Bringing Social Computing to Ambient Environments: Synergies and Challenges

Elisabeth André
Augsburg University, Germany
http://hcm-lab.de

Human-Centered Multimedia

Universität
Augsburg
University

- **Founded:** April 2001
- **Chair:** Elisabeth André
- **Research Topics:**
  - Embodied Conversational Agents
  - Perceptive User Interfaces
  - Affective Computing
  - Interactive Storytelling
- **Study Programs**
  - BSc/MSc Informatics
  - BSc/MSc Informatics and Multimedia
  - Elite Graduate Program Software Engineering
  - BA Media and Communication

National and International Projects

- **Affective Computing**
  - Humaine, CALLAS, CEEDS, Ilhaire, TARDIS
- **Technology-Enhanced Learning**
  - CUBE-G, DynaLearn, eCUTE, e-Circus, TARDIS
- **Multimodal Interaction, Behavior Analysis**
  - IRIS, OC Trust, CEEDS, TARDIS
- **E-Health**
  - Metabo
- **Smart Energy**
  - IT4SE

humaine
emotion-research.net

callas

IRIS

DynaLearn

TARDIS

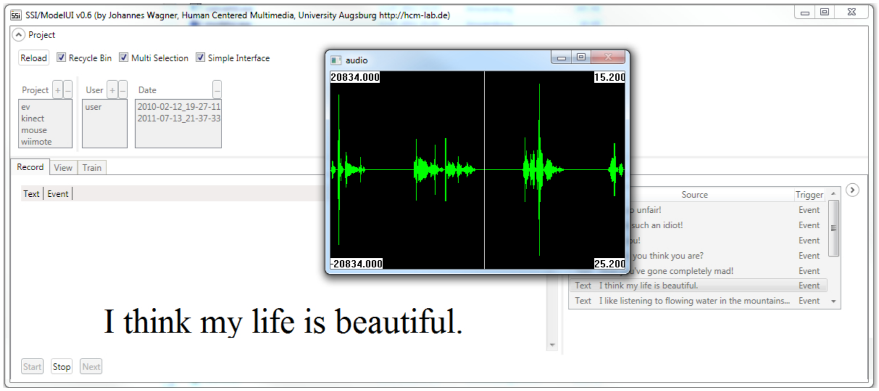
Education in Cultural Understanding, technology enhanced

TRUST

ILHAIRE
The science of laughter

Motivation

- **Mutual Gains and Benefits**
  - **AMI environments:**
    - unobtrusive sensors that let us collect subtle behavioral cues under naturalistic conditions
    - usually focus on context and user activity data
    - reasoning mechanisms
    - typically mobile environments
  - **Social Signal Processing:**
    - techniques for analyzing and interpreting behavioral cues and linking them onto higher-level psychological concepts, such as emotions and personality
    - focus on psychological user states
    - typically desktop environments



SSI/ModelUI v0.6 (by Johannes Wagner, Human Centered Multimedia, University Augsburg http://hcm-lab.de)

☑ Recycle Bin ☑ Multi Selection ☑ Simple Interface

Project: ev, kinect, mouse, wiimote

User: user

Date: 2010-02-12_19-27-11, 2011-07-13_21-37-33

Source	Trigger
...o unfair!	Event
... such an idiot!	Event
...u!	Event
... you think you are?	Event
...u've gone completely mad!	Event
Text I think my life is beautiful.	Event
Text I like listening to flowing water in the mountains...	Event

I think my life is beautiful.

Conscious vs. Unconscious Interaction

- **Conscious Interaction:**
  - Open interaction with a system where a user intentionally inputs discrete commands to explicitly express his needs
  - Example: Language, Pointing …
- **Unconscious Interaction:**
  - Continuous (often nonverbal) behavior the user does not voluntarily control, but which may be interpreted as the implicit expression of a particular need
  - Example: non-acted facial expressions and body postures
- **Role of Context:**
  - Both in the case of conscious and unconscious interaction, ambiguities need to be resolved by context modeling

Outline of the Talk

- Examples of unconscious behaviors in human-machine interaction
- Unconsciously expressed social and emotional behaviors
- Problems with traditional machine learning approaches and potential solution strategies
- Agenda for future research

Distinction between Conscious and Unconscious Signals

- We got rid of the Push-to-Talk?
- But what about …

Distinction between Conscious and Unconscious Signals

Push-to-Touch?

Push-to-Gesture?

Does the boy rest his right arm on the table or conduct a command gesture?

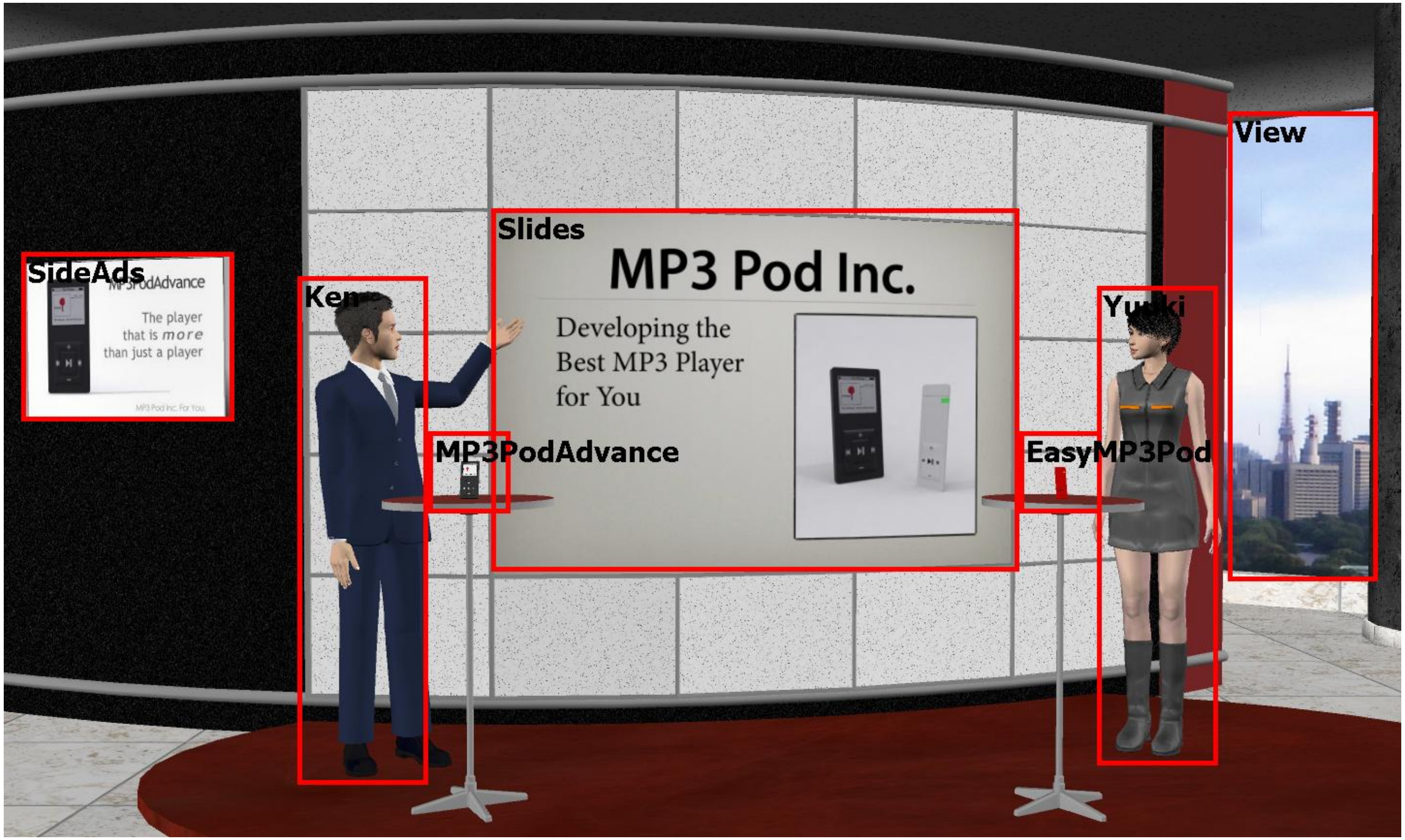
Is the left user just raising his hands (out of desperation) or conducting a command gesture?

Unconscious and Conscious Interaction

- **Home Entertainment System**
  - Distinction between command and no-command gestures (for example, greeting gestures)
  - Distinction between conscious (command) and unconscious signals (for example, scratching one’s head)
  - Automatically interrupts presentation when the users’ interest is diverted.

Unconscious Gaze-Based Interaction

- Agents adapt presentation implicitly to the user’s attention as inferred from his or her eye gaze



Michael Nischt, Helmut Prendinger, Elisabeth André, Mitsuru Ishizuka: MPML3D: A Reactive Framework for the Multimodal Presentation Markup Language. IVA 2006: 218-229

Results of a User Study

- **Human-like Conversation:**
  - Participants interacting with the gaze-based agents felt that the agents were aware of them
  - Participants interacting with “blind” agents thought that the agents react to them in a strange way

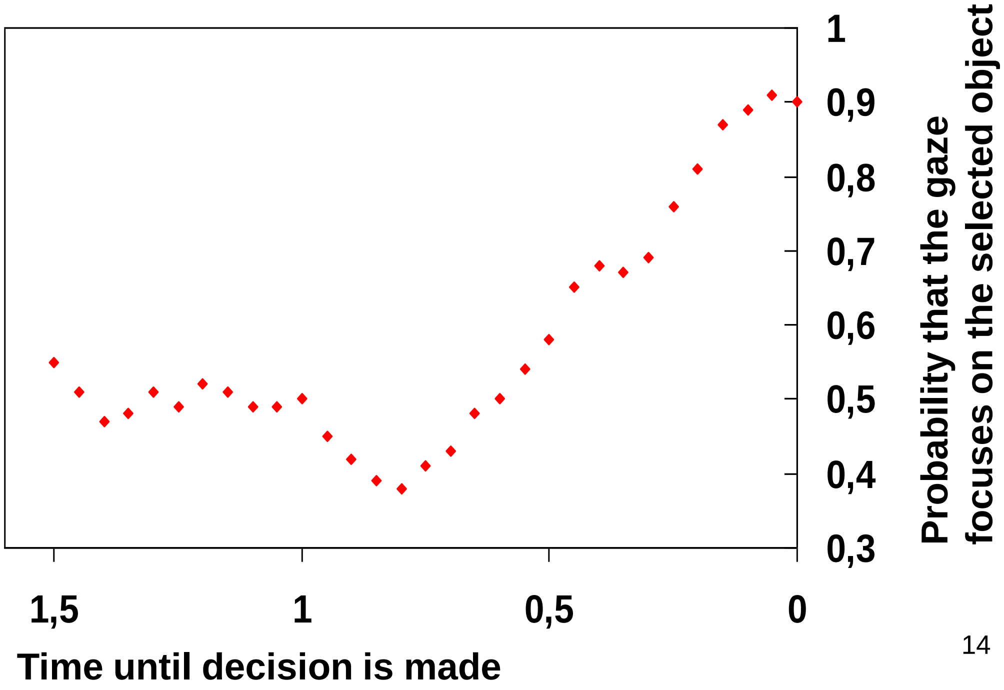
- **Midas Touch Problem:**
  - interface should not react to each change of fixation
  - risk of „overdoing“ attentiveness
  - User starts to adopt unnatural gaze behaviors

Helmut Prendinger, Tobias Eichner, Elisabeth André, Mitsuru Ishizuka: Gaze-based infotainment agents. Advances in Computer Entertainment Technology 2007: 87-90

Implicit Prediction of User Preferences from Gaze

- **Question:**
  - Is it possible to predict based on the user‘s gaze behavior which one of two objects he or she prefers?
  - Shimojo & Simion (CALTECH) analyzed gaze behaviors 1.5 s before a selection (by pressing a button) was made.

- **Gaze Cascade Effect:**
  - Probability that the user focuses on the preferred object increases continuously



Implicit Prediction of User Preferences from Gaze

- Transfer of this work on the selection of ties
- In 81% of the cases, the preferred tie was correctly predicted.
- Better results for similar than for different ties.

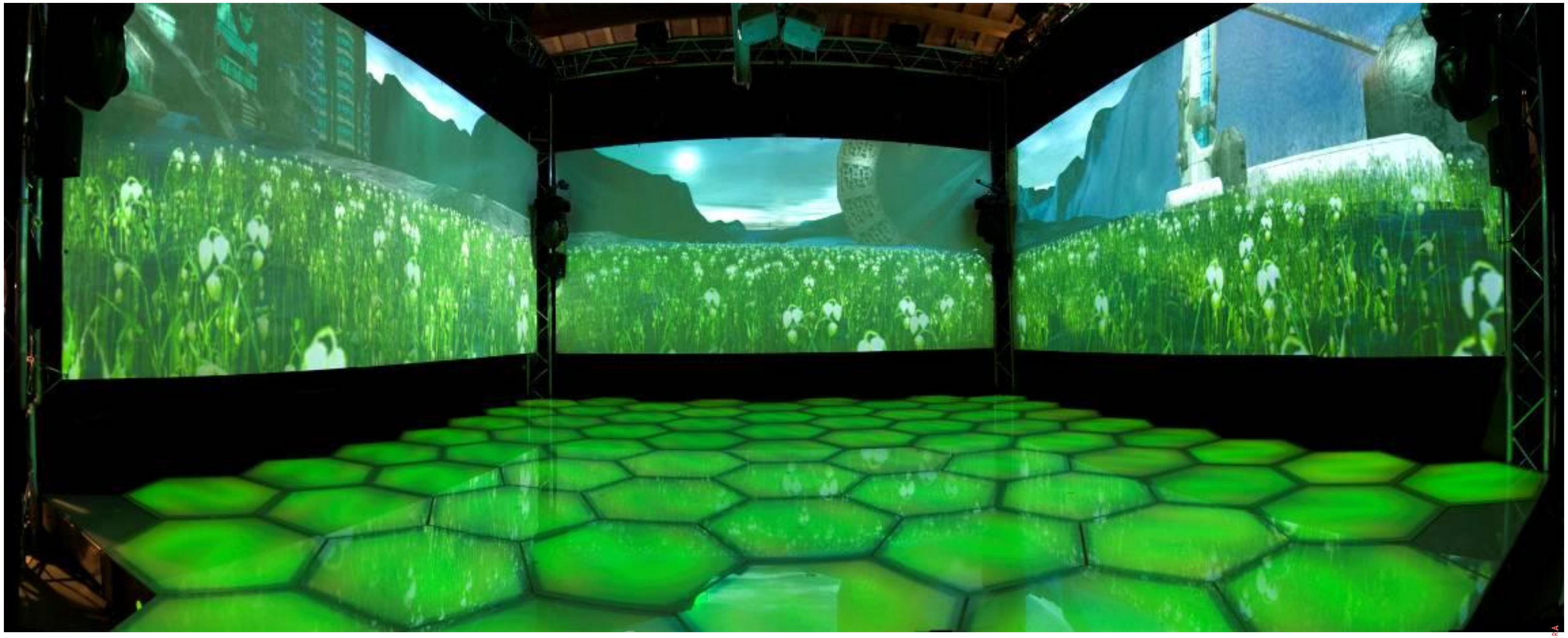
Very different ties

Similar ties

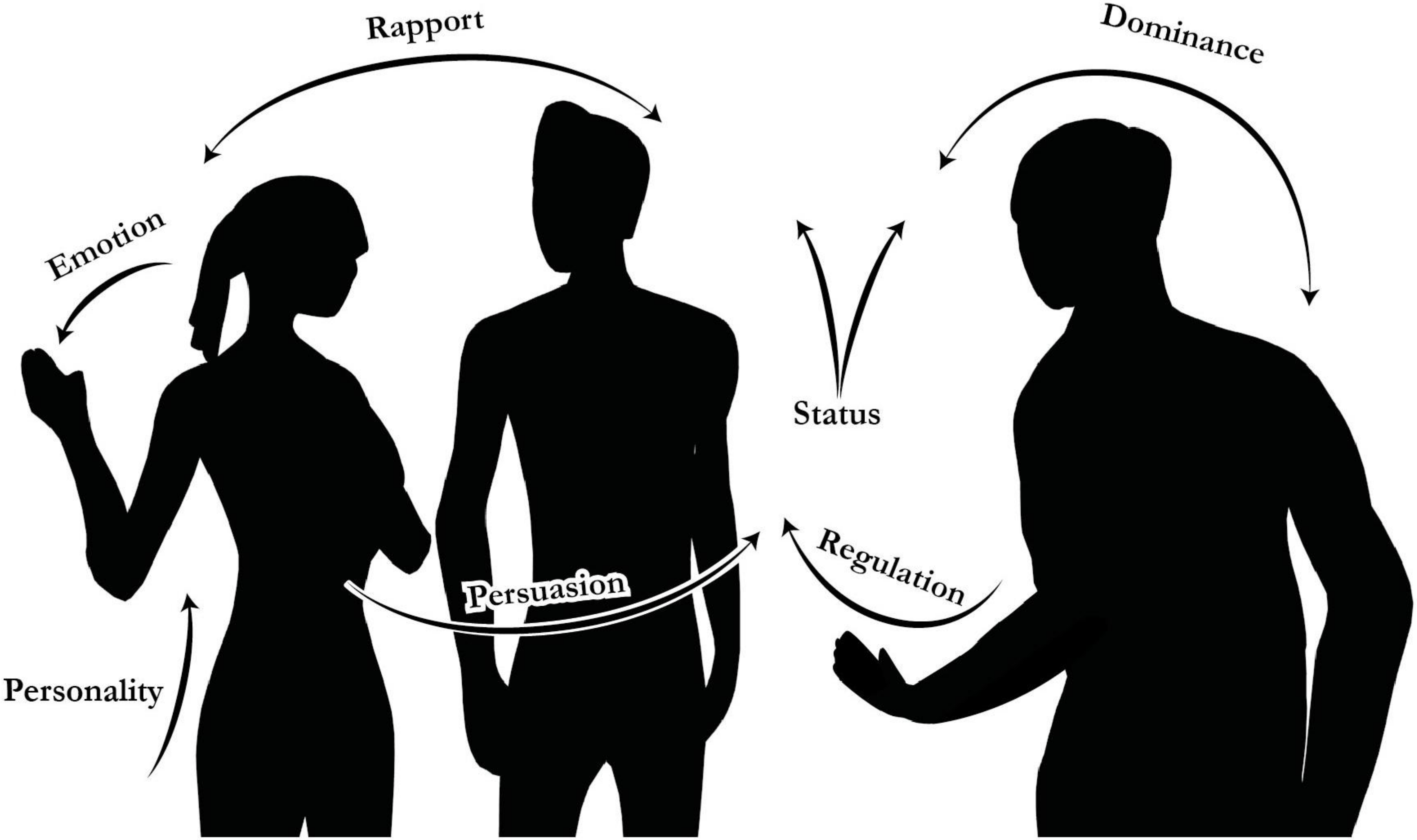
Nikolaus Bee, Helmut Prendinger, Arturo Nakasone, Elisabeth André, Mitsuru Ishizuka: AutoSelect: What You Want Is What You Get: Real-Time Processing of Visual Attention and Affect. PIT 2006: 40-52

Exploit Unconscious Signals to Optimize Data Exploration

- Users are projected into data spaces while their body suit monitors their coupling with this experience
- Exploit implicit sources of information (gaze, gestures, posture, EEG)
- Linking multiple users together to create a collective discovery system



Conscious and Unconscious Signals in Human Dialogue



Why to Care about Social Skills of Machines?

- **Traum and colleagues:**
  - Many breakdowns in man-machine communication could be avoided if the machine was sensitive to the user‘s emotions.
- **Aist and colleagues:**
  - Emotional scaffolding leads to a more persistent learning performance.
- **Prendinger and colleagues:**
  - An empathetic system led to a more positive physiological response.
- **Bosma and André:**
  - Physiological data (heart rate, skin conductance etc.) are significantly correlated to the level of commitment
  - → Resolve ambiguities in feedback signals, such as “ok”

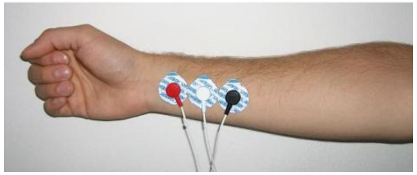
Challenges of Social Signal Interpretation

- Ambiguities of social cues
- Variations in social cues are quite high
  - Situation-specific
  - User-specific
  - …
- Social cues may be suppressed or faked

- **Even more challenges in AMI environments due to the highly unpredictable situations**

Analysis of Social Signals

- **Kinds of psychological and conversational states**
  - **Emotions** from
    - Facial expressions (Zeng et al. 2009)
    - Gestures (Caridakis et al. 2006)
    - Speech (Vogt et al. 2008)
    - Physiological measurements (Kim & André 2008)
  - **Interest** (Schuller et al. 2009)
  - **Engagement** (Nakano & Ishii 2012)
  - **Trust** (Bee et al. 2011)
  - **Personality** (Pianesi et al. 2008)
  - **Rapport** (Gratch et al. 2006)



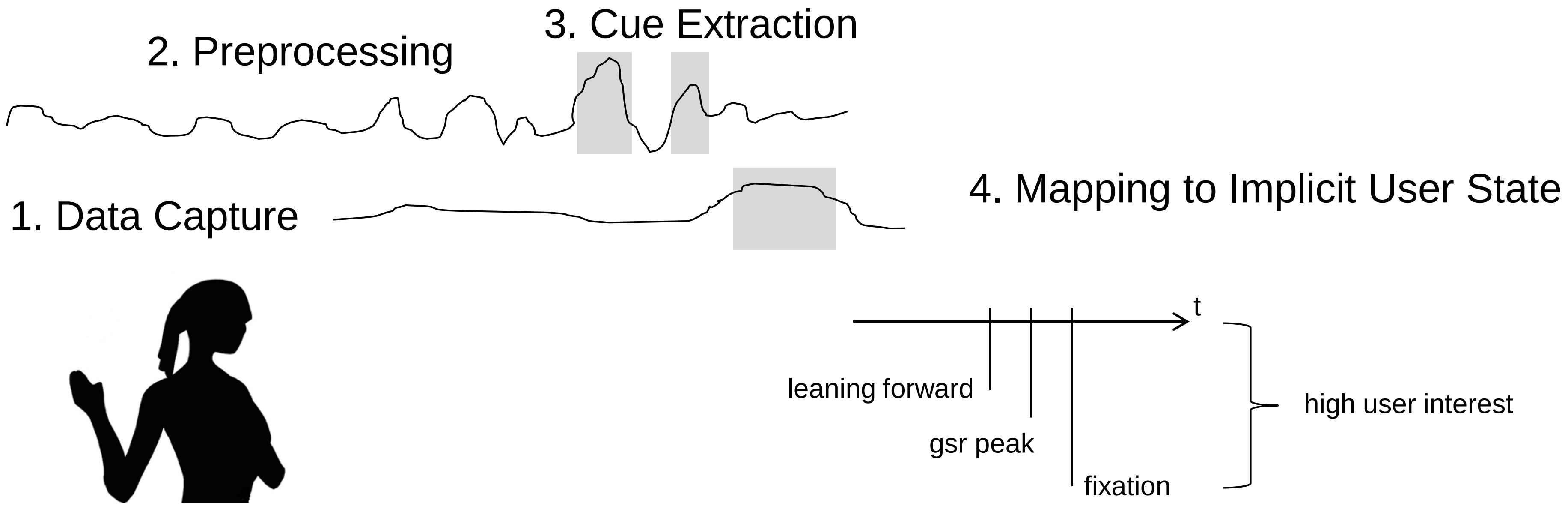
- **SSPNet** FP7 Network of Excellence on Social Signal Processing

Analysis of Social Signals in Small Groups

- Speech activity and fidgeting, i.e. amount of movement in a person‘s hands and body, to detect **functional roles** in a group (Dong et al. 2007)
- Overlapping speech, video cues, such as motion energy and audio-visual cues, such as the amount of movement during speech, to determine the **level of group cohesion** in meetings (Hung and Gatica 2010)
- Recognition of social laughter as an indicator of **emotional contagion** (Wagner et al. 2012)

From Raw Sensor Data to Implicit User States …

- Integrating work on **Embedded Computing** (University of Pisa, De Rossi) and **Social Signal Processing** (Augsburg University)



3. Cue Extraction

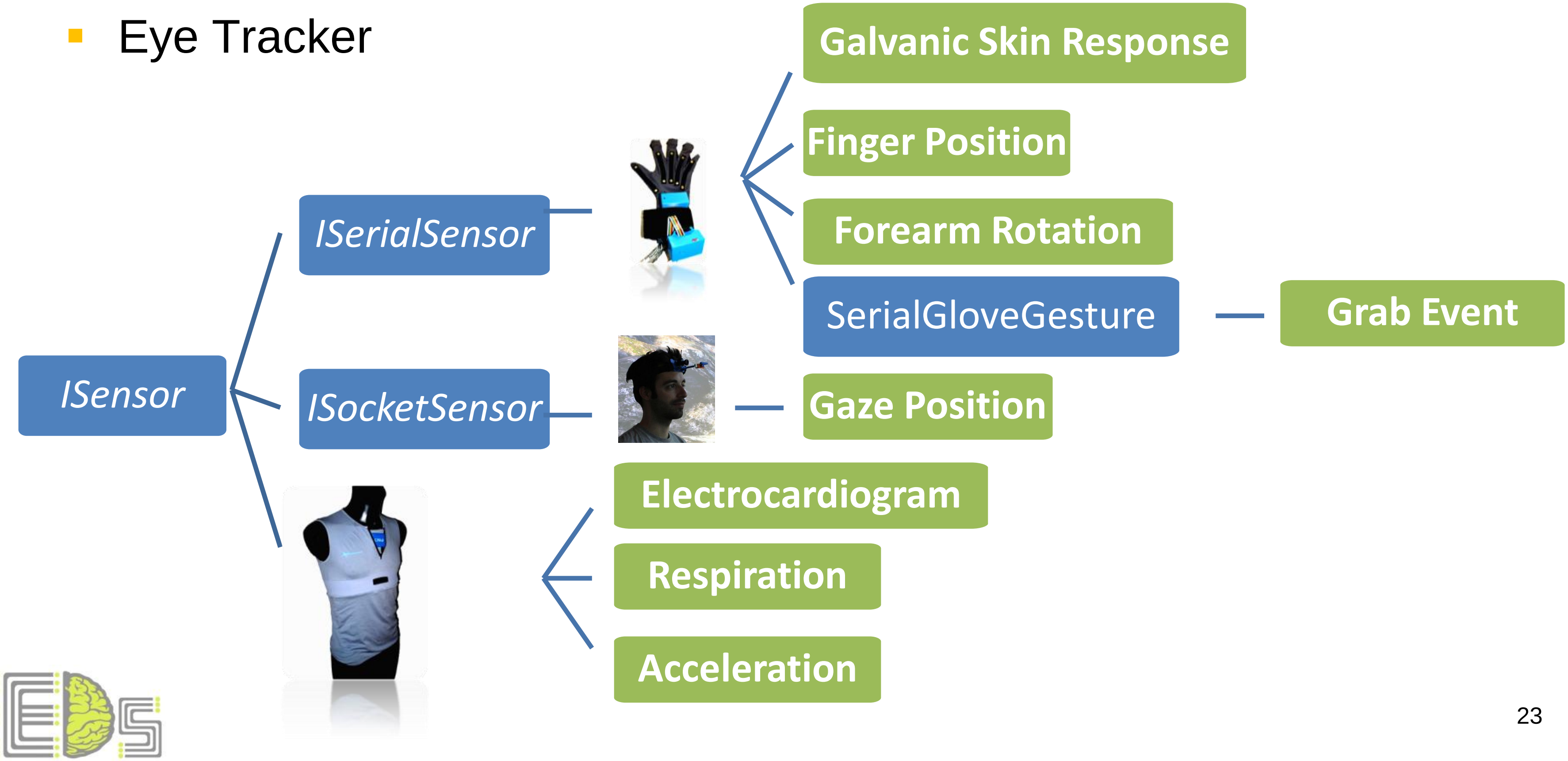
2. Preprocessing

4. Mapping to Implicit User State

1. Data Capture

Sensor Devices from University of Pisa

- Integration of sensor devices into the SSI framework in order to provide a coherent platform for sensing and processing raw signals
  - Smartex T-Shirt
  - Data Glove
  - Eye Tracker



ISensor

ISerialSensor

ISocketSensor

Galvanic Skin Response

Finger Position

Forearm Rotation

SerialGloveGesture

Grab Event

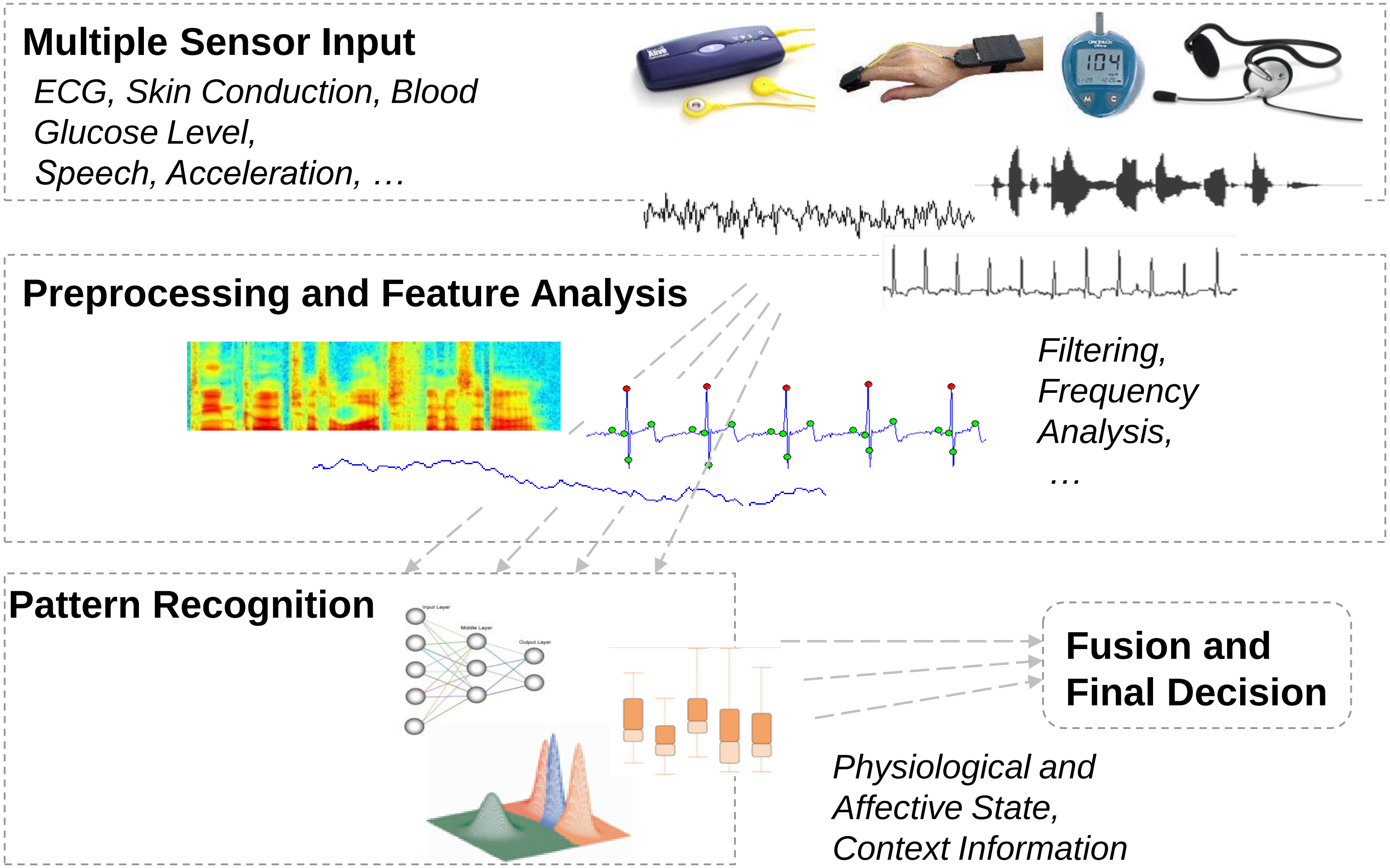
Gaze Position

Electrocardiogram

Respiration

Acceleration

Social Signal Interpretation: SSI by Augsburg University



Multiple Sensor Input

ECG, Skin Conduction, Blood Glucose Level, Speech, Acceleration, …

Preprocessing and Feature Analysis

Filtering, Frequency Analysis, …

Pattern Recognition

Fusion and Final Decision

Physiological and Affective State, Context Information

SSI is freely available under: http://www.openssi.net

Why traditional ML does not work in AMI environments

- Social cue recognition performance is overestimated
  - Most recognition systems are trained and tested on corpora that contain fixed segments with acted prototypical cues
  - Often only cues that have been labeled equally by a majority of annotators are used for classification.
- Requirements for realistic applications
  - We have to cope with non-prototypical user data.
  - Cues have to be processed frame by frame as being produced by the user.

Why traditional ML does not work in AMI environments

- **Previous approaches**
  - Segmentation-based
  - Offline
  - Classifier trained on prototypical data
  - Focus on acted data

- **Requirements for AMI**
  - Framewise
  - Online
  - Classifier trained on all data (prototypical and non-prototypical)
  - Focus on spontaneous data

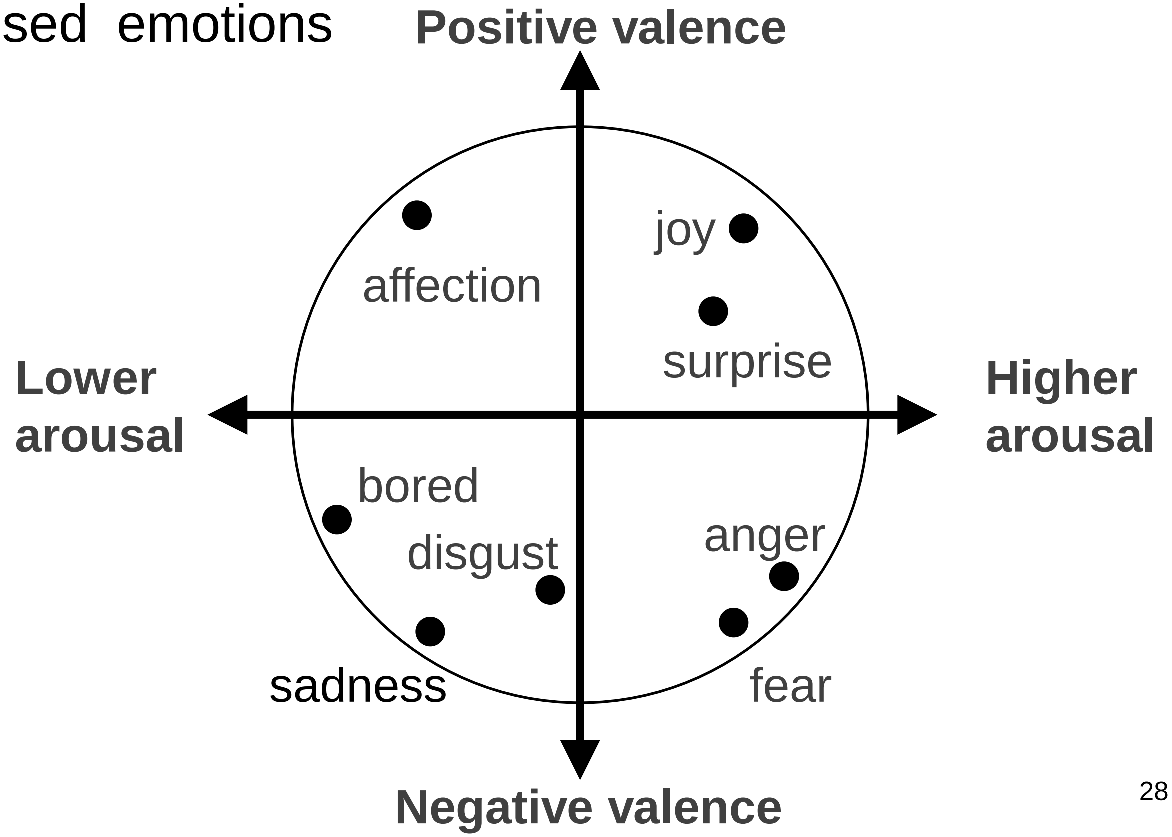
Why we need to train classifiers from spontaneous data

- **Ekman's Basic Emotions:**
  - Anger, Disgust, Fear, Happiness, Sadness, Surprise

- **Application-oriented Emotions:**
  - Call Centers: Anger
  - Meetings: Engagement, Approval, Disapproval
  - Dialog Systems: Confidence, Confusion, Frustration, Baby Talk, Politeness, Interest
  - Driver Assistance Systems: Stress
  - Smart Home: Emotions do not only depend on the application, but on the user's general situation, e.g. stress with partner, tiredness due to sickness etc.

Strategy to Cope with Unpredictable Emotional States

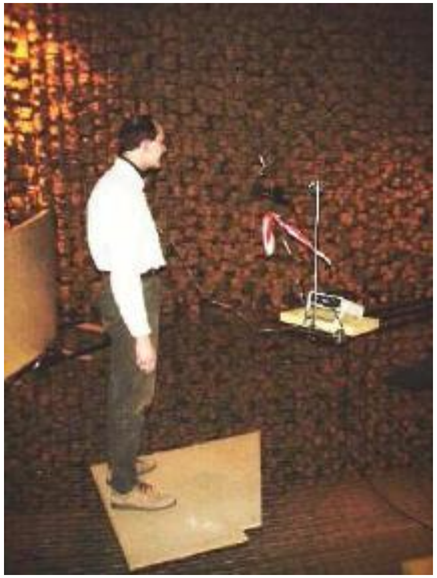
- **Strategy:**
  - train a limited set of emotion classes based on pleasure and arousal in a dimensional emotion model
  - which should then subsume the actually expressed emotions at runtime



Recognition of Emotions from Spontaneous Speech

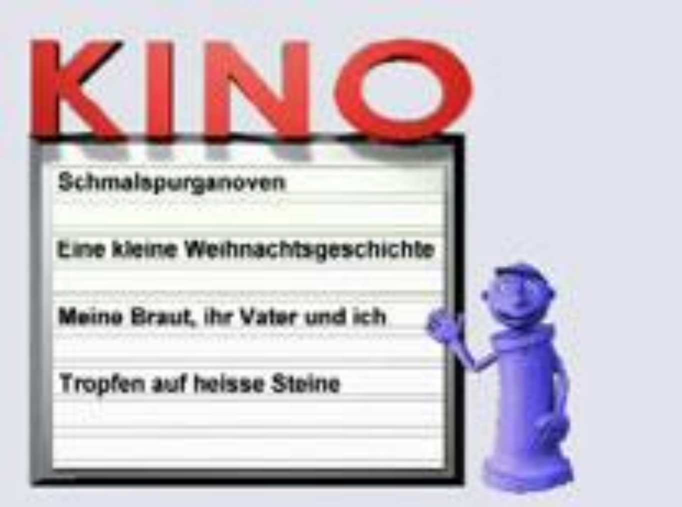
- **Corpus of TU Berlin**
  - Acted speech of 10 professional actors
  - **Recognition rates:** about 80 % for a 7-class problem (BERLIN)

Joy:	In 7 Stunden wird es soweit sein. 🔊
Anger:	In 7 hours, it will happen. 🔊



Recognition of Emotions from Spontaneous Speech

- **SmartKom Korpus of LMU**
  - Spontaneous speech of ca. 80 users, approx. equal gender distribution
  - Wizard-of-Oz setting
  - Partly emotional speech as sometimes malfunction of system was simulated
  - **Recognition Rates:** 26% for 7 emotions



KINO

Schmalspurganoven

Eine kleine Weihnachtsgeschichte

Meine Braut, ihr Vater und ich

Tropfen auf heisse Steine

Irritation: Ich möchte‘ ne Email schreiben. Email – nicht Telephon. Ok? Email.
I would like to write an email. Email – not telephone. Ok? Email.

Joy: Ja, bitte. Ich möchte telephonisch reservieren.
Yes, please. I would like to make a reservation by phone.

Recognition of Emotions from Spontaneous Speech

- **AIBO Corpus of Friedrich-Alexander-Universität Erlangen**
  - Spontaneous speech of ca. 50 children between 10 and 13 years old
  - **Recognition Rates:** ca. 60% for four emotions

Joy: Nein, Aibo, Du sollst nach links gehen.
No, Aibo, you have to go to the left.

Irritation: Aibo, Du sollst aufstehen.
Aibo, you have to get up.

Why we need to train classifiers from spontaneous data

- Accuracy for acted speech quite high
  - about 80 % for a 7-class problem (BERLIN)
- Classification of natural emotions only usable for a smaller number of classes
  - about 60 % for a 4-class problem (AIBO)
  - about 50 % for a 3-class problem (SMARTKOM)
- Feature reduction less important for natural emotions
- We cannot learn best segment length and best features for natural emotions from acted emotions

Thurid Vogt, Elisabeth André: Comparing Feature Sets for Acted and Spontaneous Speech in View of Automatic Emotion Recognition. ICME 2005: 474-477

Laughter Recognition

- **Task:**
  - explore features that are suitable to detect laughter in continuous speech
- **Challenge:**
  - Laughter consists of many distinctive sounds: evident, inaudible, song-like, grunt-like etc. many of which resemble speech (Bachorowski, Smoski and Owren)
- **Corpus Used for Classifier Training:**
  - Emotionally colored conversations from SEMAINE corpus
  - Additional YOUTUBE laughter sessions

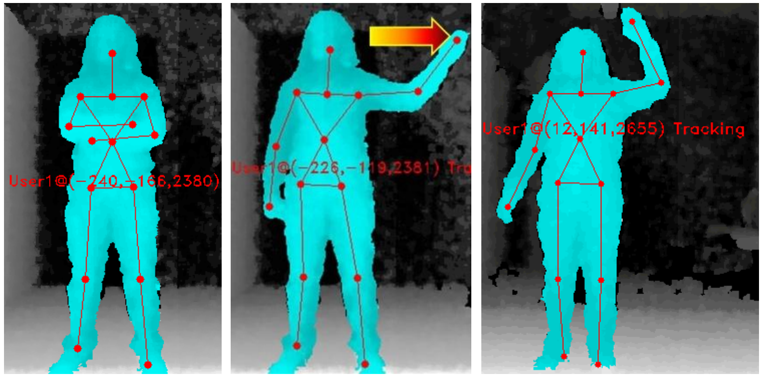
Accuracy Rates for Laughter

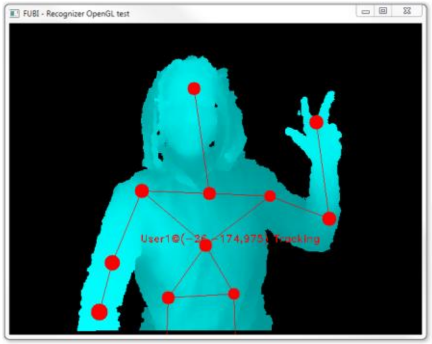
- Most studies achieve > 90 %, however, no distinction between different kinds of laughter

	Cai et al.	Konx & Mirghafori	Felkin , Terrien & Thorisson	Guirguis, Wagner, Lingenfelser & André
Classifier	HMM	SVM MLP	C4.5	SVM
Window	1.5 s	~2 s	2.5 s	1 s
Dataset	TV Programs	ICSI Meetings	Own Recordings	Semaine, Youtube Videos
Accuracy	~ 90 %	92 % ~ 96 %	89.5 %	91,2 %

Higher Level Processing

- **Kinect Gesture Recognition FUBI**
  - recognition of full body gestures and postures
  - large set of pre-defined recognizers
  - own recognizers can be defined in XML
  - finger recognition





Declarative Specification of Gestures

XML Definition

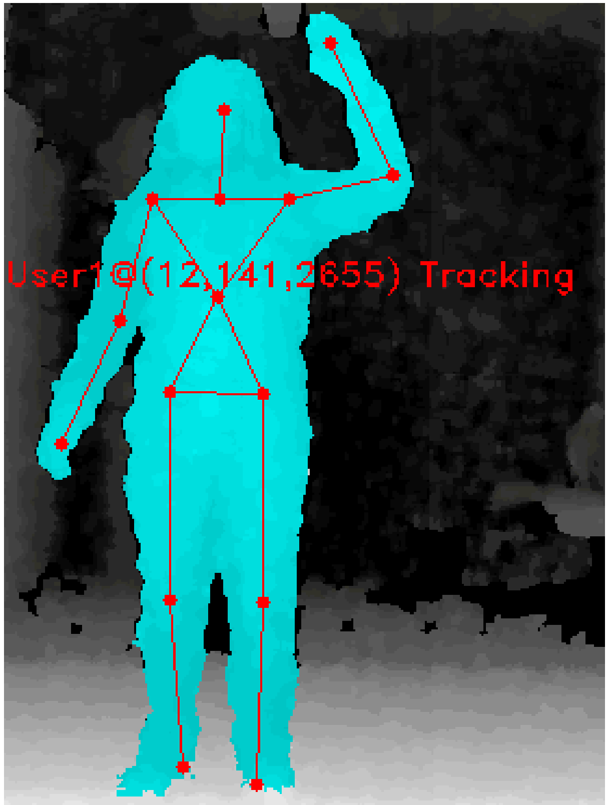
```
<PostureCombinationRecognizer
        name=“Waving">

        <Recognizer name="rightHandRight"/>
        <State maxDuration="1.2”
                  minDuration="0.05"
                  timeForTransition="0.4“/>

        <Recognizer name="rightHandLeft"/>
        <State maxDuration="1.2“
                  minDuration="0.05"
                  timeForTransition="0.4“/>

</PostureCombinationRecognizer>
```

Gesture



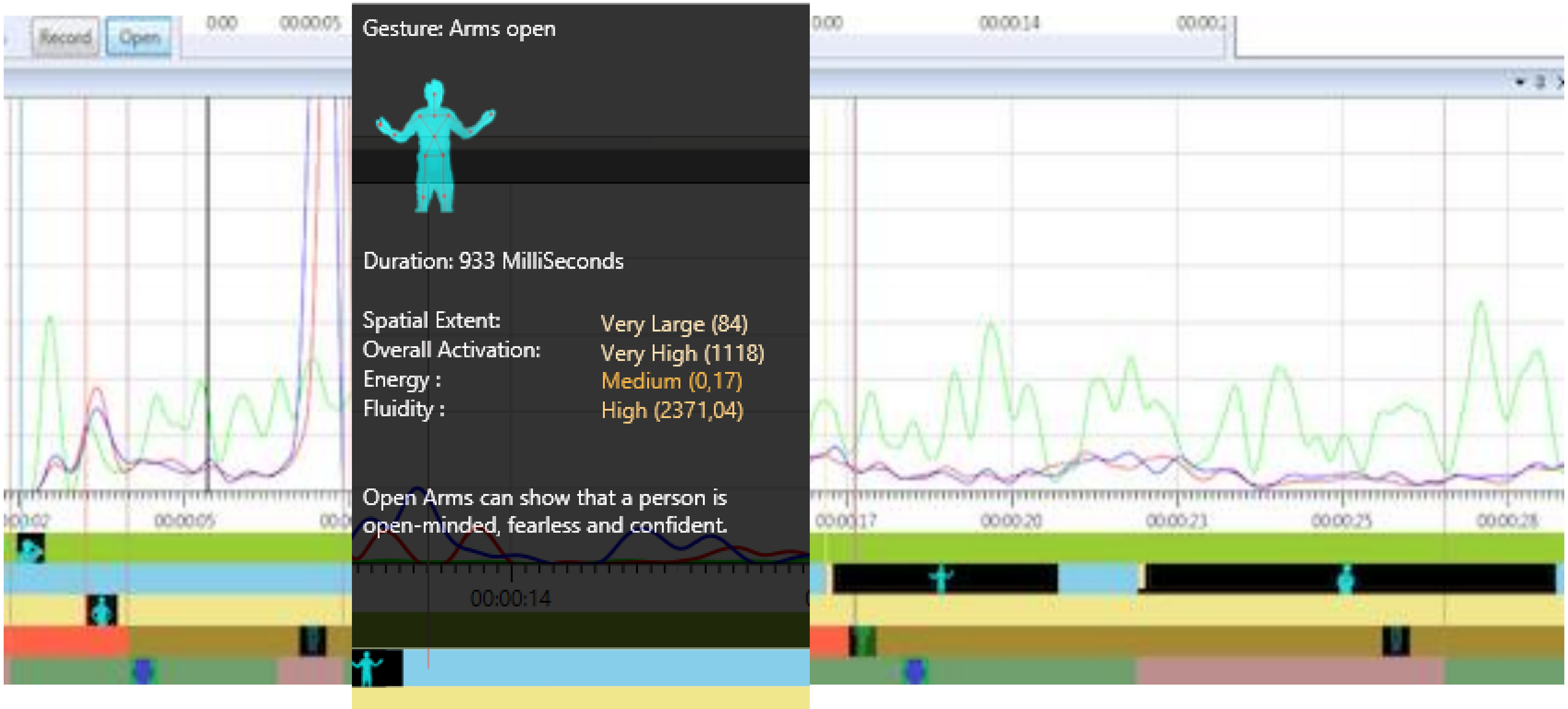
Signals and their Interpretation in Social Interactions

- **Example:** Level of Engagement

	Movement Quality	**Specific body movements**
Speaking		
High	high overall activity	orientation of the body and the face towards the interlocutor
Low	low overall activity orientation of	orientation of the body and the face away from the interlocutor
Listening		
High	low overall activity	orientation of the body and the face towards the interlocutor, head tilt, touch chin without bracing the head
Low	high overall activity	orientation of the body and the face away from the interlocutor, touch chin while fully bracing the head

Signals and their Interpretation in Social Interactions

- **Example:** Level of Engagement



Gesture: Arms open

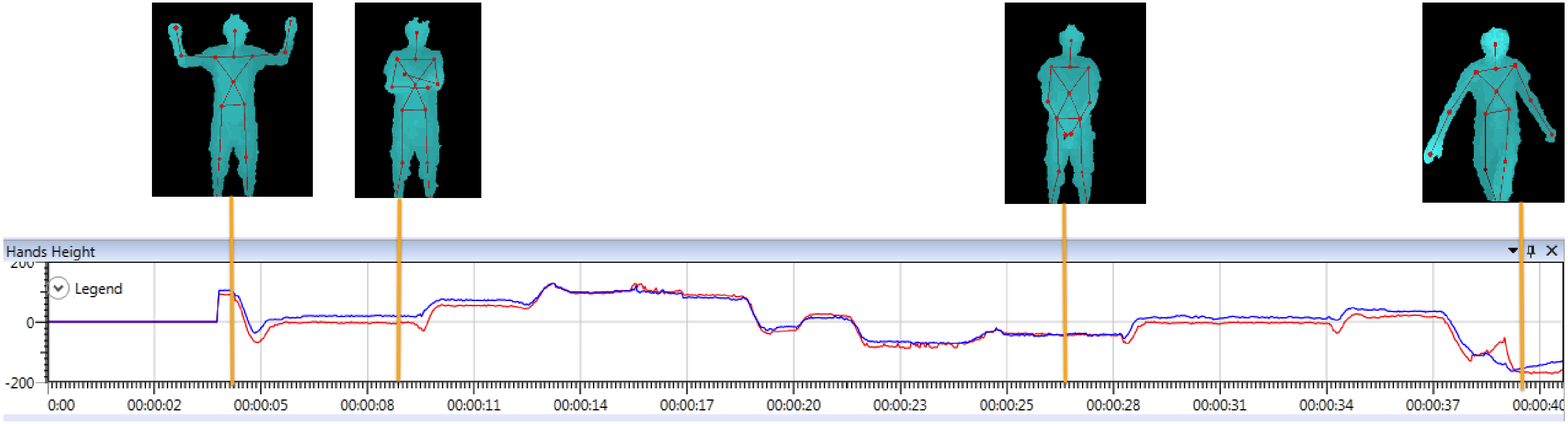
Duration: 933 MilliSeconds

Spatial Extent: Very Large (84)
Overall Activation: Very High (1118)
Energy : Medium (0,17)
Fluidity : High (2371,04)

Open Arms can show that a person is open-minded, fearless and confident.

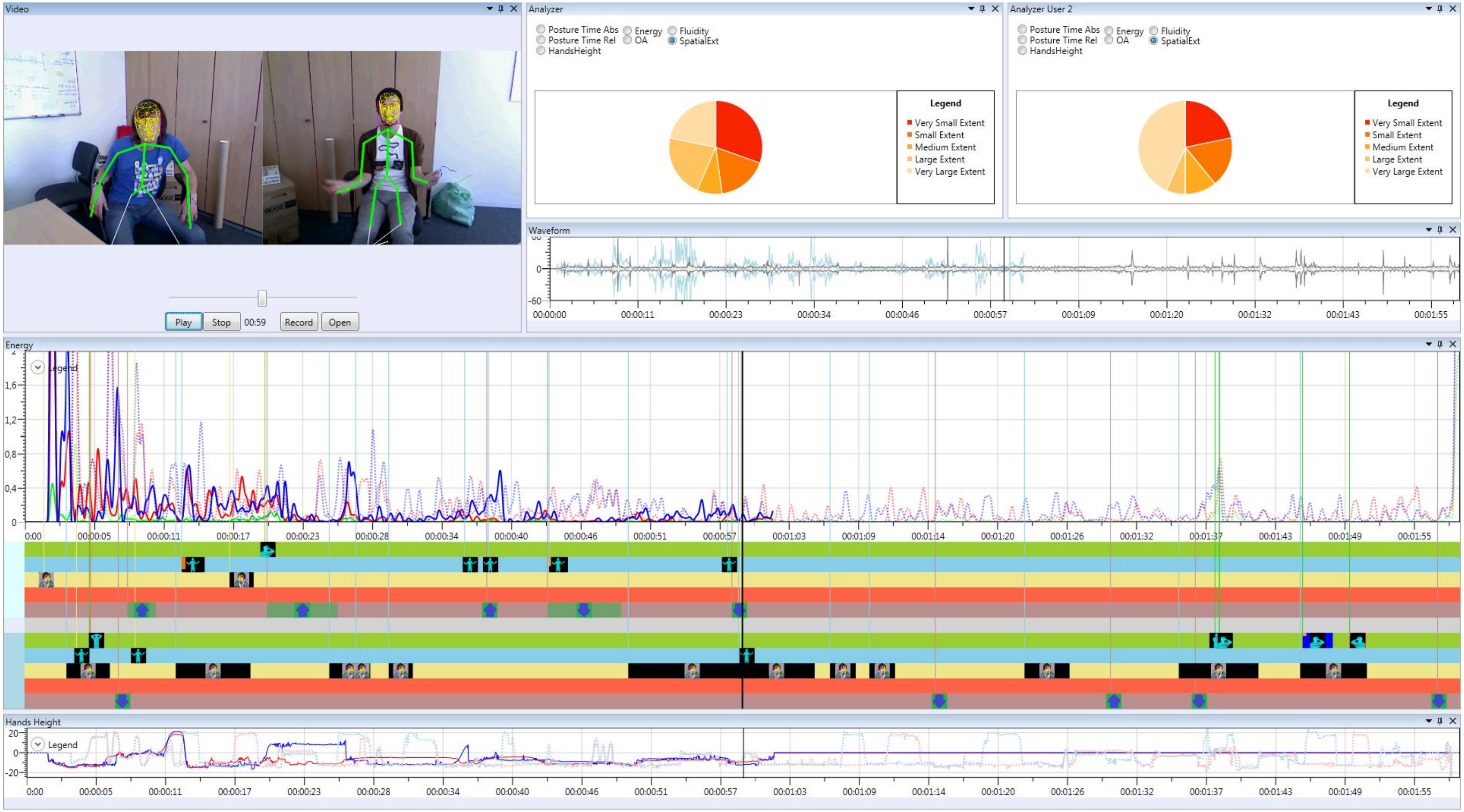
Frame-by-Frame Analysis of Bodily Behaviors

- **<u>Example:</u>** Analysis of the hands’ height in relation to the torso and to each other



Analysis of Non-Verbal Behavior

Universität
Augsburg
University


Video

Analyzer

○ Posture Time Abs ○ Energy ○ Fluidity
○ Posture Time Rel ○ OA ● SpatialExt
○ HandsHeight

Legend
- 🟥 Very Small Extent
- 🟧 Small Extent
- 🟧 Medium Extent
- 🟨 Large Extent
- 🟨 Very Large Extent

Analyzer User 2

○ Posture Time Abs ○ Energy ○ Fluidity
○ Posture Time Rel ○ OA ● SpatialExt
○ HandsHeight

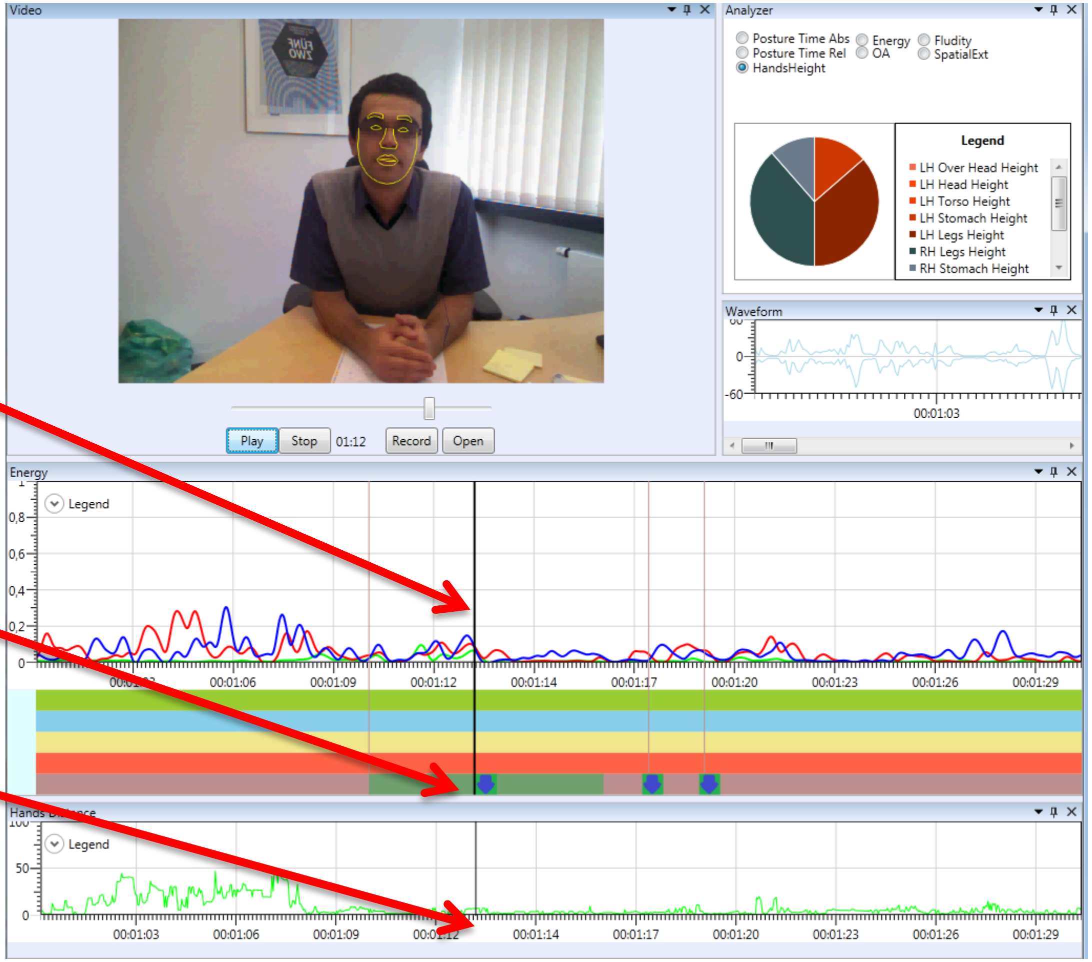
Legend
- 🟥 Very Small Extent
- 🟧 Small Extent
- 🟧 Medium Extent
- 🟨 Large Extent
- 🟨 Very Large Extent

Waveform

Energy

Hands Height

Analysis of Non-Verbal Behavior



○ Posture Time Abs ○ Energy ○ Fludity
○ Posture Time Rel ○ OA ○ SpatialExt
● HandsHeight

Calm hand-movement

Lean forward posture detected

Hands together, visualized in the graph

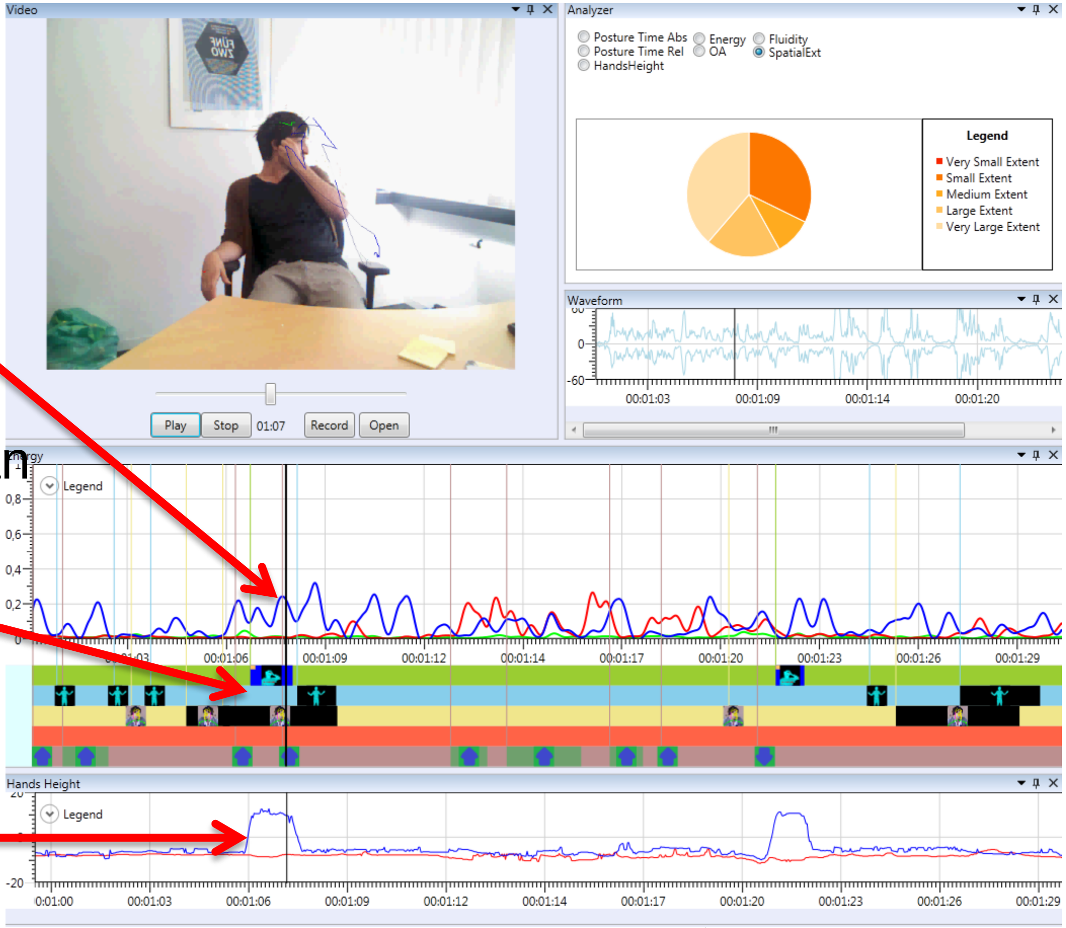
Increased Interest

Analysis of Non-Verbal Behavior

Sudden hand movements

Head touch with the left hand, Look away to the left side & Lean backward posture, detected at the same time

Left hand in head height, visualized in the graph



Analyzer

○ Posture Time Abs ○ Posture Time Rel ○ HandsHeight ○ Energy ○ OA ○ Fluidity ● SpatialExt

Legend

- 🟥 Very Small Extent
- 🟧 Small Extent
- 🟧 Medium Extent
- 🟨 Large Extent
- 🟨 Very Large Extent

Low Interest

Multi-Modal Social Signals

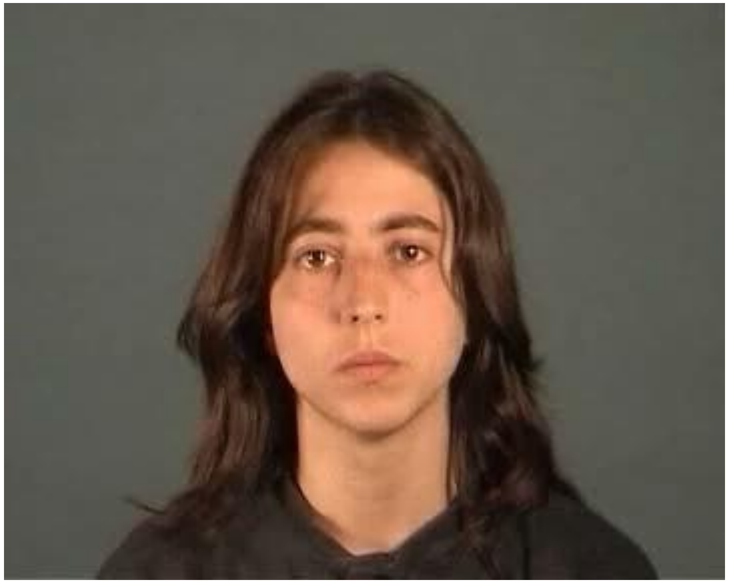
- Emotions are generally expressed through multiple modalities
- Emotions can be illustrated by a combination of vocal behavior, facial expressions, gestures and postures
- Humans base and refine their classifications of observed affective states on more than one modality - machines that try to recognize
  emotions should do so too

Experimental Comparison

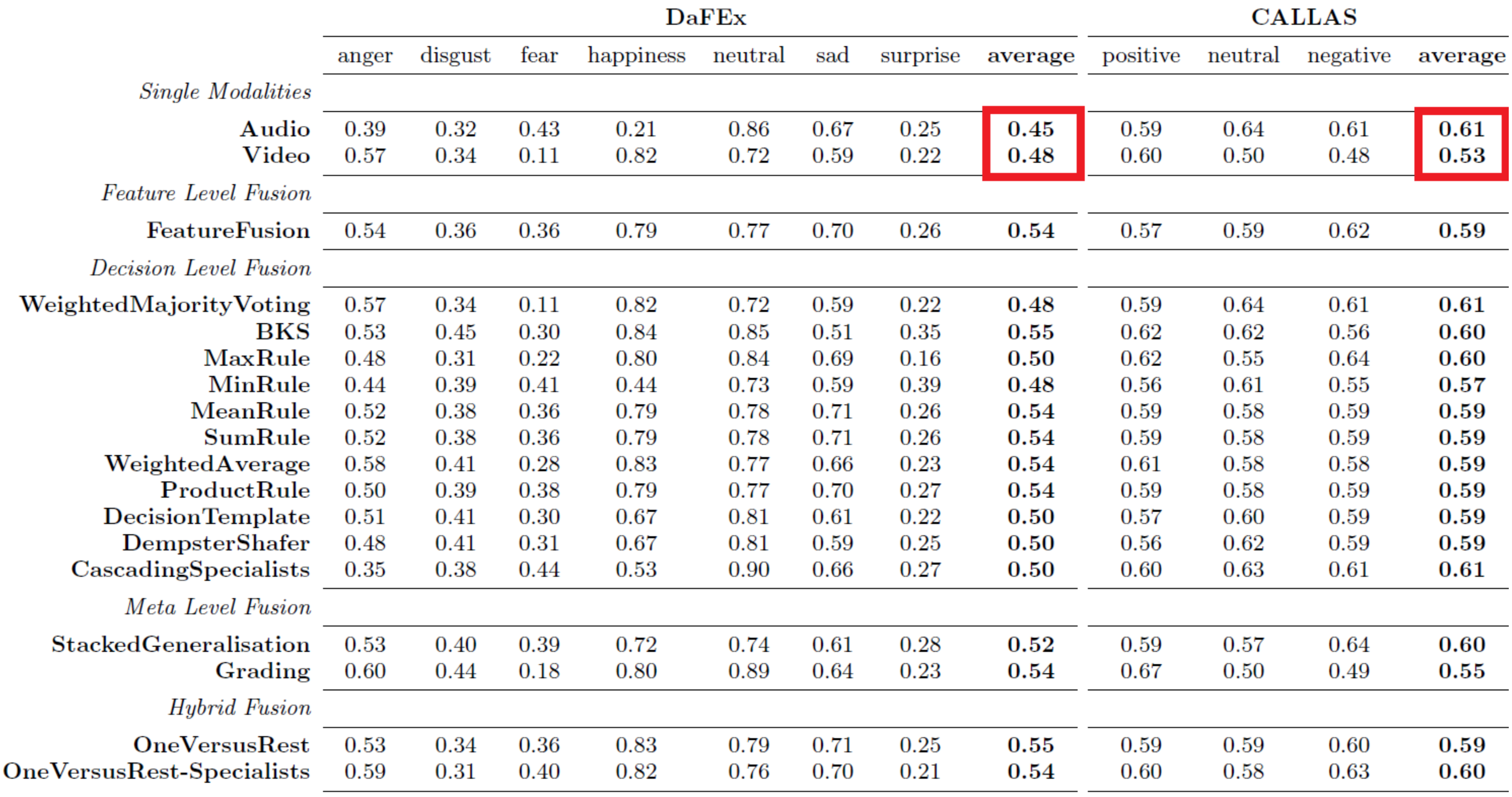
- Evaluation of fusion schemes on two Italian emotion corpora

DaFEx Corpus
acted and exaggerated

CALLAS Corpus
non-acted and natural



Results for Various Fusion Mechanisms

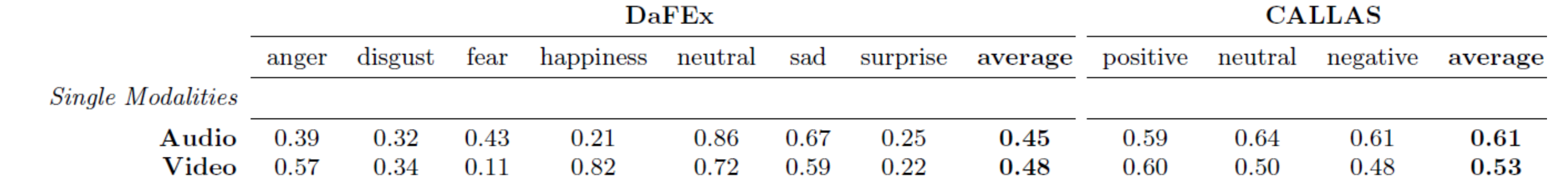


	DaFEx								CALLAS			
	anger	disgust	fear	happiness	neutral	sad	surprise	average	positive	neutral	negative	average
Single Modalities												
Audio	0.39	0.32	0.43	0.21	0.86	0.67	0.25	**0.45**	0.59	0.64	0.61	**0.61**
Video	0.57	0.34	0.11	0.82	0.72	0.59	0.22	**0.48**	0.60	0.50	0.48	**0.53**
Feature Level Fusion												
FeatureFusion	0.54	0.36	0.36	0.79	0.77	0.70	0.26	**0.54**	0.57	0.59	0.62	**0.59**
Decision Level Fusion												
WeightedMajorityVoting	0.57	0.34	0.11	0.82	0.72	0.59	0.22	**0.48**	0.59	0.64	0.61	**0.61**
BKS	0.53	0.45	0.30	0.84	0.85	0.51	0.35	**0.55**	0.62	0.62	0.56	**0.60**
MaxRule	0.48	0.31	0.22	0.80	0.84	0.69	0.16	**0.50**	0.62	0.55	0.64	**0.60**
MinRule	0.44	0.39	0.41	0.44	0.73	0.59	0.39	**0.48**	0.56	0.61	0.55	**0.57**
MeanRule	0.52	0.38	0.36	0.79	0.78	0.71	0.26	**0.54**	0.59	0.58	0.59	**0.59**
SumRule	0.52	0.38	0.36	0.79	0.78	0.71	0.26	**0.54**	0.59	0.58	0.59	**0.59**
WeightedAverage	0.58	0.41	0.28	0.83	0.77	0.66	0.23	**0.54**	0.61	0.58	0.58	**0.59**
ProductRule	0.50	0.39	0.38	0.79	0.77	0.70	0.27	**0.54**	0.59	0.58	0.59	**0.59**
DecisionTemplate	0.51	0.41	0.30	0.67	0.81	0.61	0.22	**0.50**	0.57	0.60	0.59	**0.59**
DempsterShafer	0.48	0.41	0.31	0.67	0.81	0.59	0.25	**0.50**	0.56	0.62	0.59	**0.59**
CascadingSpecialists	0.35	0.38	0.44	0.53	0.90	0.66	0.27	**0.50**	0.60	0.63	0.61	**0.61**
Meta Level Fusion												
StackedGeneralisation	0.53	0.40	0.39	0.72	0.74	0.61	0.28	**0.52**	0.59	0.57	0.64	**0.60**
Grading	0.60	0.44	0.18	0.80	0.89	0.64	0.23	**0.54**	0.67	0.50	0.49	**0.55**
Hybrid Fusion												
OneVersusRest	0.53	0.34	0.36	0.83	0.79	0.71	0.25	**0.55**	0.59	0.59	0.60	**0.59**
OneVersusRest-Specialists	0.59	0.31	0.40	0.82	0.76	0.70	0.21	**0.54**	0.60	0.58	0.63	**0.60**

Florian Lingenfelser, Johannes Wagner, Elisabeth André: A systematic discussion of fusion techniques for multi-modal affect recognition tasks. ICMI 2011: 19-26

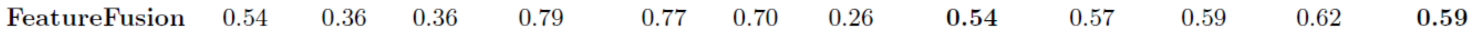
Observations

- Enhanced results only on the acted DaFEx corpus (acted emotions seem to lead to more consistent modalities)



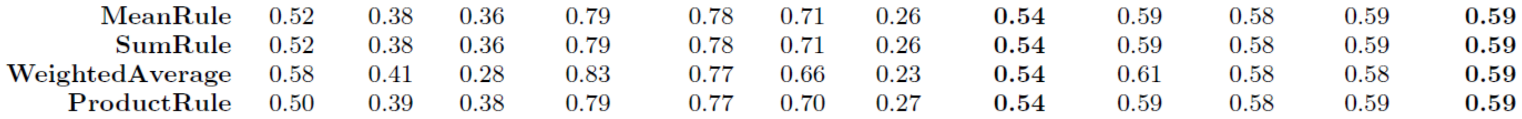
	DaFEx								CALLAS			
	anger	disgust	fear	happiness	neutral	sad	surprise	**average**	positive	neutral	negative	**average**
Single Modalities												
Audio	0.39	0.32	0.43	0.21	0.86	0.67	0.25	**0.45**	0.59	0.64	0.61	**0.61**
Video	0.57	0.34	0.11	0.82	0.72	0.59	0.22	**0.48**	0.60	0.50	0.48	**0.53**

- Feature Level Fusion: stable and straightforward, acceptable results



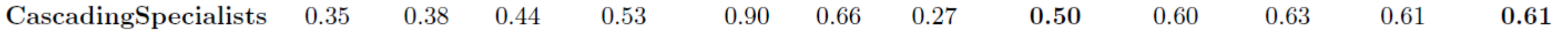
FeatureFusion	0.54	0.36	0.36	0.79	0.77	0.70	0.26	**0.54**	0.57	0.59	0.62	**0.59**

- Decision Level Fusion: impression of interchangeability



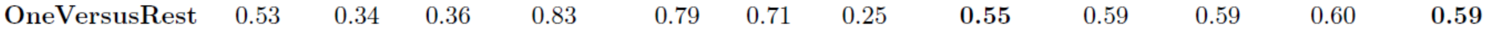
MeanRule	0.52	0.38	0.36	0.79	0.78	0.71	0.26	**0.54**	0.59	0.58	0.59	**0.59**
SumRule	0.52	0.38	0.36	0.79	0.78	0.71	0.26	**0.54**	0.59	0.58	0.59	**0.59**
WeightedAverage	0.58	0.41	0.28	0.83	0.77	0.66	0.23	**0.54**	0.61	0.58	0.58	**0.59**
ProductRule	0.50	0.39	0.38	0.79	0.77	0.70	0.27	**0.54**	0.59	0.58	0.59	**0.59**

- Specialist selection fails if parameterization fails (user-independent evaluation)



CascadingSpecialists	0.35	0.38	0.44	0.53	0.90	0.66	0.27	**0.50**	0.60	0.63	0.61	**0.61**

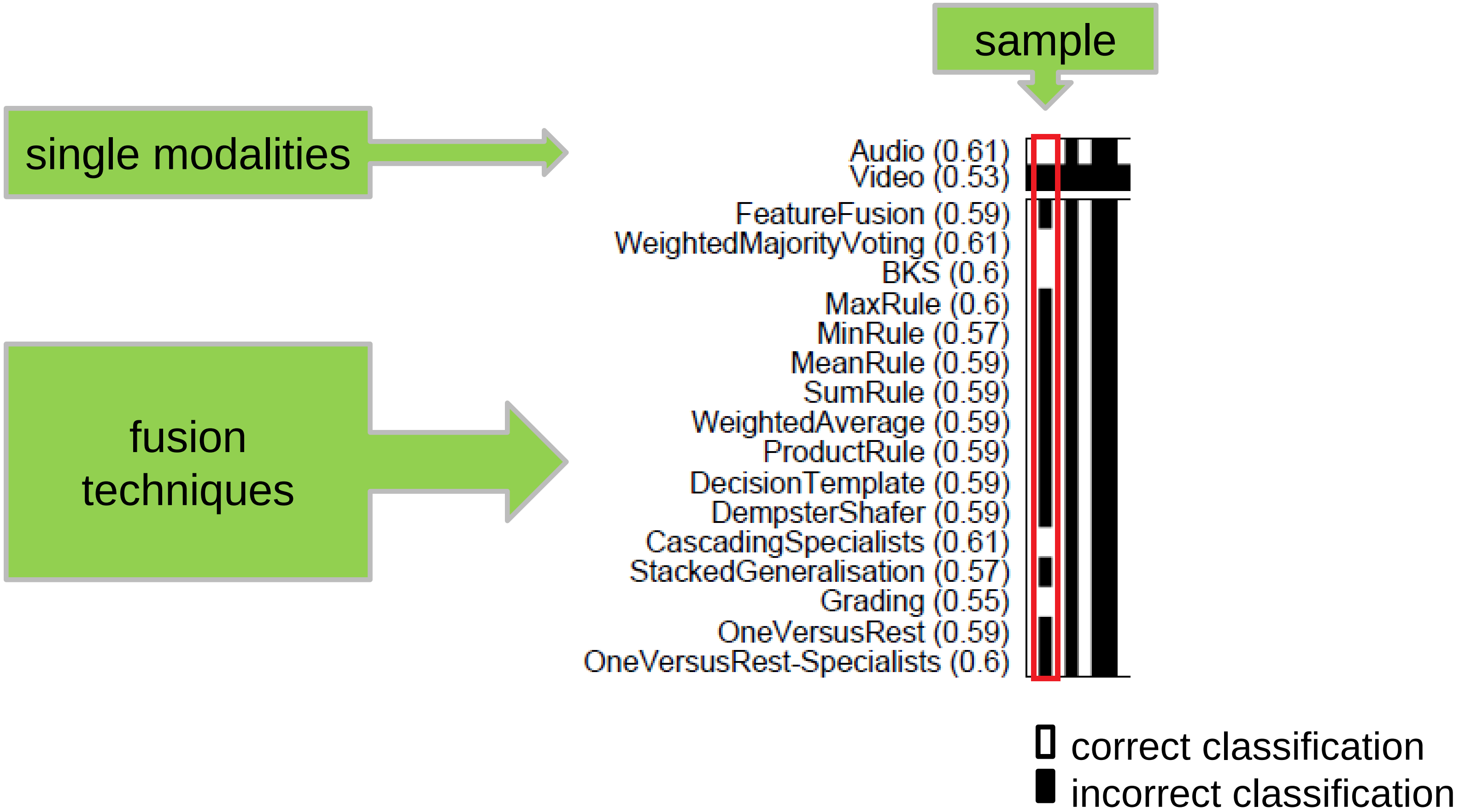
- Hybrid Fusion: more complex than simple feature fusion, but slightly better results



OneVersusRest	0.53	0.34	0.36	0.83	0.79	0.71	0.25	**0.55**	0.59	0.59	0.60	**0.59**

Visualization of Results

- Numbers not capable of representing where single fusion techniques gain or loose recognition accuracy



sample

single modalities

fusion techniques

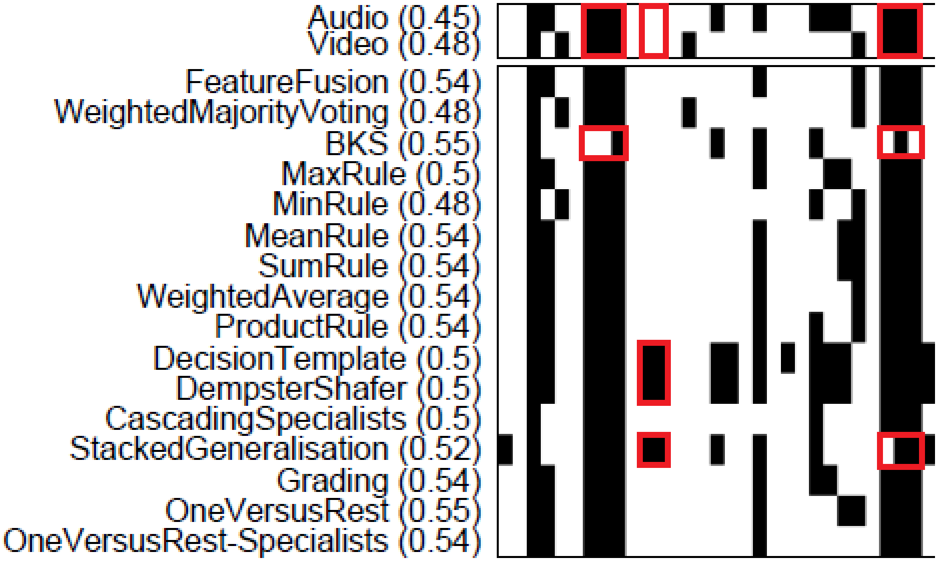
▯ correct classification

▮ incorrect classification

⇒ We need to consider context information!

Error Learning

- Error Learning (Meta Fusion Schemes, BKS, Decision Template, Dempster Shafer)



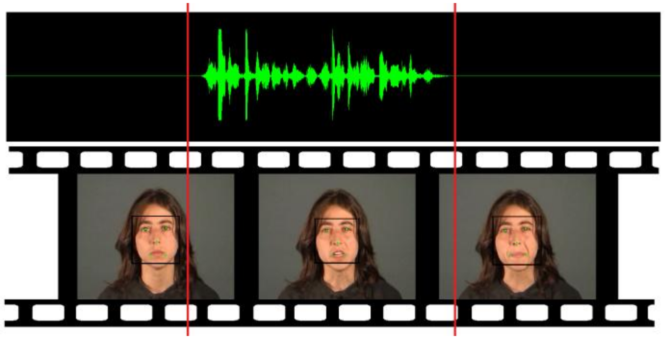
- despite wrong predictions in both modalities, correct prediction possible

BUT ALSO

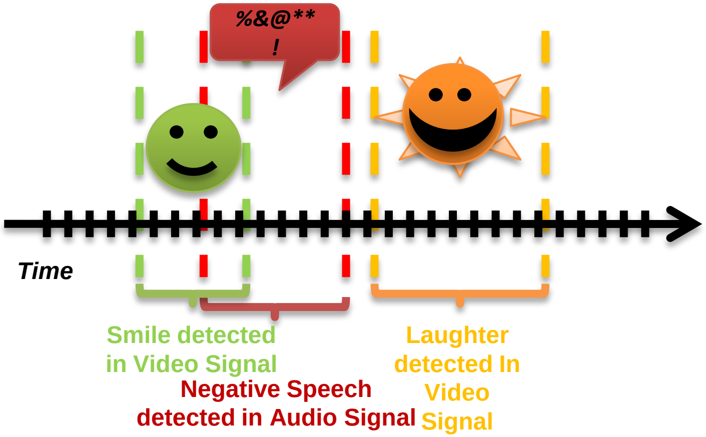
- incorrect fusion result despite correct predictions by single modalities

Discussion

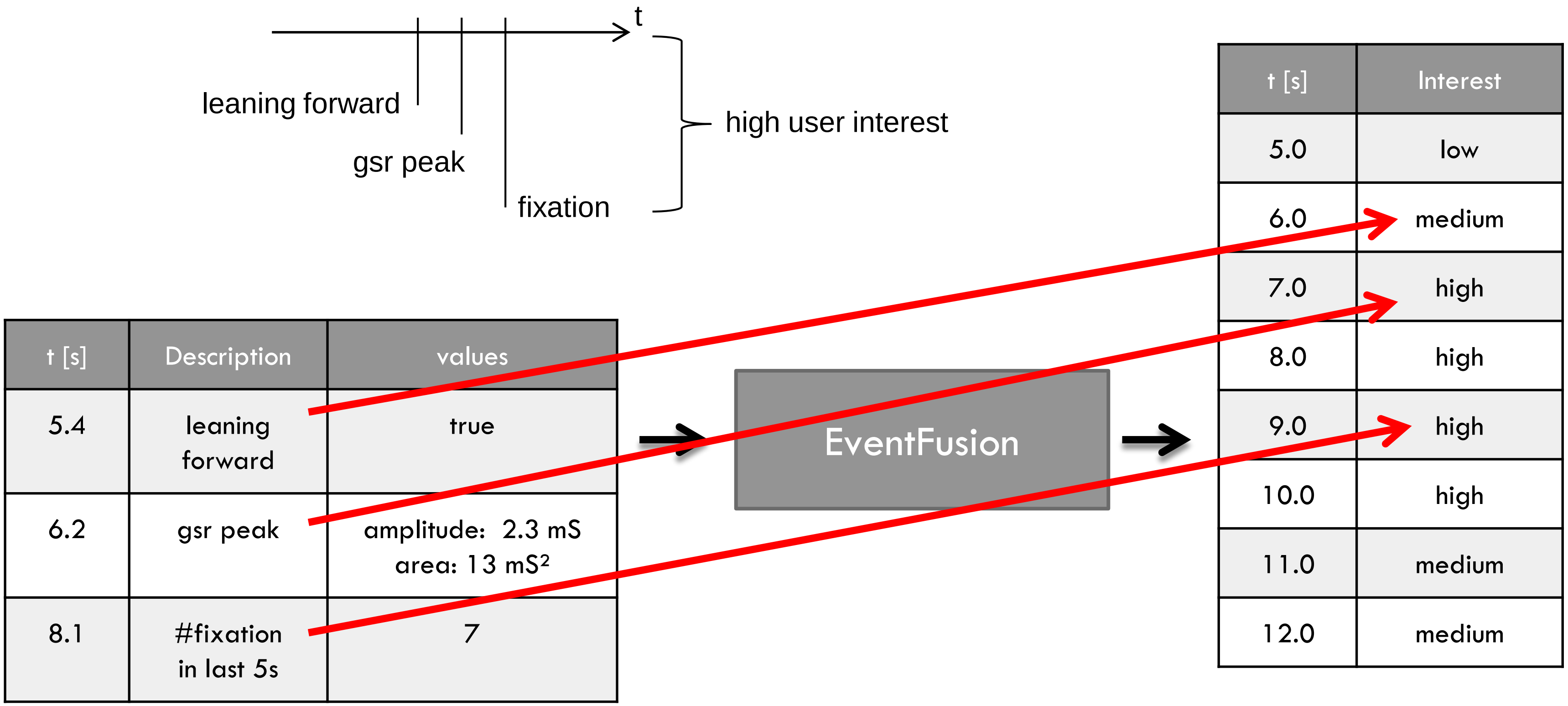
- **Hypothesis:**
  - Segmentation Problem
  - Analysis of further modalities is triggered by spoken sentences in the vocal modality – meaningful information in other modalities is assumed, but not guaranteed



- **Possible solution:**
  - Reject assumption „all relevant events happen at the same time in all modalities“
  - Separate treatment of events in different modalities
  - Incorporate temporal component


%&@**!
Time
Smile detected in Video Signal
Negative Speech detected in Audio Signal
Laughter detected In Video Signal

Event-Based Fusion



t [s]	Description	values
5.4	leaning forward	true
6.2	gsr peak	amplitude: 2.3 mS area: 13 mS²
8.1	#fixation in last 5s	7

t [s]	Interest
5.0	low
6.0	medium
7.0	high
8.0	high
9.0	high
10.0	high
11.0	medium
12.0	medium

Conclusions

- Social and emotional sensitivity may provide an added value to many AMI applications.
- Bringing Social Signal Processing to AMI leads to new requirements:
  - Frame-by-frame analysis instead of segment-based analysis
  - Online analysis (while the users are interacting) instead of offline analysis
  - Classifiers need to provide acceptable results for ALL data (prototypical and non-prototypical)
- Social and emotional signals are particularly difficult to interpret requiring to understand and model the causes and consequences of them.
- Realizing social and emotional intelligence requires a fully integrated loop consisting of perception, reasoning, learning and responding.
- ⇒ Exploit context sensing and reasoning technologies from AMI

Future Priorities

- Multisensory fusion
  - Integrating sensing technology in natural open environments
    - Distinguishing between command and no-command signals → get rid of push-to-command interfaces
  - Exploit information on context and psychological user states
    - to improve personalization
    - to increase robustness
- Fully integrated loop consisting of perception, reasoning, learning and responding → symbiotic human-machine interaction

Future Priorities

- Affective User Models
  - Focusing on unconscious signals to create and maintain
    - Rapport
    - Engagement
    - Common ground
    - User experience
  - Mechanisms to cope with uncertainties
  - Models of cognition
  - Long-term user modeling

Thanks to