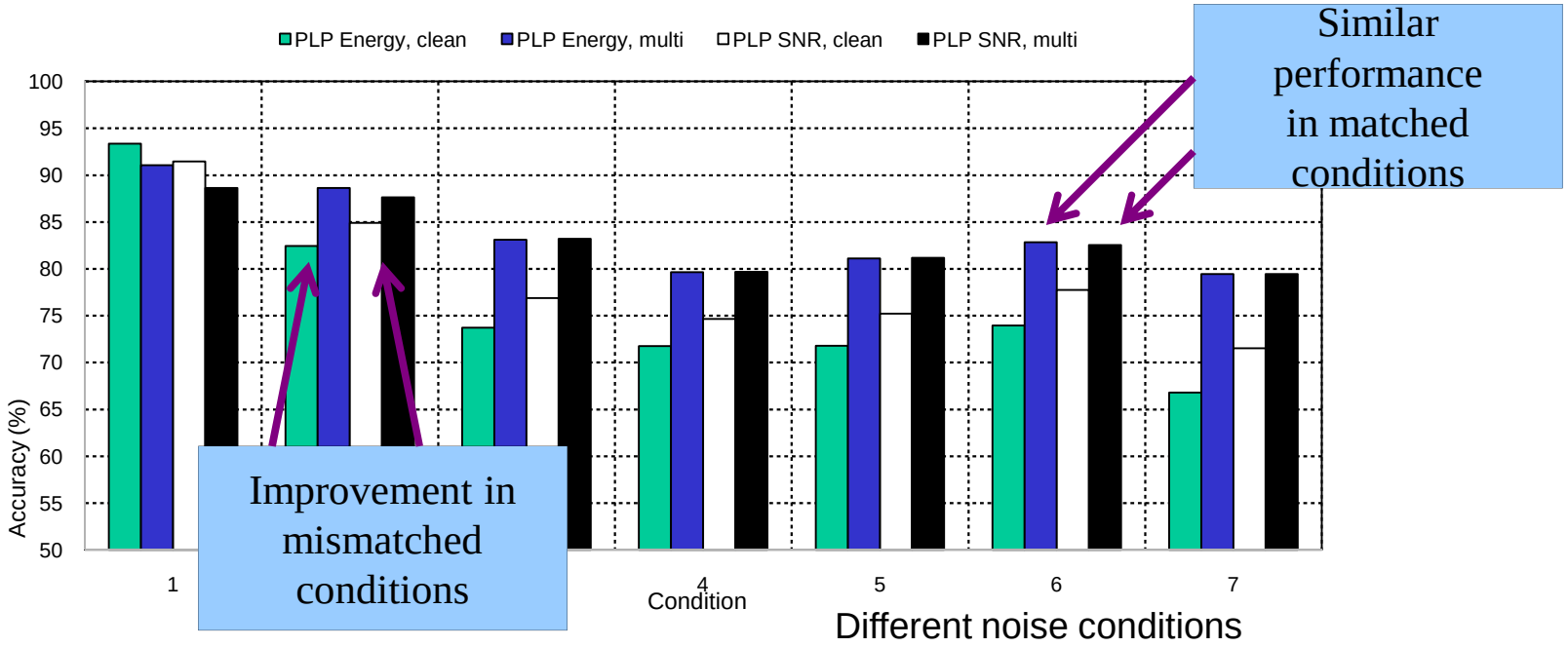
2. Scalable video coding



Subjective quality assessment of scalability options

SNR cepstrum for ASR

- The SNR cepstrum arises naturally from cepstral normalisation.
- Implies a way to handle ASR features in noise
  - The work will appear in a journal in the autumn



MLP-based speech processing

- Hierarchical MRASTA features for Mandarin ASR
- Hierarchical multilayer perceptron (MLP) based language identification system
- Mixed language speech recognition using universal phones
- HMM-based and template-based ASR using posterior features

ASR using posterior features

Acoustic class (e.g. phone) conditional probabilities estimated by MLP is directly used as feature observation in Kullback-Leibler divergence based hidden Markov model

- Grapheme-based ASR
- Integration of articulatory features
- Incorporation of multilingual information using multilingual phone posteriors
- Rapid training/adaptation of acoustic models

Research targeted towards monolingual, non-native, and multilingual speech recognition

Summary

- Advanced extensions of previous IM2 works
- Integration with applications (eg C-Board in IM2.IP2)
- Several demos of working systems
- Still fundamental work to support more applicative advances
  - ASR, Face/Object tracking, data mining,…

- We will continue converging towards applications
- And maintain the strong link with IP2

Thank you for your attention!

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

iM INTERACTIVE MULTIMODAL INFORMATION MANAGEMENT

Human Centered Design and Evaluation (IP2 of IM2 Phase 3)

D. Lalanne[1], A. Popescu-Belis[2], P. Dillenbourg[3], J. Sauer[1]

[1]University of Fribourg, [2]Idiap Research Institute, [3]EPFL

IM2 Advisory Board,
September 2, 2011

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

IM INTERACTIVE MULTIMODAL INFORMATION MANAGEMENT

Participants

- New IM2 partners
  - Prof. Pierre Dillenbourg, CRAFT / EPFL
  - Prof. Jürgen Sauer, Cognitive Ergonomics Group, University of Fribourg
- Partners from Phases I and II
  - to ensure smooth integration of new partners and sharing of technology between IP1 and IP2
  - Dr. Denis Lalanne, U. of Fribourg (*IP head*)
  - Dr. Andrei Popescu-Belis, Idiap (*deputy*)

Objectives of IP2

Overall goal

- Generalize and validate the technologies developed through research in Phases I and II of IM2
- Provide feedback and guidance to research in IP1

Approach

1. Develop new lightweight applications, mainly oriented towards teamwork spaces and learning
   - "Augmented Teams"
   - "Communication Board" (Cboard)
2. Carry out formal user-centered evaluations of new applications and IM2 multimodal technologies

The EmotiBoard, a Communication Board (CBoard) application

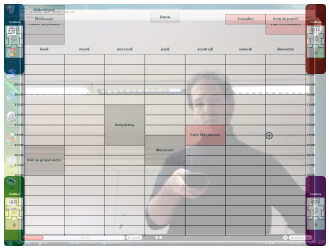
D. Lalanne[1], F. Ringeval[1], D. B. Jayagopi[2], D. Gatica-Perez[2], A. Sonderegger[1], J. Sauer[1]

[1]University of Fribourg, [2]Idiap Research Institute

IM2 Advisory Board,
September 2, 2011

CBoard

- The Communication Board is a vertical interactive surface on which multiple users can interact, in collocation or remotely
- An applicative framework that uses IM2 technologies
  - To set up novel real time technologies
  - To analyze users interactions to facilitate evaluation
- A framework to evaluate factors influencing collaboration (cognitive ergonomics group)



CBoard - Team

- The CBoard application involves three partners:
  - The DIVA group at Fribourg (F. Ringeval, D. Lalanne) aims at developing novel interactive technologies to facilitate team collaboration
  - The Cognitive Ergonomics group at Fribourg (A. Sonderegger, J. Sauer) is interested in studying the interaction of users with technology and with each other in a team context
  - The Social Computing group at Idiap (D. B. Jayagopi and D. Gatica-Perez), looks to move towards large-scale studies of group characterization and productivity analysis

CBoard - Set up

- A large (1m62 x 1m22) plexiglass display surface (with back-projection), on which users can interact simultaneously

- The CBoard framework supports the development of multi-user applications that can be accessed through multi-machines
  - Can interact using wiimotes or hand gestures, from the wall, a table or PC.
  - The system captures and transmits pointing devices positions and events between the machines (as well as audio/video)
  - A java library for easily developing CBoard applications has been created.

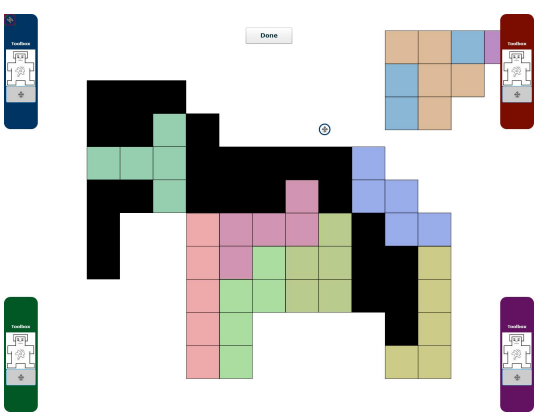
- First experiment comparing CBoard with paper-pencil interface suggested that the CBoard is beneficial for collaborative group work
  - with regard to group performance, task completion, frequency of interaction with system, team satisfaction, subjective evaluation of team performance

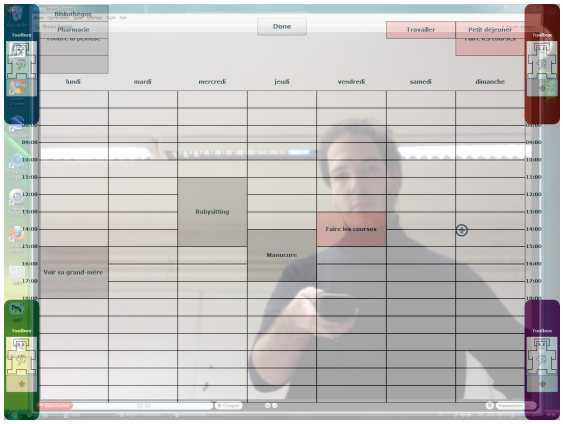
CBoard application: EmotiBoard – Study (1)

- Evaluate the influence of continuous feedback of team members' emotion on team performance and team satisfaction
- Experiment contains one highly aroused team member expressing bad mood (trained confederate)
- A 2 x 2 between-subjects design:
  - Emotion feedback (feedback provided vs. no feedback)
  - Remote group work (remote group work vs. collocated group work)
- Measures:
  - performance, satisfaction with teamwork, task load, conversational behavior (Microcone)
- Hypotheses:
  - Teams having an emotion feedback tool available will perform better
  - The bad mood of one team member will influence the outcomes of teamwork less in remote than collocated condition

EmotiBoard application – Study (2)

- Participants: 40 groups of 3 students
  - Each group must perform 3 tasks recorded using microcone
    1. Sensori-motor task (connecting dots)
    2. Spatial reasoning (placing jigsaw pieces into a figure)
    3. Coordination & planning (planning a day)





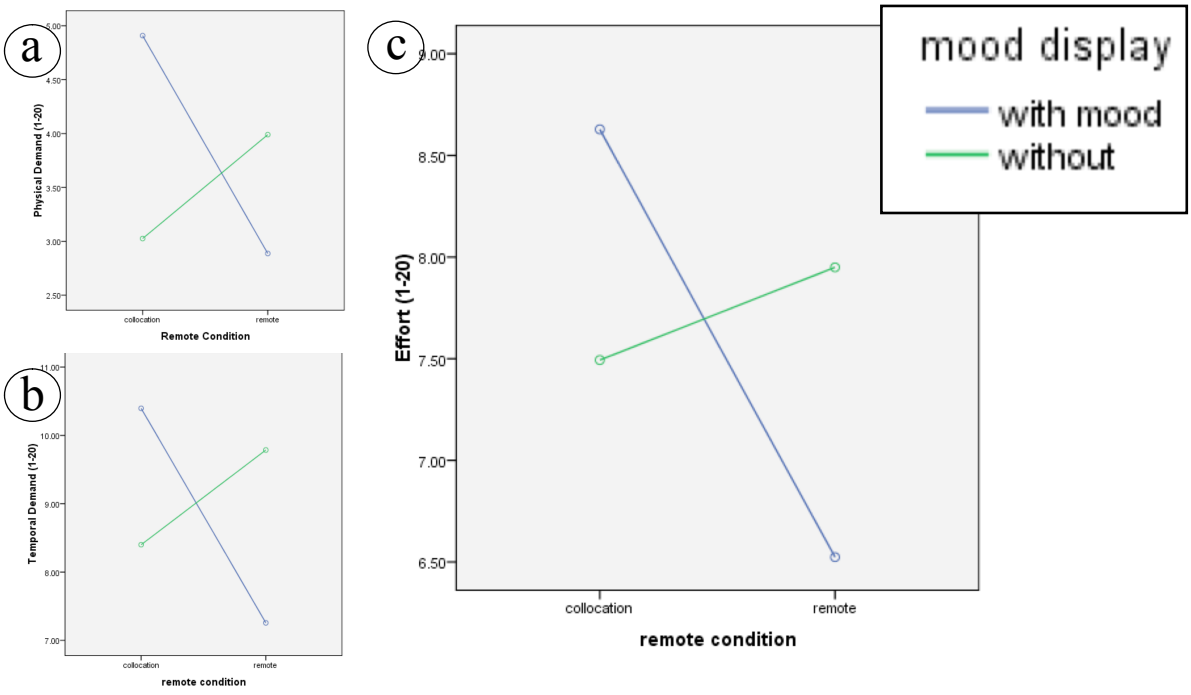
Conversational behavior analysis (Idiap)

- Automatic extraction of individual conversational behavior:
  - Turn-taking cues – speaking time, turns, interruptions and backchannels
  - Prosodic features – mean and variation of pitch and speaking rate

- Prediction tasks:
  - Using individual conversational behavior to predict bad mood
  - Using group conversational behavior to predict groups aware of the mood of its participants or (self-perceived) performance or satisfaction of participants

EmotiBoard – Preliminary results (1)

Interaction effects of mood display and remote group work on task load measures



EmotiBoard – Preliminary results (2)

- Conversational behavior analysis
  - Individual and group mood prediction difficult. In task 1 & 2, conversations too 'sparse'
  - In the future,
    - Analyze only segments where the group solves Task3 (planning a day)
    - Predicting self-perceived performance or satisfaction using group conversational behavior (to compare with results of questionnaires)

CBoard/EmotiBoard - Future work

- Analyze all the data (quantitative and also results from conversational behavior) from the experiment
- Recognize and display emotional state, and mood in real-time (F. Ringeval, UniFr)
- Track and display eye-gaze (collab with B. Noris, A. Billard, EPFL)?
- Enhance remote collaboration with real-time mood display (and visual focus of attention)
  - Will consider tasks requiring more conversations
- Run novel experiments

Augmented Teams

Idiap Research Institute: *A. Popescu-Belis, A. Nanchen*
EPFL CRAFT: *P. Dillenbourg, F. Kaplan, O. Mubin, N. Li*

IM2 Advisory Board,
September 2, 2011

FNSNF Fonds national suisse Schweizerischer Nationalfonds Fondo nazionale svizzero Swiss National Science Foundation
IM INTERACTIVE MULTIMODAL INFORMATION MANAGEMENT

IP2 Augmented Teams

- Objective
  - assess the utility of the Automatic Content Linking Device in a collaborative environment at the EPFL Rolex Learning Center

- AT application = ASR + ACLD + tabletop interface
  - listen to a discussion: student group work in a dedicated space
  - display a representation of words recognized using ASR
  - allow participants to use these words for queries
  - launch queries automatically or on demand to various repositories such as course material or the Web
  - display results and let participants use them

Software developments for the ACLD

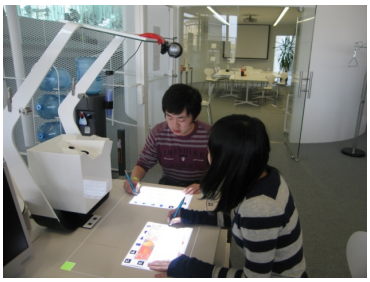
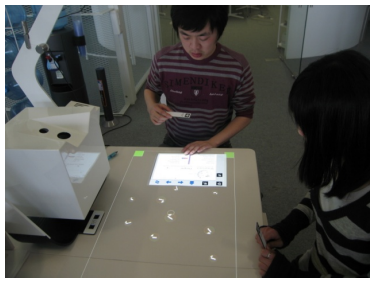
- Tuning the Idiap ASR system for the ACLD / CRAFT
  - documenting low-level settings for real-time ASR
  - experiments with tuning to improve accuracy at CRAFT
  - diagnosis of low accuracy observed in the CRAFT setting
  - vocabulary adaptation: with ‘lexgen’, awaits new word lists

- Streamlining architecture of ACLD
  - UI merged with other modules, no longer uses Hub
  - improved performances, removed undesired behaviors

  - Installation on powerful laptop (MacBook Pro, i7, 8 GB)
  - solving compilation issues for ASR / audio libraries
  - comparing microphones for EPFL setting

Augmented Teams experimental setup

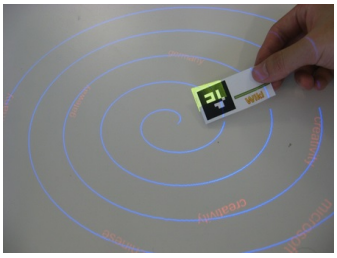
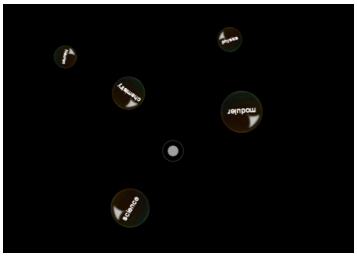
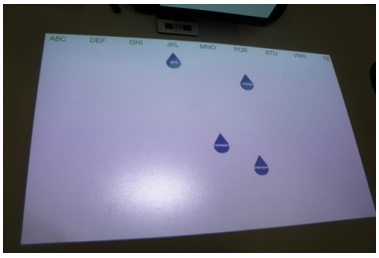
- Tabletop interface: CRAFT Tinkerlamp projector (displays words and documents) and camera (recognizes paper and tags), plus an infrared camera (recognizes pens)
- Words recognized by the ASR are projected on the table, where they can be manipulated using tags, then used for search
  - alternatively, a keyboard can be used to enter words too
- Results are projected on a paper browser where they can be consulted, stored, etc. (infrared pens: scroll, click, etc.)


Words in the tabletop interface

- **Goal**: find experimental evidence for the use of just-in-time speech-based retrieval (ACLD)
- **Method**: compare manual keyboard-based search *vs.* on-demand ASR-based *vs.* automatic ASR-based

➔ Importance of word display and manipulation
  – spiral | bubbles | raindrops (sorted)

Initial experiments: conditions and results

- **First run**: two people discussing a text about neurons (2 groups)
  - Condition 1: only keyboard to enter queries for documents
  - Condition 2: automatic searches with words from the conversation (+kb)
    - Wizard-of-Oz instead of ASR to ensure accuracy
- Observations
  - "automatically" (WOz) recognized words almost never used for search
    - difficulties with word representation? (= bubbles moving randomly on the table)
  - paper browsers found natural and convenient, some issues to solve
- **Second run**: two people placing new power plants on a map
  - foldable paper display, keyboard can be "connected" to a display
  - words as raindrops, alphabetical order
- Observations
  - captured words still did not lead to many useful queries
  - users often intentionally uttered words for queries, or used the keyboard
    - need for using multi-word phrases for search?
  - overall interaction flow seems promising, more issues were solved

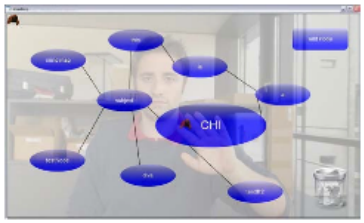
Augmented Teams: perspectives

- Experiments at CRAFT, with assistance from Idiap
  - reach optimal setup of ASR for best possible recognition
  - test ACLD mechanism with improved visualization of results
  - give well-justified quantitative assessments of the utility (in the Augmented Teams setting) of:
    - ASR for capturing words from a conversation
    - ACLD for running automatic queries
    - tabletop interface, paper browser, keyboard-based search, etc.
- Developments at Idiap
  - PhD student on IP2 starts in September 2011
  - improve ACLD interactivity: human-computer dialogue
  - benefit from CRAFT experience in user-centric design

Use of IM2.IP1 technology in IP2

- Several joint meetings have been organized between IP1 and IP2 in order to support the collaborations going on between the two IPs.

- The following technologies are integrated into IP2 applications:
  - Automatic Speech Recognition (in Augmented Teams)
  - Content Linking - ACLD (in Augmented Teams)
  - Indexing and retrieval (in Augmented Teams)
  - Multimodal interaction (in CBoard & Augmented Teams)
  - Pointing devices and gestural interaction (in CBoard)
  - Speaker turns (in CBoard)
  - Behavioral data mining (in CBoard)
- Other technologies planned to be integrated:
  - Real-time Emotion Recognition (in CBoard)
  - Visual Focus of Attention (in CBoard)



Wrap Up

- Two novel applications were set up and have started producing evaluation results from user-oriented experiments
  - Communication Board
  - Augmented Teams

- IP2 encourages researchers from IP1 to provide technologies in order to:
  - Improve applications
  - Help analyzing interaction sessions with applications

Thank you for your attention!

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

iM INTERACTIVE MULTIMODAL INFORMATION MANAGEMENT

IM2 Phase III
IP3: Social Signal Processing

IP Head: A.Vinciarelli (Idiap Research Institute)

Deputy IP Head: F.Valente (Idiap Research Institute)

IM2 Advisory Board,
September 2, 2011

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

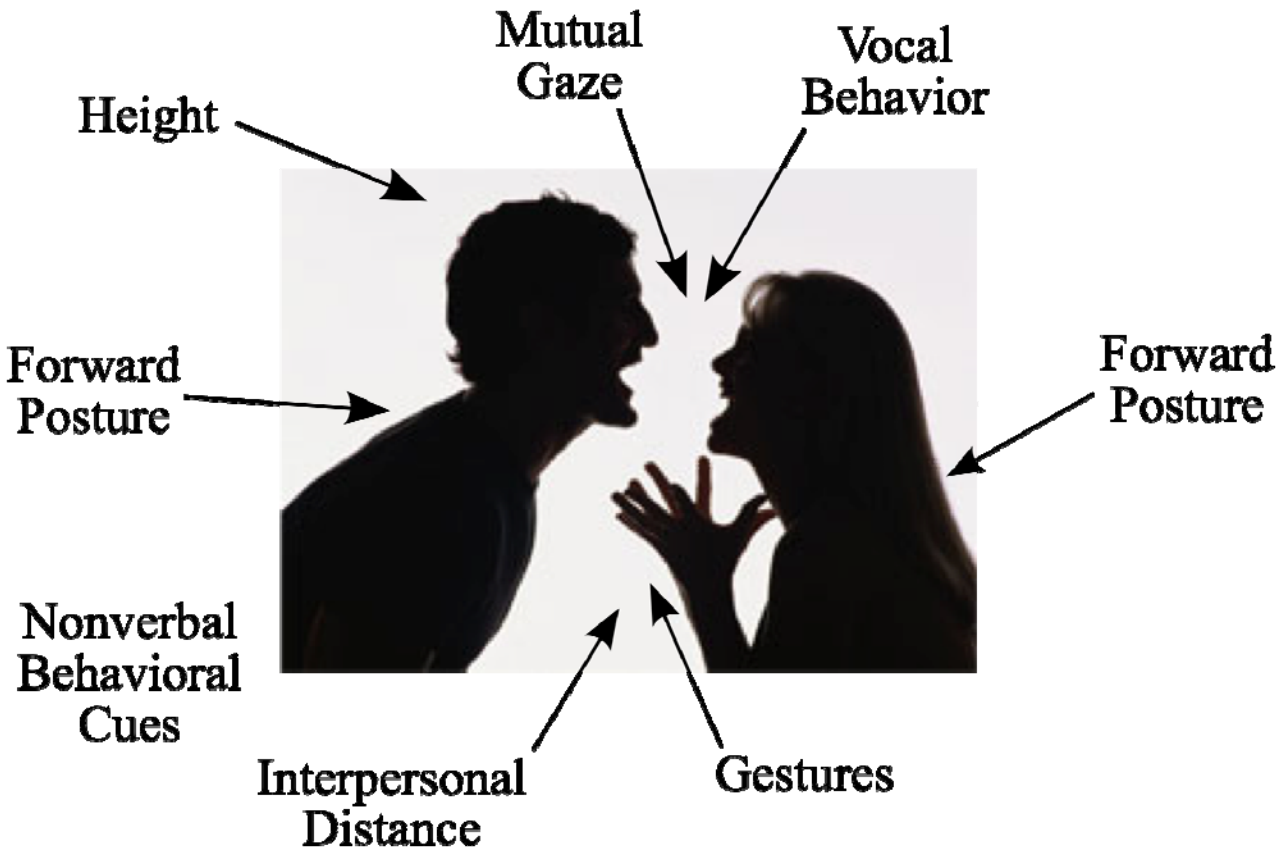
iM INTERACTIVE MULTIMODAL INFORMATION MANAGEMENT

Outline

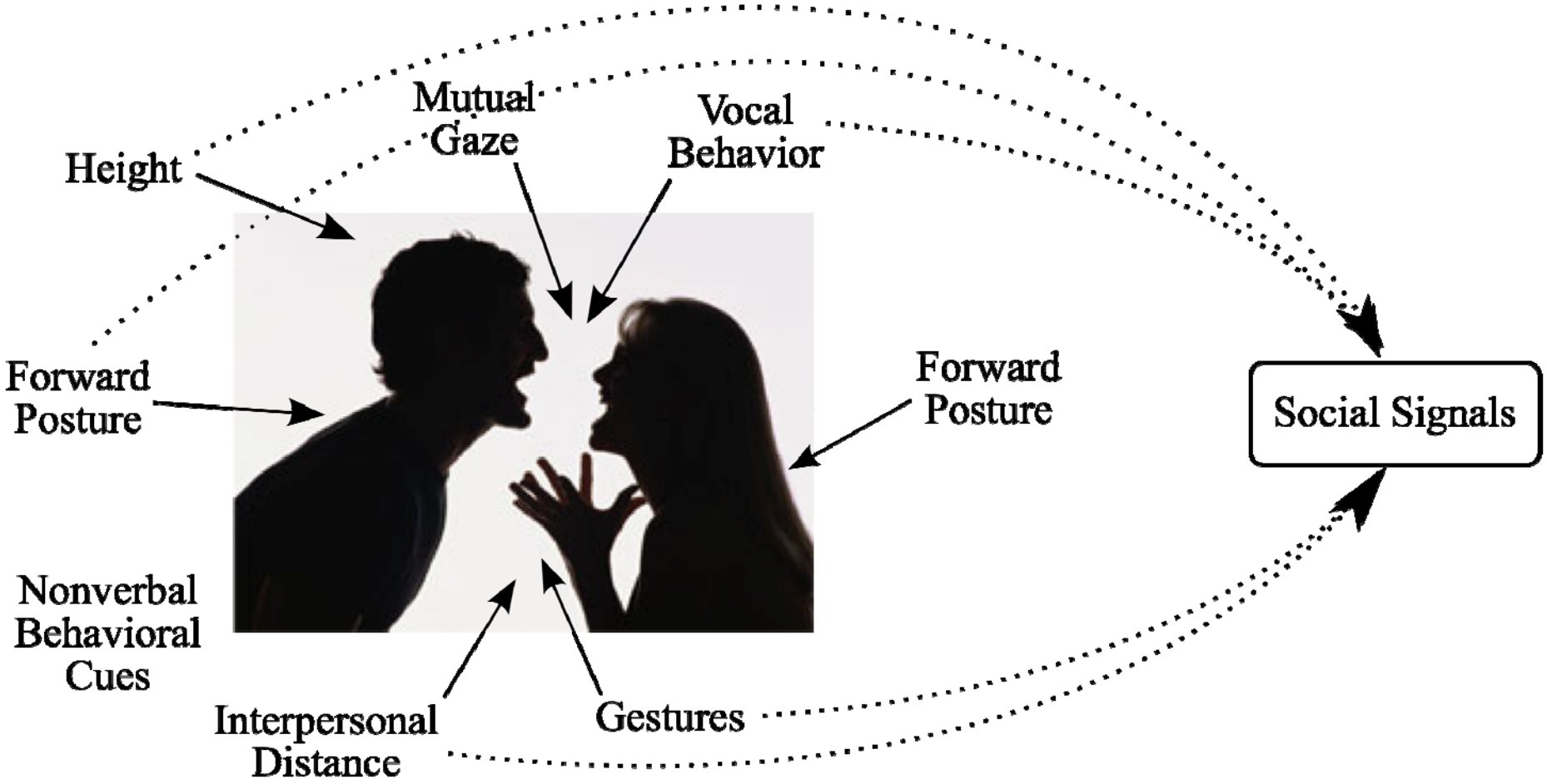
- What is Social Signal Processing?
- Participants and Goals
- Main Results
- Networking and References

Nonverbal Behavior

Nonverbal Behavior



Nonverbal Behavior



Social Signal Processing

SSP studies signals that:

- are produced during social interactions
- that either play a part in the formation and adjustment of relationships and interactions between agents (human and artificial)
- or provide information about the agents
- and that can be addressed by technologies of signal processing and synthesis.

The \Belfast Declaration": http://sspnet.eu/about/

Social Signal Processing

SSP studies signals that:

- •are produced during social interactions
- •that either play a part in the formation and adjustment of relationships and interactions between agents (human and artificial)
- •or provide information about the agents
- •and that can be addressed by technologies of signal processing and synthesis.

The \Belfast Declaration": http://sspnet.eu/about/

Participants and Collaborators

IM2 Participants:

- Idiap Research Institute (F.Valente)
- ETHZ (A.Fossati)

External Bodies:

- NCCR Affective Sciences (M.Mortillaro)
- EC NoE SSPNet (A.Vinciarelli)
- Italian Institute of Technology (V.Murino)
- U. of Verona (M.Cristani)

Research Tracks and Goals

Two main research tracks:

- Engagement
  - •Recognition of social roles
  - •Proxemics Understanding
- Effectiveness of Delivery
  - •Automatic Personality Perception

Main Goal: Developing unimodal approaches relevant to the tracks

Main Results

- Role recognition approaches for several settings, (broadcast data, political debates and meetings)
- Prosody based approaches for Automatic Personality Perception
- Speaker diarization approaches based on the combination of acoustic and role related information
- Approaches for automatic proxemics understanding (F-formations and social distance)

Roles

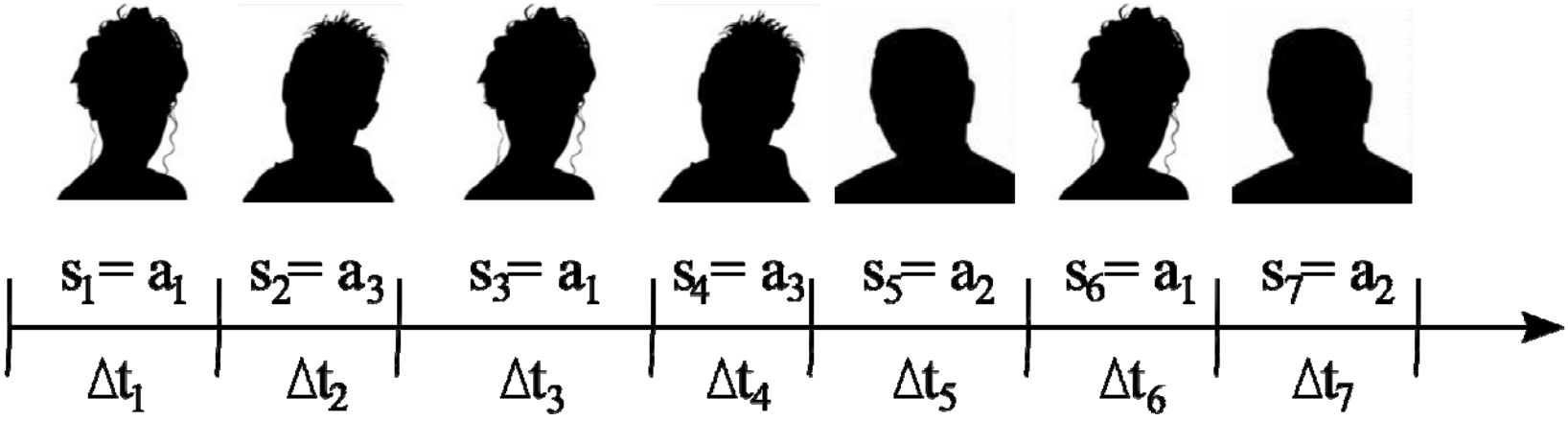
"*Role theory concerns one of the most important features of social life, characteristic behavior patterns or roles [...] persons are members of social positions and hold expectations for their own behaviors and those of other persons*"

B.J.Biddle, "*Recent Developments of Role Theory*", Annual Review of Sociology, 12:67-92 (1986)

Expectations

- Norms: explicit prescriptions about the behavioral patterns associated to a given role (evident in broadcast data)
- Beliefs: subjective choices on how a role should be performed (evident in AMI meeting roles)
- Preferences: spontaneous choices based on personality traits or attitudes (evident in social roles)

Roles in Broadcast Data



The sequence of turns is mapped into a sequence of roles using a Conditional Random Field:

$$R^* = \arg\max_R p(R \mid S)$$

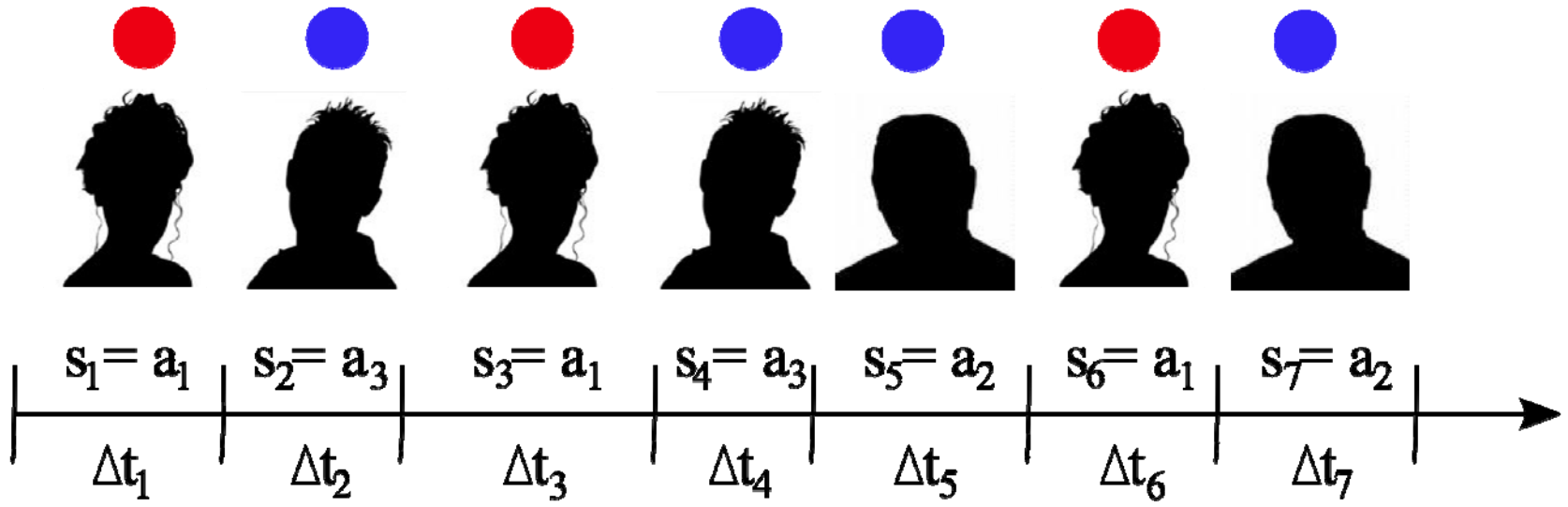
Where $S = \{(s_i, t_i, \Delta t_i, X_i)\}$

Results

Corpus	P	T	P+T
C1 (A)	83.0%	89.7%	89.3%
C2 (A)	69.5%	84.2%	**87.0%**
C1 (M)	87.1%	99.1%	99.1%
C2 (M)	76.2%	96.9%	96.2%

- C1: News (20 hours)
- C2: Talk-Shows (26 hours)
- Roles: Anchorman, Guest, Weatherman, Interviewer(-ee), Headline Reader

Roles in Meetings



Consider a richer description of the turns:

$$S = \{(s_i, t_i, \Delta t_i, X_i, f_i)\}$$

Where f_i corresponds to the "*AMI role*"

Social Roles

Corpus	All	PR	SU	NE	GA
AMI-5	68.0%	72.0%	65.0%	80.0%	15.0%

- The sequence of turns is mapped into a sequence of social roles (Protagonist, Supporter, Neutral, Gatekeeper) using Dynamic Bayesian Networks
- AMI-5 is a collection of 5 AMI meetings annotated in terms of social roles

Personality

"*[Personality is the latent construct accounting for] individuals' characteristic patterns of thought, emotion, and behavior together with the psychological mechanisms - hidden or not - behind those patterns*"

D.Funder, "Personality", Annual Review of Psychology, 52:197-221 (2001).

The Big-5 (I)

"*The Big Five Personality Factors appear to provide a set of highly replicable dimensions that parsimoniously and comprehensively describe most phenotypic individual differences*"

G.Saucier and L.R.Goldberg, "The Language of Personality: Lexical Perspectives on the Five-Factor Model", in "The Five-Factor Model of Personality", J.S.Wiggins (ed.), pp. 21-50 (1996)

The Big-5 (II)

- Extraversion: Active, Assertive, Energetic, Outgoing, Talkative
- Agreeableness: Appreciative, Forgiving, Generous, Kind, Sympathetic, Trusting
- Conscientiousness: Efficient, Organized, Planful, Reliable, Responsible, Thorough
- Neuroticism: Anxious, Self-pitying, Tense, Touchy, Unstable, Worrying
- Openness: Artistic, Curious, Imaginative, Insightful, Original, Wide interests

Results

Trait	N≥6	N≥7	N≥8
Extraversion	73.0 (100.0)	76.3 (77.6)	78.7 (57.3)
Agreeableness	62.7 (100.0)	63.7 (67.2)	69.5 (40.9)
Conscientiousness	72.7 (100.0)	78.8 (67.0)	82.4 (38.1)
Neuroticism	67.7 (100.0)	70.3 (68.4)	74.3 (38.9)
Openness	59.8 (100.0)	68.8 (55.1)	74.0 (22.8)

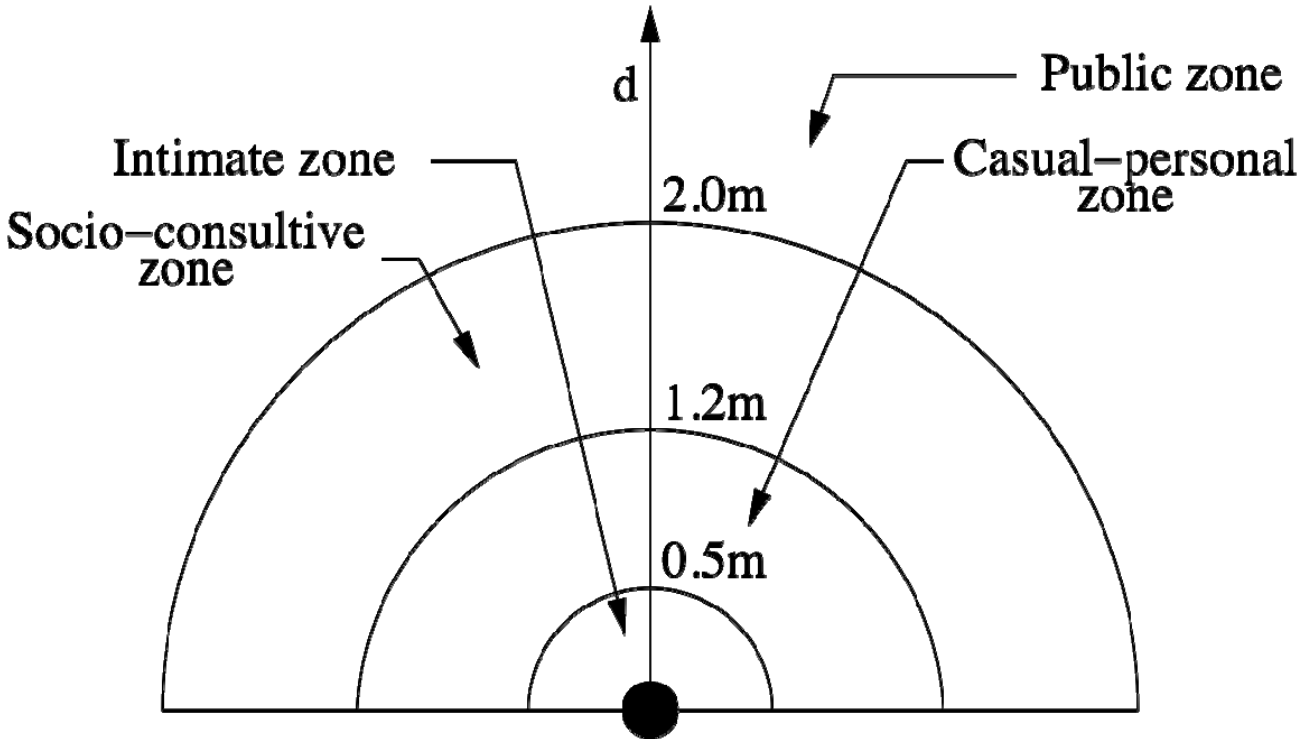
640 Audio Clips (10 seconds) assessed by 11 assessors (the same for all clips) using the BFI-10 questionnaire

The Big-5 (I)

"*[...] the study of man's transactions as he perceives and uses intimate, personal, social and public space in various settings [...]*"

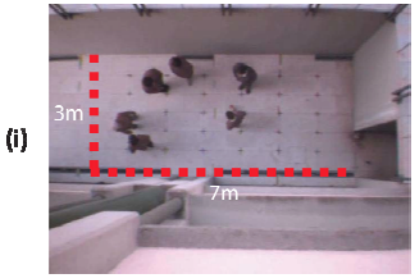
E. Hall, "The hidden dimension", Doubleday, NY, (1966)

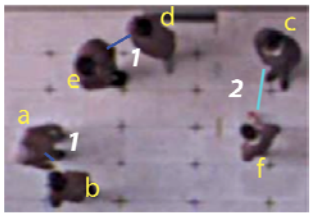
The Four Zones

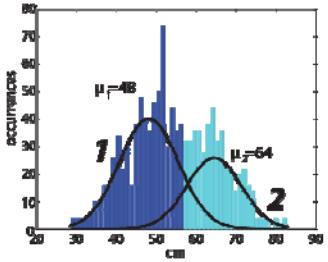


People divide their personal space in four concentric areas corresponding to different levels of intimacy

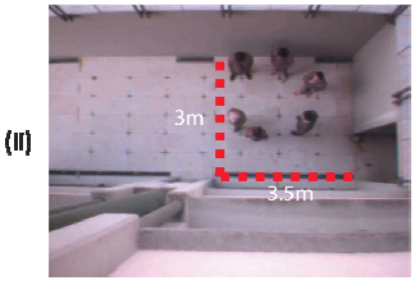
F-formations and Social Distance

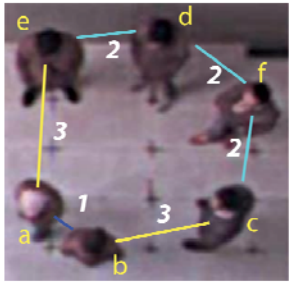


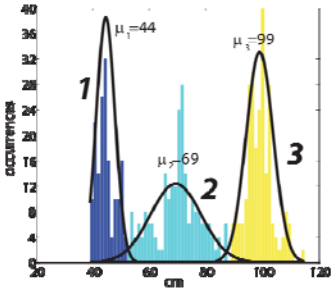




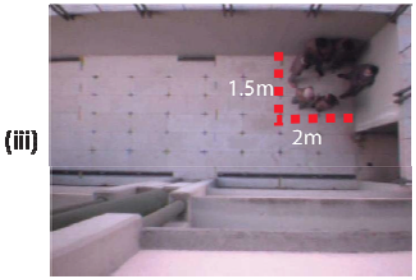
Class / Couple	1μ= 48cm, [29, 55])	2μ= 64cm, [55, 82])
ab	0.68	0.32
cf	0.35	0.65
de	0.85	0.15



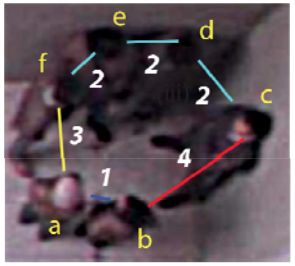




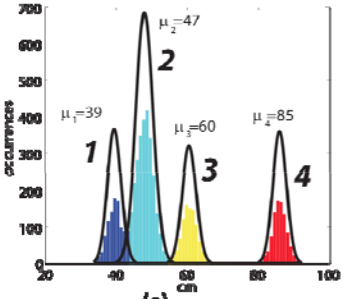
Class / Couple	1μ= 44 cm, [38, 51])	2μ= 69cm, [51, 86]	3μ= 98cm, [86,114])
ab	0.96	0.04	0
bc	0	0.04	0.96
cf	0.28	0.70	0.02
fd	0.26	0.39	0.35
de	0	0.89	0.11
ea	0	0	1



(a)



(b)



(c)

Class / Couple	1μ= 39cm, [33, 43])	2μ= 47cm, [43, 54])	3μ= 60cm, [54, 65])	4μ=85cm, [80, 92])
ab	0.96	0.04	0	0
bc	0	0	0	1
cd	0.01	0.99	0	0
de	0.08	0.92	0	0
ef	0.06	0.94	0	0
fa	0	0.01	0.99	0

(d)

Networking Activities

- Several activities result from the collaboration including both IM2 and external bodies
- The IP head has setup a IEEE Technical Committee on SSP (in collaboration with M.Pantic and A.Pentland)
- Idiap,IIT and U. of Verona co-organize the International Workshop on Socially Intelligent Surveillance and Monitoring
- Limited collaboration with other IPs

References

- F.Valente et al., "*Speaker Diarization of Meetings based on Speaker Role N-gram Models*", Proc. of ICASSP, 2011
- A.Vinciarelli et al., "*Understanding Social Signals in Multi-party Conversations: Automatic Recognition of Socio-Emotional Roles in the AMI Meeting Corpus*", Proc. of IEEE Intl. Conf. on SMC, 2011
- M.Cristani et al., "*Towards Computational Proxemics: Inferring Social Relations from Interpersonal Distances*", Proc. of IEEE Intl. Conf. on Social Computing, 2011
- S.H.Yella et al., "*Information Bottleneck Features for HMM/GMM Speaker Diarization of Meetings Recordings*", Proc. of Interspeech (nominated best paper), 2011

Thank you for your attention!

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

INTERACTIVE
MULTIMODAL
INFORMATION
MANAGEMENT

Structural impact and activities beyond the NCCR (AIM2)

Hervé Bourlard, IM2 Director
Idiap Research Institute
Touradj Ebrahimi, IM2 Deputy Director
EPFL

IM2 Advisory Board,
September 2, 2011

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

INTERACTIVE
MULTIMODAL
INFORMATION
MANAGEMENT

Outline

- Structural impact achieved in Phase I & II
- Expected structural impact in Phase III
- AIM2: Beyond IM2

Structural Impact in Phase I&II

- On the Leading House (idiap):
  - From small to large, international, research institute
  - Decision by the Federal Government (SER, Art. 16 of the Law on Research) to establish for Idiap a strategic alliance with the EPF-ETH domain, since January 2008, with significant federal funding increase. This was conditional on the negotiation and signature of a joint Idiap-EPFL development plan.
  - Idiap-EPFL joint development plan (signed in July 2008).
  - Nomination of a first joint Idiap-EPFL Professor Assistant Tenure Track (PATT) with the clear intention of both Idiap and EPFL to open a new call for additional PATT joint positions;
  - Creation of a new academic title at EPFL, the external MER, especially targeted at young, promising Idiap scientists;
  - Various improvements in the academic conditions for Idiap PhD students and researchers, etc.
- At EPFL:
  - Creation of a new Electrical Engineering Institute within the School of Engineering at EPFL, featuring almost all EPFL IM2 partners;
  - Creation of a new Electrical Engineering Doctoral Program at EPFL (EDEE), partly aligned on the research and education topics of IM2;
  - One new professor chair in Brain-Machine Interface
  - Several tenured-track professors involved in IM2 got tenured
  - Young EPFL professors involved in IM2 (more in IM2-III)
  - Set up an audiovisual multimodal test and evaluation laboratory at the Multimedia Signal Processing Group

Structural Impact in Phase I&II

- At ETHZ:
  - Prof. Vittorio Ferrari, new ETHZ/Computer Vision Lab professor, integrated into the IM2 team for the 3rd phase.
  - New computer infrastructure for multi-modal processing will be fully installed and operational, for calculation intensive processing of longer videos and movies
- At UniFri:
  - New Human-IST institute conducting multidisciplinary research in the field of human-machine Interaction
- At UniGe:
  - Computer Science Department: creation of a permanent "Maître d'enseignement et de recherche" position;
  - Faculty of Sciences and UniGe: associate professor salary for Jan. 2009 to Oct. 2010 until the next CS Dpt. professor retirement, to target a chair in the IM2 domain.

Structural Impact in Phase I&II

- All IM2 partners are involved in multiple relevant and IM2-related major EU projects, collaboration with industry, and spin-offs. Most recently:
  - SSPNet: Social Signal Processing Network (EU-NoE) to be started in Feb.09, is the first and only one big project in this new area
  - PetaMedia has IM2 itself as one of its partner.
  - Submission of a very competitive EU KIC proposal

    (budget: 100 MEuros/year over 12 years)

Structural impact expected in IM2-III

- **Leveraging on already achieved scientific and technical assests and know-how, to establish visible and ambitious collaborative projects**
  - EPFL Learning Center as a living lab environment to implement and evaluate ideas
  - Positioning as the EU provider of multimodal data, software and evaluation measures
  - Establishment of multimodal test and evaluation infrastructures

- **Increase excellence in science, TT and education**
  - Quantity and quality of publications,
  - Number of national and international research projects
  - Technology transfer activities, creation of successful spin-offs
  - Contribute to already existing and establish new educational curricula

THREE QUESTIONS TO IM2 ADVISORY BOARD:
- What other structural impacts should be targeted as outcome of IM2-III?
- Which strategies to successfully achieve them?
- How best communicate them to SNSF?

Beyond IM2

- Started developing ideas already in Phase II
- All partners fully motivated to build a lasting community around IM2
- The community should:
  - Be larger than the current IM2 partners,
  - Welcome and provide support/services to all interested parties involved in multimodal information management,
  - Promote activities in the IM2 related fields.
  - Start already during IM2-III to increase chances of survival beyond IM2

AIM2

– Association for Interactive Multimodal Information Management

AIM2 Objectives

“ Promote the field of Interactive Multimodal Information Management by strenghtening the corresponding education, research, development and dissemination activities”

- Supports its community by:
  - Setting up and conducting research and development projects
  - Encouraging consulting and technology transfer
  - Mediating exchanges of data corpora, software and know-how
  - Promoting teaching and educational activities
  - Facilitating dissemination of knowledge through organization of events and publications

AIM2: Some initial thoughts on needs

- Generic: Develop more business and research opportunities
- For academia:
  - Having access to a network and needs/technologies of industries
  - Support of grant proposals
  - PhD students:
    - Reduced conference fees? Student grants/scholarship?
    - Networking/easier to get a job; possibility to post CVs, etc.
  - Awards
    - Recommendation to conference awards, etc.
- For industry:
  - Having access to state-of-the-art research, network, manpower, pool of highly-qualified personnel
  - (Preferential?) access to research results for commercial activities
    - Should be beyond what they already have access to through conferences, etc.
    - Possibility of steering research directions
    - Mini-projects
    - AIM2 support statements for CTI projects
  - Swiss vs international (Switzerland too small)
- For individuals: mix of the above

Current status of AIM2

- The AIM2 Manifesto
  - Goals
  - Major Actions
    - R&D projects
    - Technology transfer
    - Data and technology exchanges
    - Education
    - Dissemination
  - Membership types and fees
    - Institutional (1'000.- CHF)
    - Corporate members (300.- CHF to 2'000 CHF)
    - Individual members (200.- CHF)
    - Honorary members (free)

Current status of AIM2

- The AIM2 Statutes
  - Name and Seat
  - Purpose
  - Membership issues
    - Categories
    - Admission
    - Resignation
    - Duration
  - Organizational issues
    - Offices
    - The administrative assembly
    - Invitation to administrative assembly
    - Duties of administrative assembly
    - Voting rights
    - The Board
    - Duties of the Board and its operation
    - Auditor
    - Supporting entities
    - Property rights
    - Finances
    - Dissolution

Current status of AIM2

- Domain name purchased
  - www.aim2.org
  - www.aim2.ch
- Website in progress (80% completed)
- Transfer of 50 KCHF to set up and initiate AIM2 during Phase III
- To be done:
  - Complete website and make publicly accessible
  - Identify potential key actors (head, board, treasurer, etc.)
  - Leaflet
  - Dissemination and launch

QUESTION TO IM2 ADVISORY BOARD:

- What are your feedback and recommendations on AIM2?

Thank you for your attention!

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

INTERACTIVE
MULTIMODAL
INFORMATION
MANAGEMENT