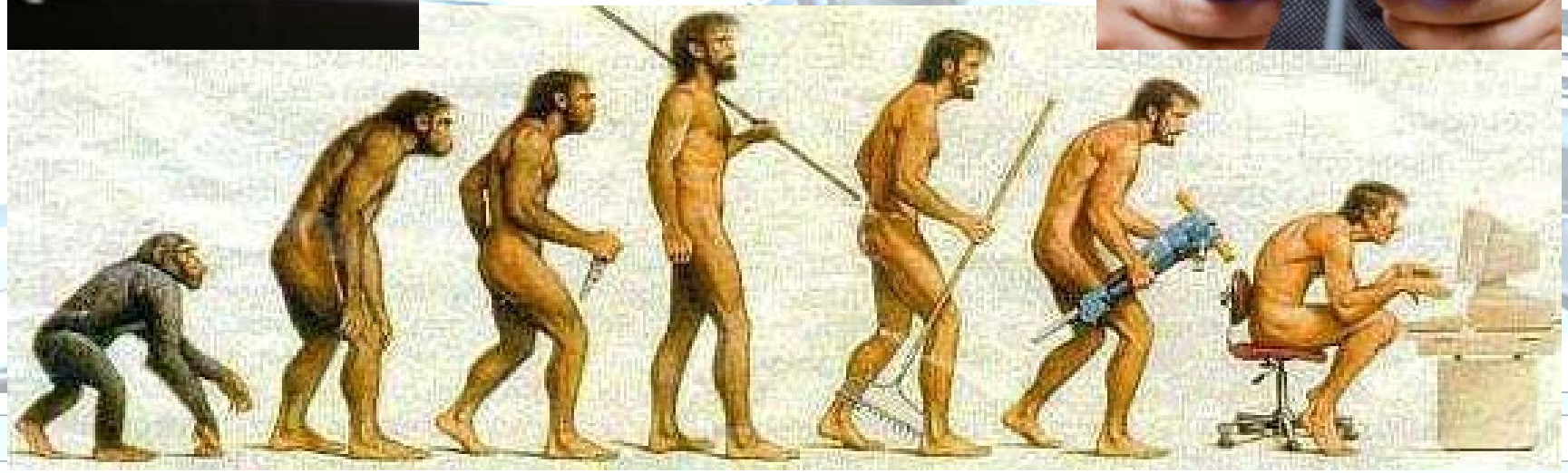
# The Future of User Interfaces: Human-Behaviour-Centred Designs

**Maja Pantic**

***Imperial College London***

***Computing Dept.***

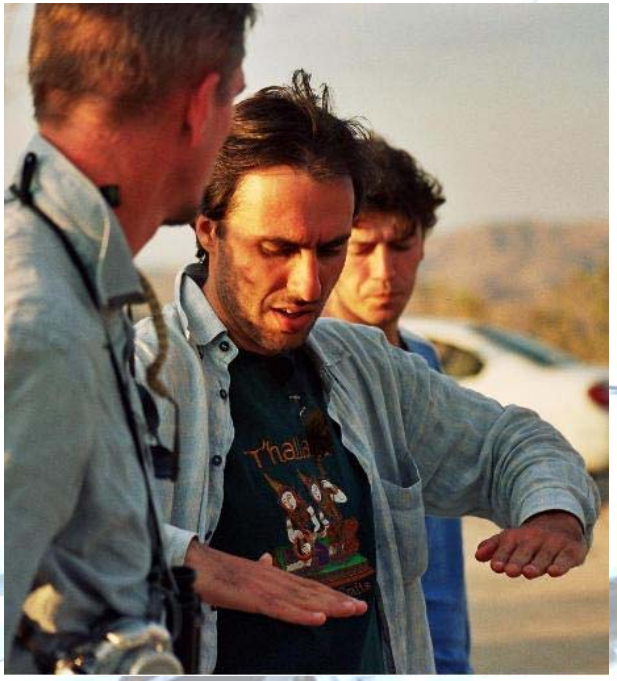
# Humans & Computers: Present

# Humans & Computers: Future

# Humans & Computers: Present

***Human-Human Interaction:***



Simultaneous employment of sight, sound and touch

***Human-Computer Interaction:***

keyboard

mouse

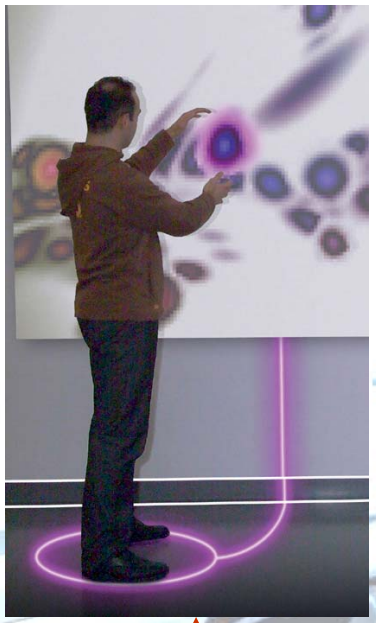
touch screen

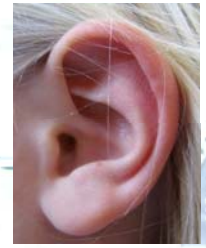
joystick

**Direct manipulation**

Current HCI-designs are single-modal and context-insensitive

# Computers – Future: Human-Centred Computing





**Visual processing**

**Audio processing**

**Tactile processing**

Who the user is?
What his/her task is?
Where he/she is?

**Context-sensitive human behaviour interpretation**

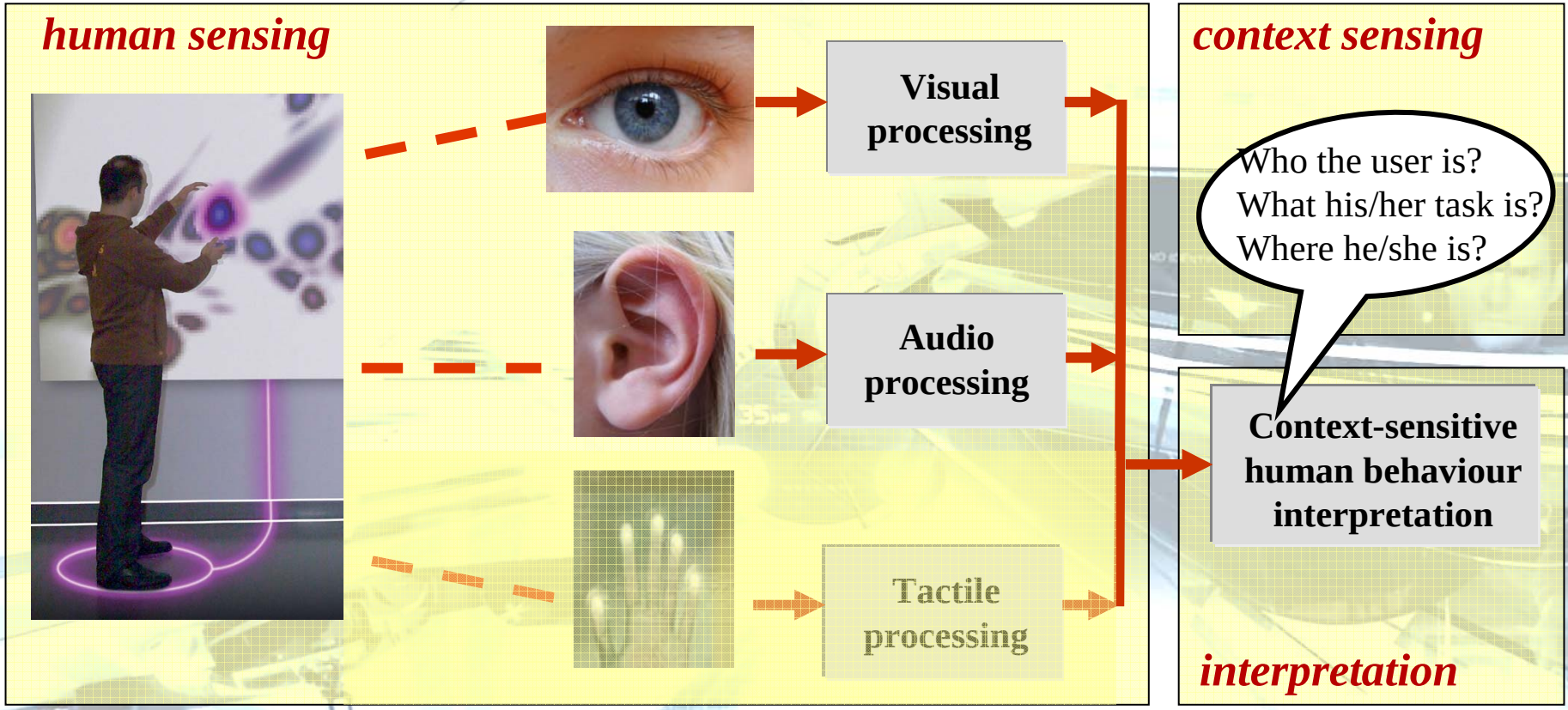
**Context-sensitive responding**

# Human Behaviour Understanding

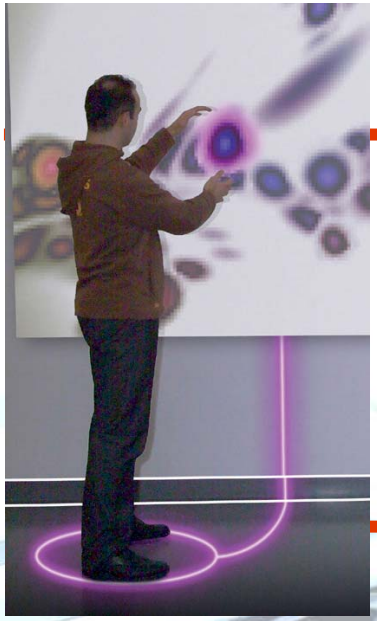
- **Messages**:
  - affect / attitude / cognitive states
  - manipulators
  - emblems
  - illustrators
  - regulators

  *Social Signals*
- **Informative Modalities**:
  - visual (face-, body-, hand gestures)
  - auditory (speech, intonation, non-verbal vocalizations)
  - tactile
- **Context Dependency:**
  - who? where? what? when? why?

# Machine Understanding of Human Behaviour



# Human Sensing



**Visual processing**

**detection, tracking, expression recognition**

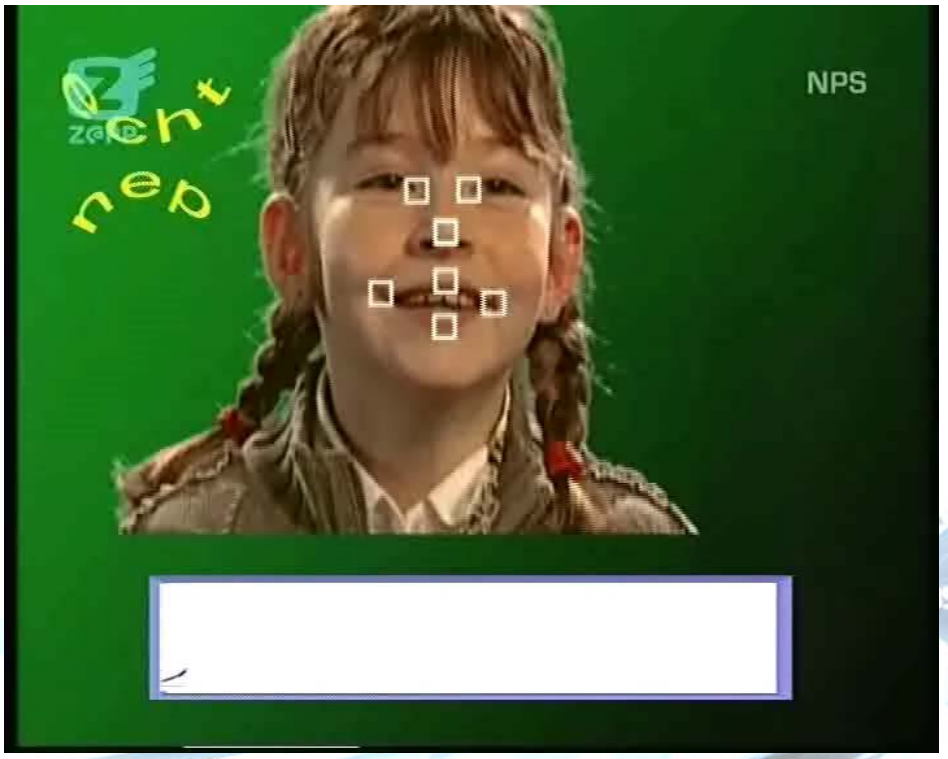
**Audio processing**

**detection, tracking, gesture recognition**

# Vision based Human Sensing – Face

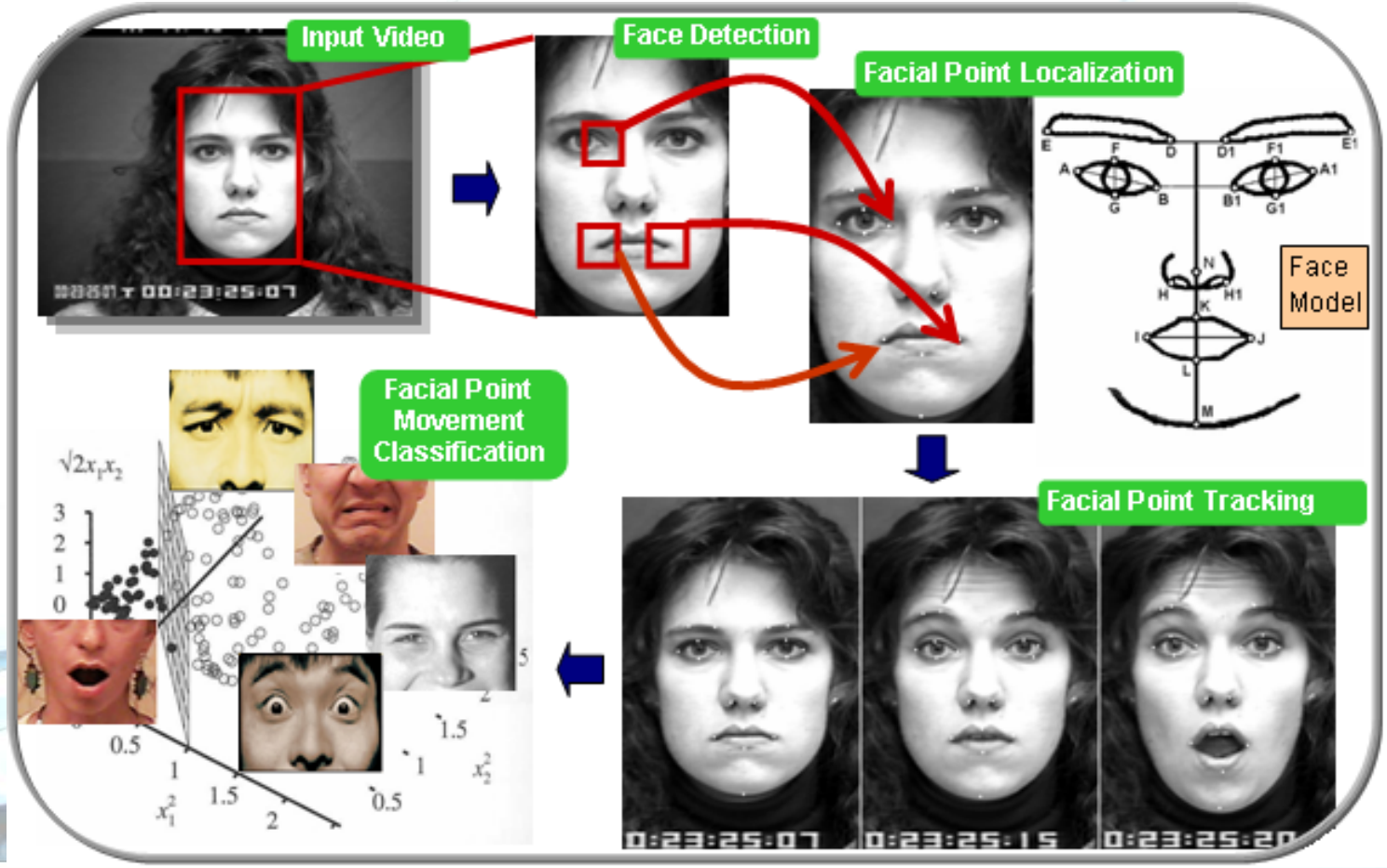
- Bibliography: Keith Price Bibliography (http://iris.usc.edu/Vision-Notes/bibliography/)
- Software: UCSD (mplab.ucsd.edu), CMU (www.ri.cmu.edu/labs/lab_51), …
- Difficulties with: tilted faces, sudden movements, occlusions

# Vision based Human Sensing – Face




- Methodology: optical flow, sequential state estimation (Kalman filtering, Particle filtering)
- Difficulties with: very fast movements, (self-) occlusions, unconstrained environments
- Bibliography: Haykin & Freitas (Eds., 2004), Pantic & Bartlett (2007)

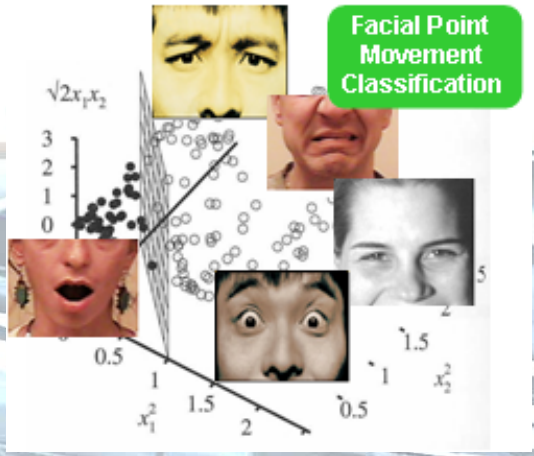
# Vision based Human Sensing – Facial Expression




**A feature-based facial expression analyzer**

# Vision based Human Sensing – Facial Expression

Anger Surprise Sadness Disgust Fear Happiness

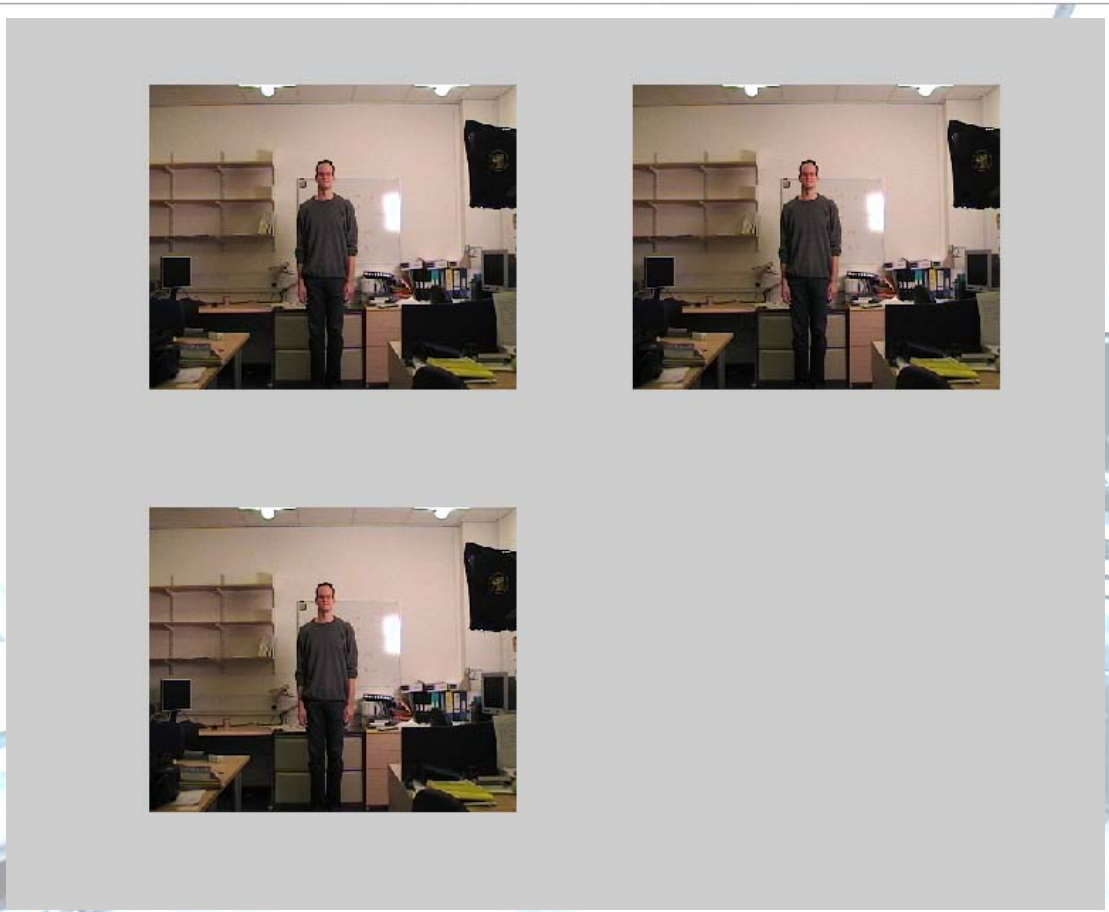




- Methodology: feature / appearance-based; Bayesian, statistical and ensemble learning
- Limited performance and robustness, disregarding temporal dynamics, posed data
- Bibliography: Tian et al. (2005), Pantic & Bartlett (2007), Zeng & Pantic & Huang (2008)

# Vision based Human Sensing – Body

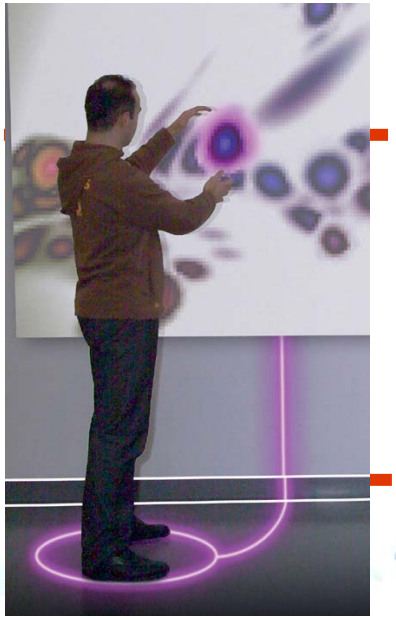
- Methodology: optical flow, sequential state estimation (Kalman filtering, Particle filtering)
- Difficulties with: very fast movements, (self-) occlusions, unconstrained environments
- Bibliography: Wang & Singh (2003), Haykin & Freitas (Eds., 2004), Poppe (2007)

# Vision based Human Sensing – Body

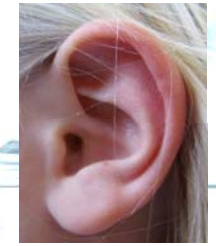


- Methodology: model / appearance-based, probabilistic or statistical learning
- Limited performance and robustness, context insensitivity (Wang et al. 2003; Poppe, 2007)

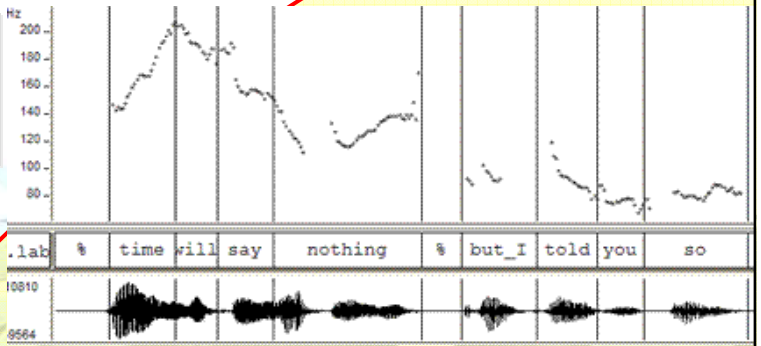
# Audio based Human Sensing



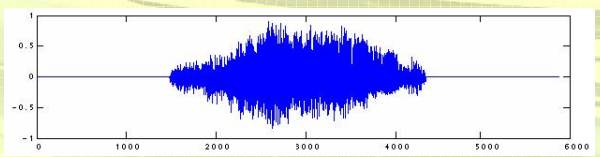
**Visual processing**



**Audio processing**




**extracting auditory features (pitch, intensity…)**




**nonlinguistic vocalizations detection (laughs, cries, sighs, coughs…)**

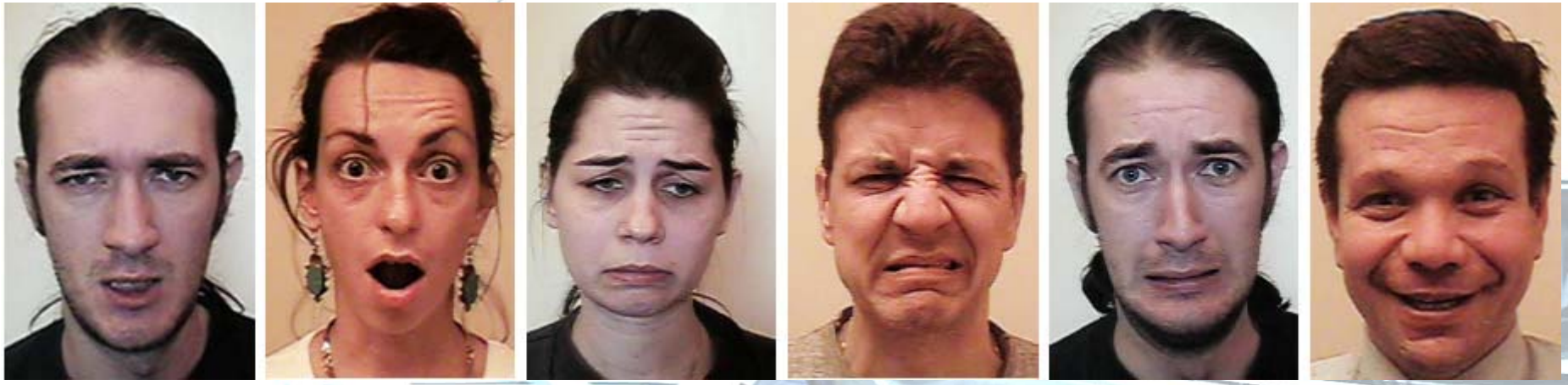
- Auditory features extraction software: Praat, Rasta
- Recent developments: attempts to automatically recognise nonlinguistic vocalizations
- Difficulties with: ambient noise, vocal outbursts detection in speech, naturalistic contexts

# Human Behaviour Understanding

- **Messages**:
  - affect / attitude / cognitive states
  - manipulators
  - emblems
  - illustrators
  - regulators

  *Social Signals*
- **Informative Modalities**:
  - visual (face-, body-, hand gestures)
  - auditory (speech, intonation, non-verbal vocalizations)
  - tactile
- **Context Dependency:**
  - who? where? what? when? why?

# Human Behaviour Understanding - Affect



Anger Surprise Sadness Disgust Fear Happiness

- Six basic emotions: introduced by Charles Darwin (1872), elaborated by Ekman
- Methodology: single- or bi-modal, decision-level data fusion, supervised learning
- Limited performance and robustness, disregarding temporal dynamics, posed data
- Bibliography: Pantic & Rothkrantz (2003), Zeng & Pantic & Huang (2008)

# Human Behaviour Understanding - Affect

**Person-dependent display of anger**

**Spontaneous vs. Posed**

- Recent developments: attempts to automatically recognise non-basic emotions (fatigue, pain), to learn user-profiled interpretations, to discern spontaneous from posed (Zeng et al., 2008)

# Human Behaviour Understanding – Social Signals

- Pentland (MIT): tone of voice → activity level, stress, engagement, mirroring
- Gatica Perez (IDIAP): voice, hand, head movements → level of interest
- El Kaliouby (MIT): facial and head gestures → (dis)agreement, level of interest
- Very limited performance and robustness

# Human Behaviour Understanding

- **Messages**:
  - affect / attitude / cognitive states
  - manipulators
  - emblems
  - illustrators
  - regulators

  *Social Signals*
- **Informative Modalities**:
  - visual (face-, body-, hand gestures)
  - auditory (intonation, non-verbal vocalizations)

  1×
- **Context & Dynamics:**
  - who? where? what? when? why?
  - temporal analysis
  - context-sensitive multimodal fusion

# Human Behaviour Understanding – Challenges

- **Scientific:**
  - **Multimodality**: Which modalities? Fusion level?
  - **Fusion & Dynamics**: Fusion and temporal correlations within and between modalities?
  - **Fusion & Context**: Context-dependent fusion and discordance handling?
  - **Learning vs. Education**: Lazy and unsupervised learning?
- **Technical**:
  - Initialisation
  - Robustness
  - Speed
  - Training & Validation Issues