Social robotics: what is new? Case study with Nadine social robot

Prof. Nadia Magnenat Thalmann

Institute for Media Innovation (IMI)

Nanyang Technological University, Singapore

nadiathalmann@ntu.edu.sg

MIRALab-University of Geneva

thalmann@miralab.ch

Welfenlab, LUH, Hanover , 20 April 2016

Turing test : 1950

Can a machine think?

Put a machine and a human in a room and send in written questions. If we cannot tell which answers are from the machine or the human, the machine is thinking…

What first passed the Turing Test and is it enough?

- The first was **ELIZA**, a program written by the American Computer scientist, **Joseph Weizenbaum (1976**)

```
> Hello, I am Eliza.
* I am afraid.
> How long have you been afraid?
* Since midterms.
> Tell me more...
* I am afraid that I will fail
my course.
> Do you enjoy being afraid that
you will fail your course?
```

- BUT anything like human intelligence must be able to engage with the **real world, with social interaction** , and the Turing Test doesn't test for that..

Where are we today?

- After 40 years:

 automatic answer systems, GPS guiding drivers , software controlled cars, drones, planes and many other things…

 misunderstanding of language, limited reasoning, social clues…

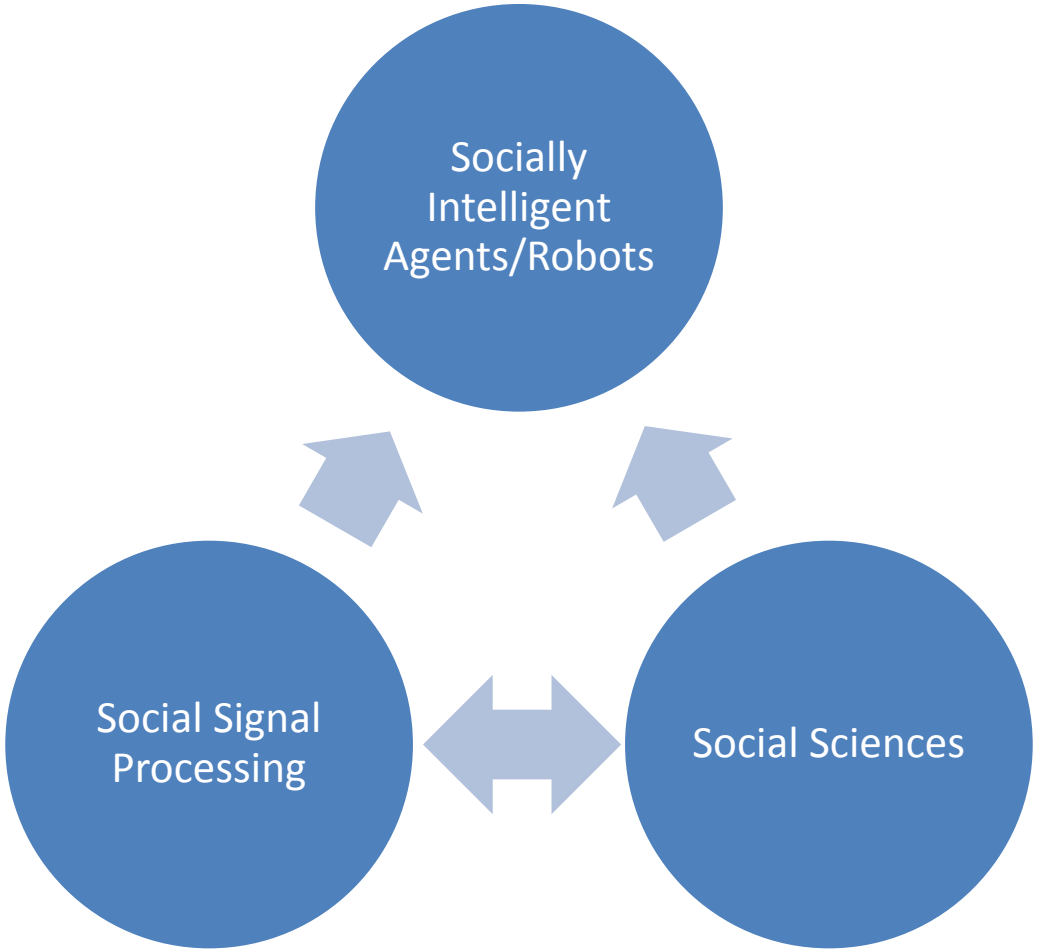
What has changed in computers?

- **40 years ago**, computers: mainly CPU + Memory + very limited I/O.
- **Today**: CPU and memory are much faster AND incredible possibilities of interfacing with people through **sensors and actuators**.

What has changed in computers?

- Hardware/software tools allowing to capture, understand, **reproduce a lot of signals**: speech, sounds, gestures, shapes, forces etc…
- **Generate Big Data** that allows us to analyse and model events or predict the future using deep learning algorithms

Three closely related disciplines



Challenges in autonomous behavior generation

- Virtual characters and robots interacting with people in social contexts
  - should understand the other users' behaviors,
  - and respond back with gestures, facial expressions and gaze.

- **Challenges:**
  - Sensing and interpreting other users' behaviors, intentions
  - Making decisions appropriate to the social situation based on partial sensory input
  - Rendering synchronized and timely multi-modal behaviors

Challenges for behavior understanding

- **Context-dependent**: W5+ (where, what, when, who, why and how)
  - e.g. "Smile can mean different
    - a display of politeness (social signal
    - Being happy (affective cue)
    - Joy of seeing a friend (affective cue/social signal)
    - Empathy (emotional response/social signal)
    - Greeting (social signal)
    - Irony/irritation (affective cue/social signal)

rather unexplored area of research: Recognizing communicative intention

A. Vinciarelli, M. Pantic, H. Bourlard. **Social Signal Processing: Survey of an Emerging Domain**, Image and Vision Computing Journal, Vol. 27, No. 12, pp. 1743-1759, 2009.

Remembering past interactions

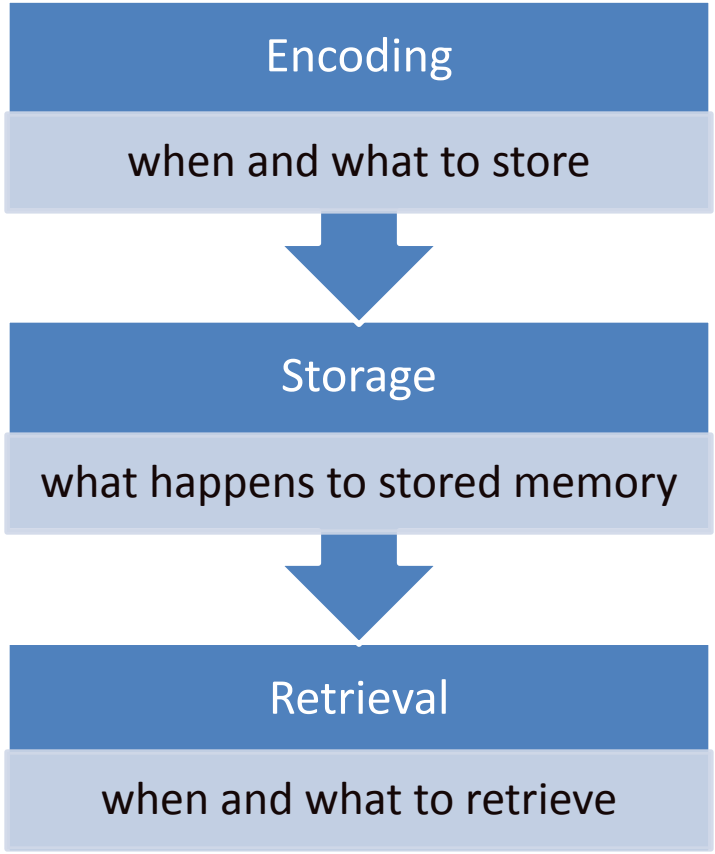
- Episodic memory is the memory of autobiographical events (times, places, associated emotions, and other contextual who, what, when, where, why knowledge) that can be explicitly stated
  - keeping the course of dialogue
  - planning long-term goals
  - explaining reasons for actions
  - learning from past experiences
  - requires a personal history of an entity

[Tul72] E. Tulving, "Episodic and semantic memory," In E. Tulving and W. Donaldson (Eds.), Organization of memory. New York: Academic Press,1972, pp. 381–403.

Episodic Memory

- Conceptual definitions (Tulving [Tul72] , Schank [SA77]: collects past personal experience with specific time, location and context

  inspiring but lack of details for implementation of robots/VH
- Findings from social sciences
  - Three phases of EM
  - Forgetting and recency effect
    - emotional memories are remembered more



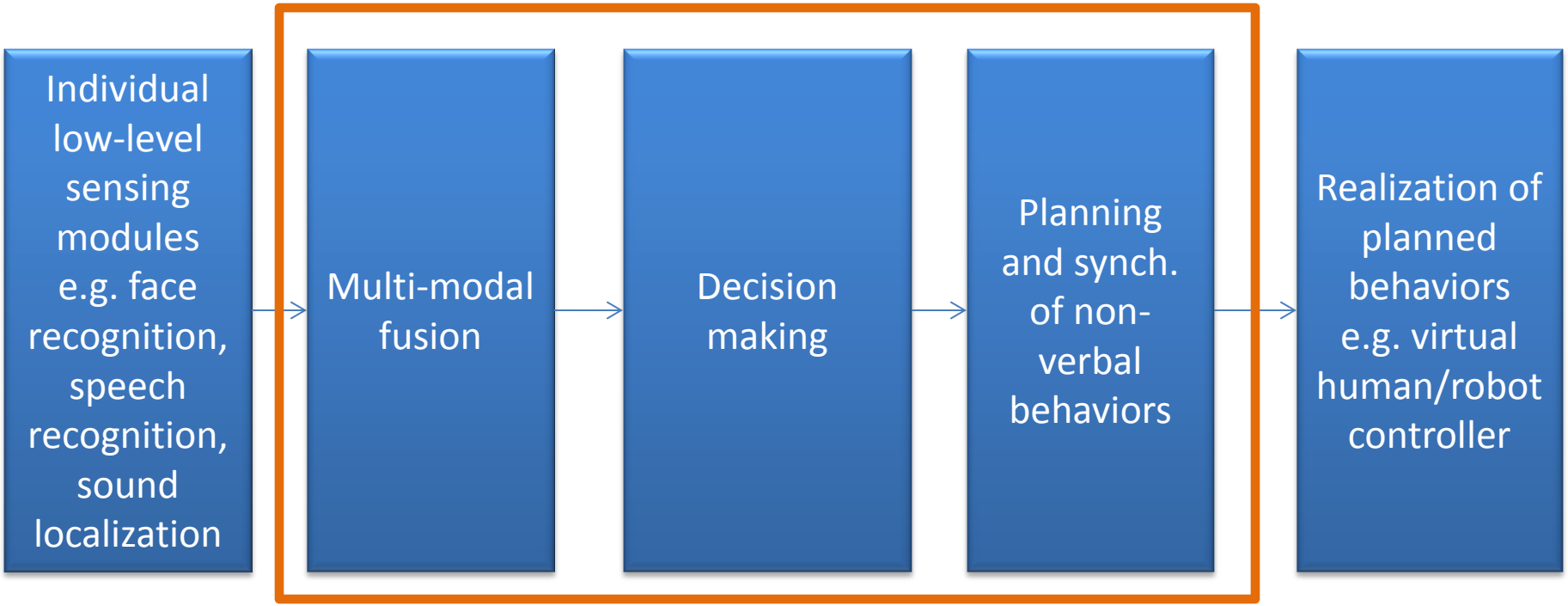
[Tul72] E. Tulving, "**Episodic and semantic memory**," In E. Tulving and W. Donaldson (Eds.), Organization of memory. New York: Academic Press,1972, pp. 381–403.
[SA77] R. C. Schank and R. P. Abelson, **Scripts, Plans, Goals and Understanding: an Inquiry into Human Knowledge Structures**. L. Erlbaum, 1977.

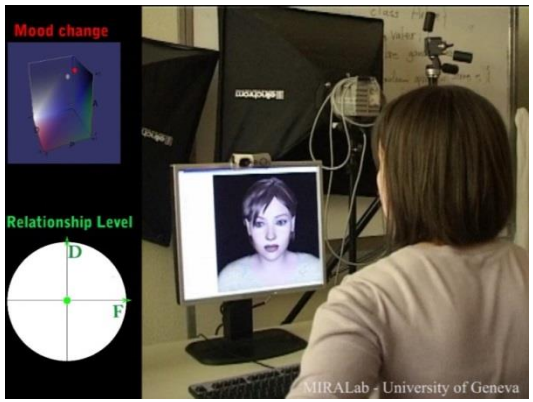
Research with Social Robot EVA (MIRALab-Unige (2008-2012)

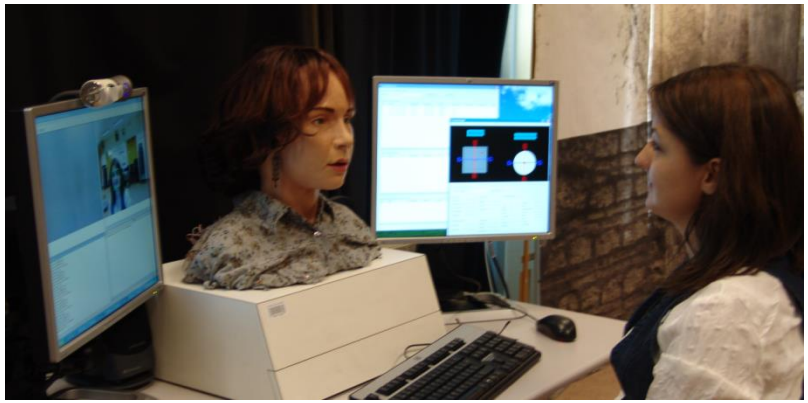
- Overall Goal: long-term social interaction framework with a human-like robot or Virtual Human: **modeling emotions, episodic memory and expressive behaviour**
- Goal: **remembering individuals** (faces and names) and **past exchanges** over multiple interactions

Eva teaching introductory computer networks concepts

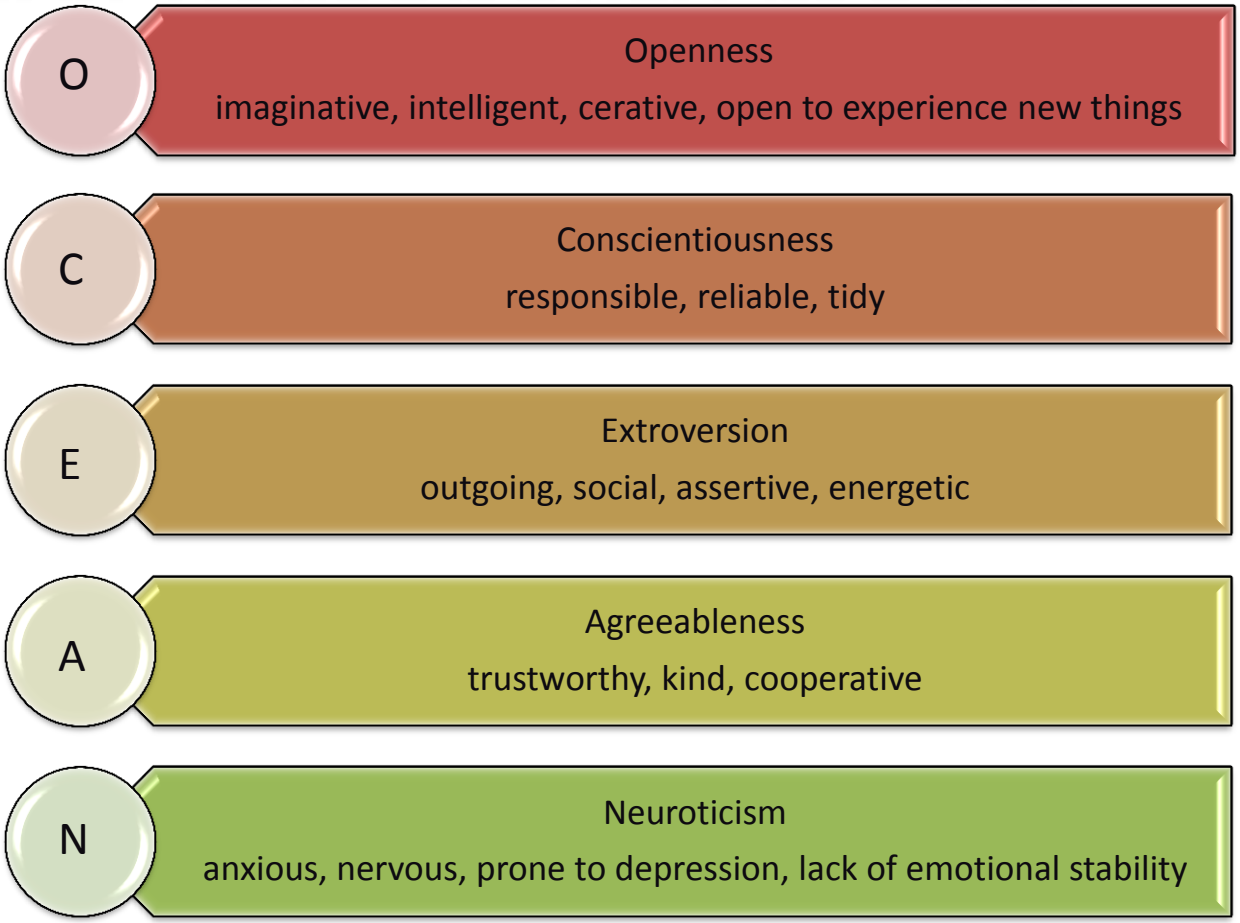
Steps in multi-modal interaction







OCEAN Personality model



Mccrae R. R. and John P.O. (1982) An introduction to the five-factor model and its applications. Journal of Personality, 60:175–215.

Mood and Personality

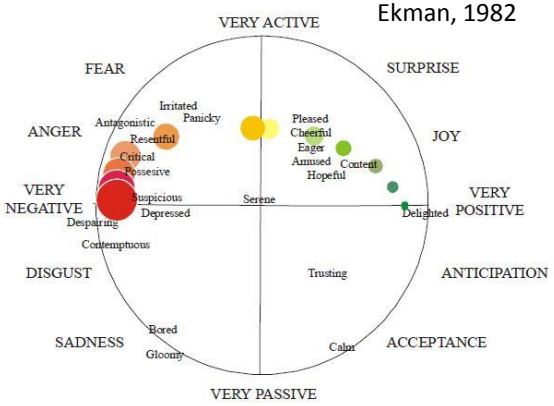
- Mood differ from emotion in three aspects (Steed et al., 2006):
  - temporal: last longer than emotions
  - expression: emotions are associated with a facial expression, moods are not
  - cause: not associated with a specific event
- Personality differ from emotion in two aspects (Steed et al., 2006):
  - duration: remain constant, do not change over time
  - focus: not specific to particular events

Steed et al. (2006) "Building Expression into Virtual Characters," *Annual Conference European Association for Computer Graphics, State of the Art Reports.*

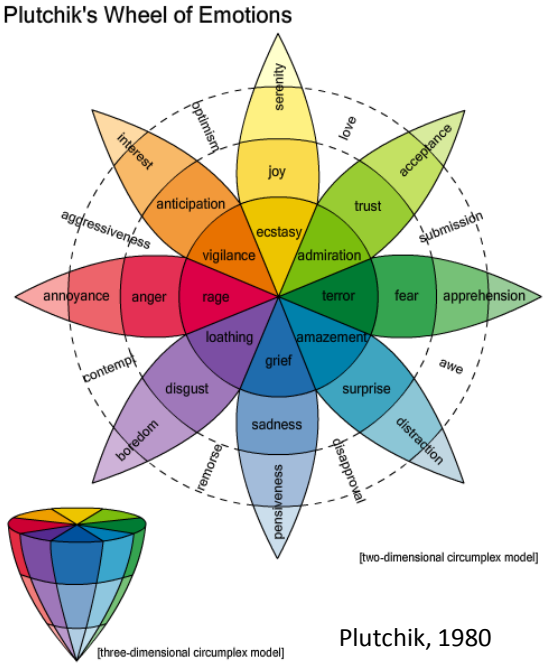
Modelling Emotions

Different representations of emotions: discrete, dimensional

Ekman, 1982



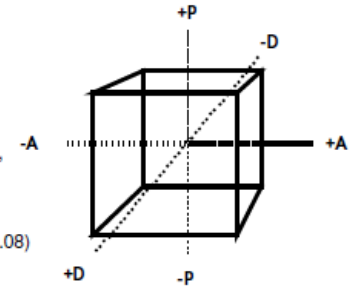
Whissel, 1989



Plutchik, 1980

The following sample ratings illustrate definitions of various emotion terms when scores on each PAD scale range from -1 to +1:

angry (-.51, .59, .25)
bored (-.65, -.62, -.33)
curious (.22, .62, -.01)
dignified (.55, .22, .61)
elated (.50, .42, .23)
hungry (-.44, .14, -.21)
inhibited (-.54, -.04, -.41),
loved (.87, .54, -.18)
puzzled (-.41, .48, -.33)
sleepy (.20, -.70, -.44)
unconcerned (-.13, -.41, .08)
violent (-.50, .62, .38).



The emotional state "angry" is a highly unpleasant, highly aroused, and moderately dominant emotional state. The "bored" state implies a highly unpleasant, highly unaroused, and moderately submissive state.

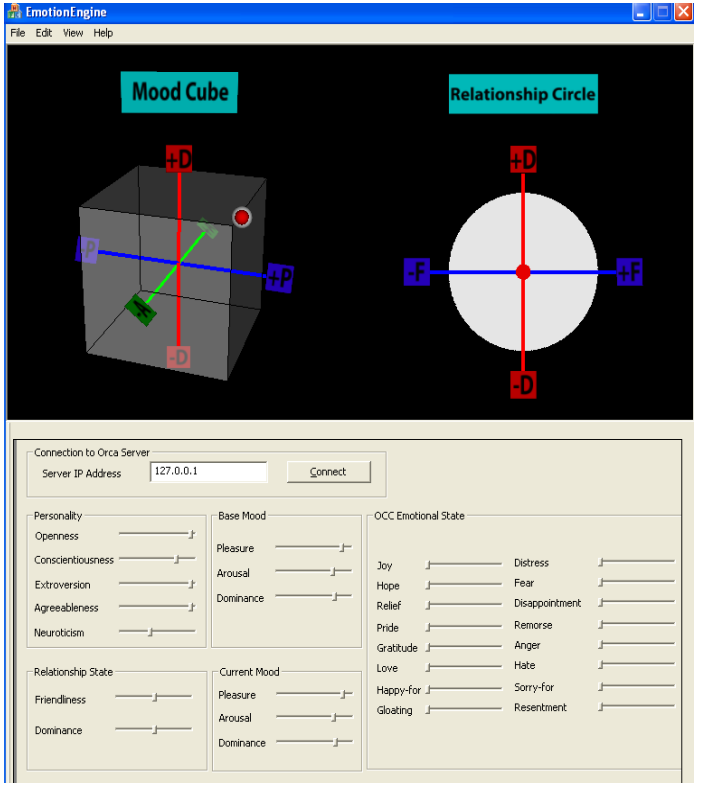
The Mehrabian P-A-D Temperament Scale

Mehrabian, 1996

[1] P. Ekman, **Emotion in the Human Face**. Cambridge University Press, NewYork, 1982.
[2] A. Mehrabian, **Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament**, Current Psychology, vol. 14, pp. 261–292, 1996.
[3] R. Plutchik, **A general psycho-evolutionary theory of emotion**, R. Plutchik and H. Kellerman (Eds.), Emotion: Theory, research, and experience,vol. 1, pp. 2–33, 1980.
[4] C. M. Whissel, **The dictionary of affect in language**, R. Plutchik and H. Kellerman (Eds.), Emotion: Theory, research, and experience, vol. 4, 1989.

Theory of affective states

- OCEAN personality [1] big five
  - Openness, conscientiousness, extroversion, agreeableness, neuroticism
- 16 OCC emotions [2]
  - Joy, hope, relief, pride, gratitude, love, happy-for, gloating, distress, fear, disappointment, remorse, anger, hate, sorry-for, resentment
- Pleasure-Arousal-Dominance(PAD) model for mood [3]
  - Positive: Exuberant, dependant, relaxed, docile
  - Negative: Bored, disdainful, anxious, hostile
- Two-dimensional relationship model [4]
  - Friendliness and dominance



[1] R. R. Mccrae and P. John, **An introduction to the five-factor model and its applications**, Journal of Personality, vol. 60, pp. 175–215, 1992.

[2] A. Ortony, G. L. Clore, and A. Collins, **The Cognitive Structure of Emotions**. Cambridge University Press, 1988.

[3] A. Mehrabian, **Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament**, Current Psychology, vol. 14, pp. 261–292, 1996.

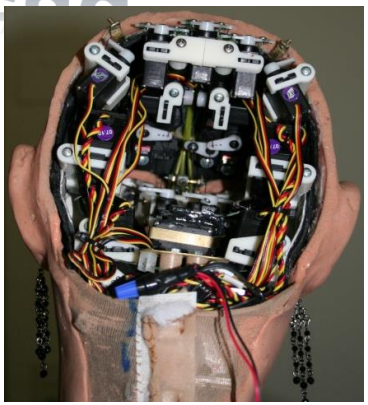
[4] M. Argyle, **Bodily communication**, second ed. Methuen and Co Ltd, 1998.

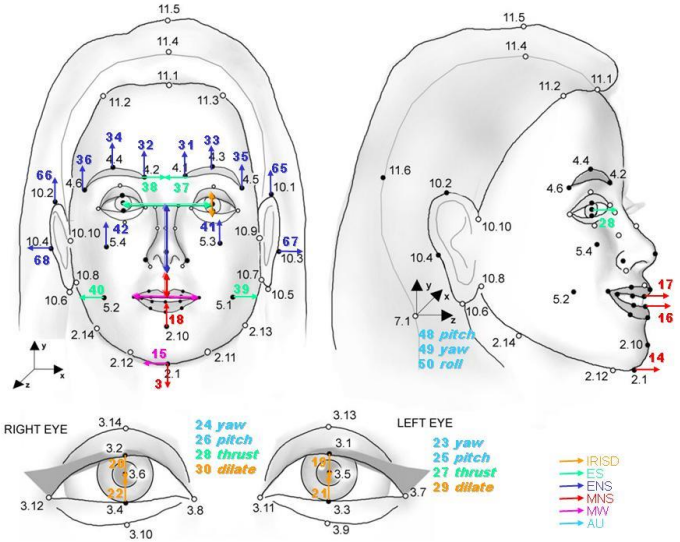
Emotion and Mood Update (PAD model)

- Four cases :
  - Case 1: Initialization of base mood with personality
  - Case 2: At the beginning of each interaction session when a person is recognized
  - Case 3: At the end of each interaction session when a person leaves
  - Case 4: At each emotional impulse during dialogue

Animating our MIRALab robot head

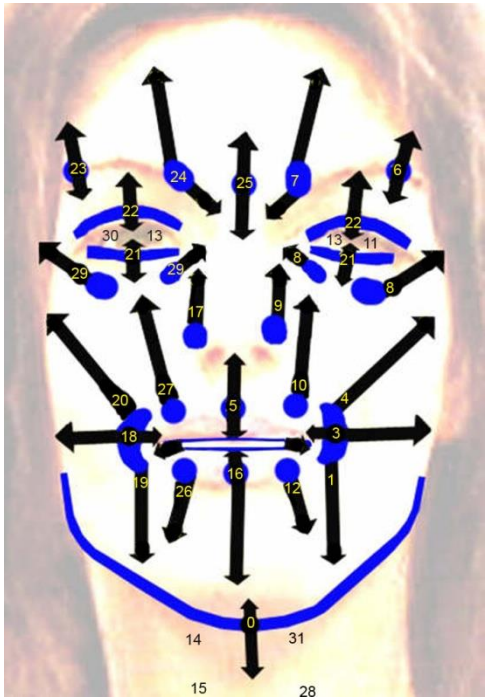
- Controlled by 32 Servo-motors
  - Bi-directional
  - Inside the head
  - Values between 0 and 250
- Skin deformed by Servo-motors pulling and pushing the skin





MPEG-4 based face engine*

MPEG-4 FAPs to Robot Servo-motors

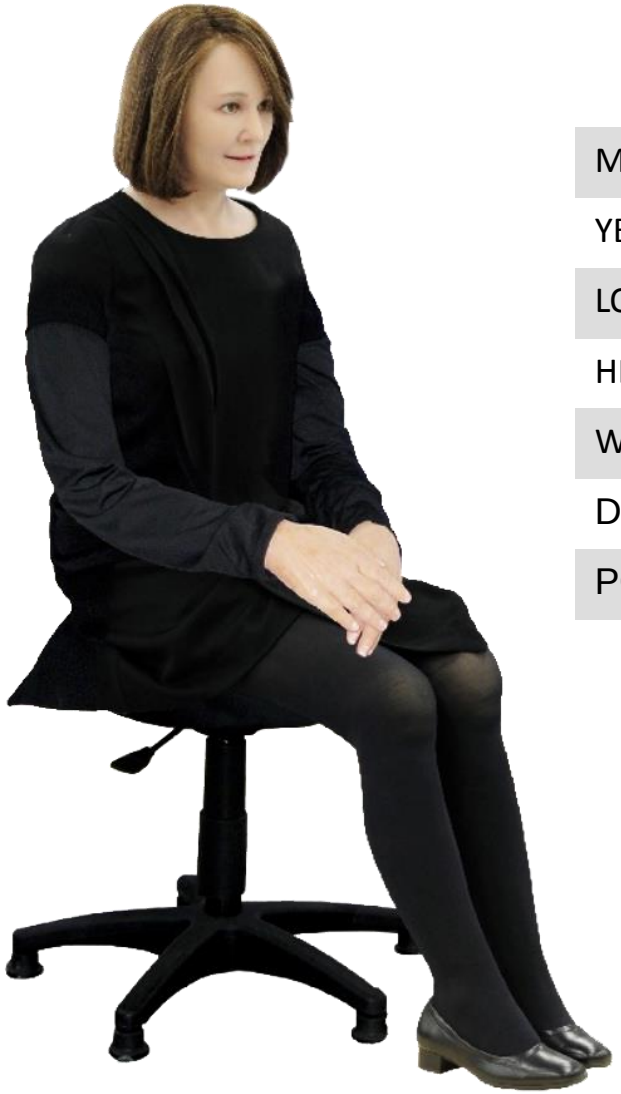


MIRALab robotic tutor (2008-2012)

Mixing real people with autonomous virtual humans and social robots (BeingThere Centre)

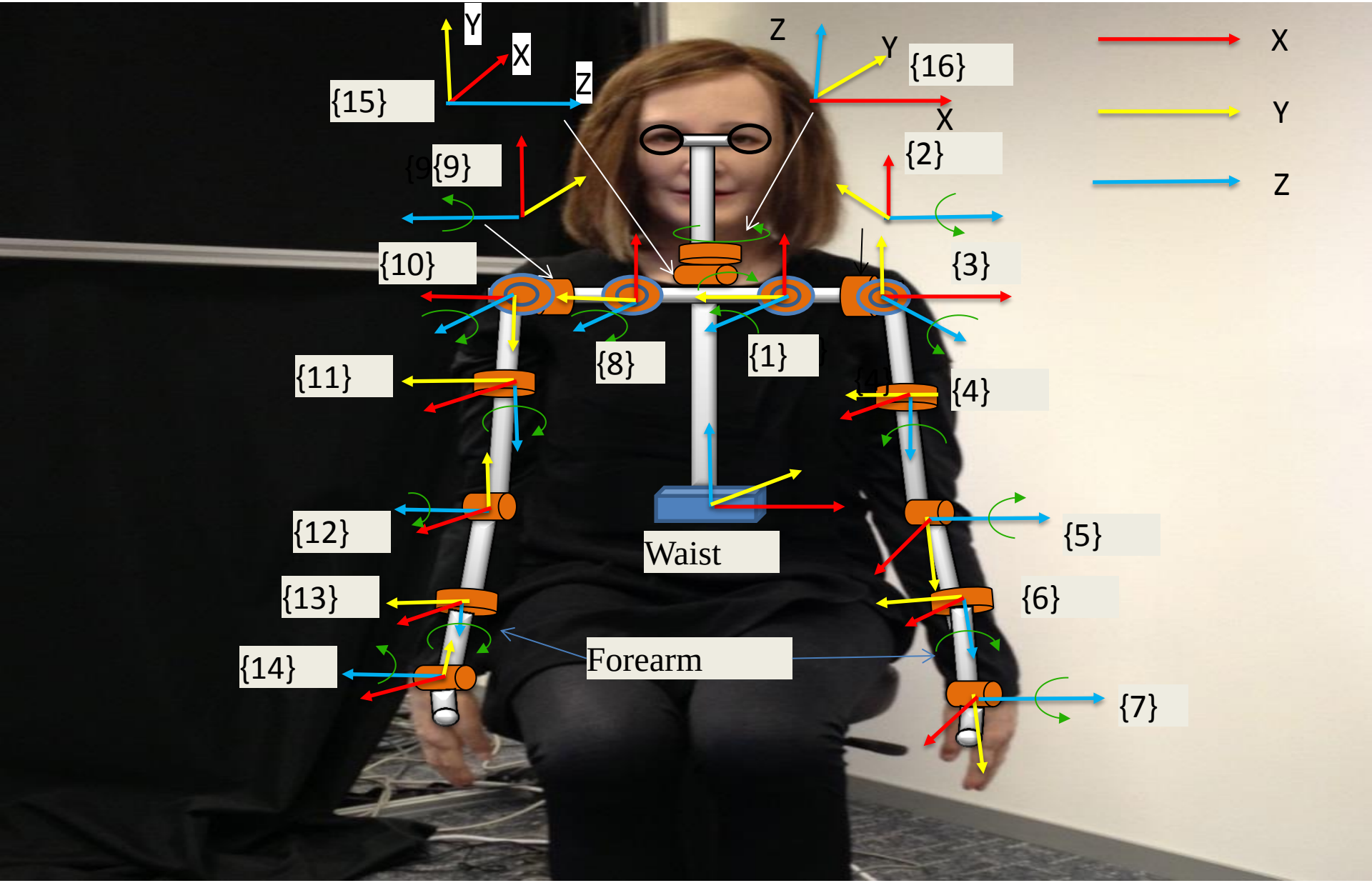
http://imi.ntu.edu.sg/BeingThereCentre/Pages/BTChome.aspx

Nadine

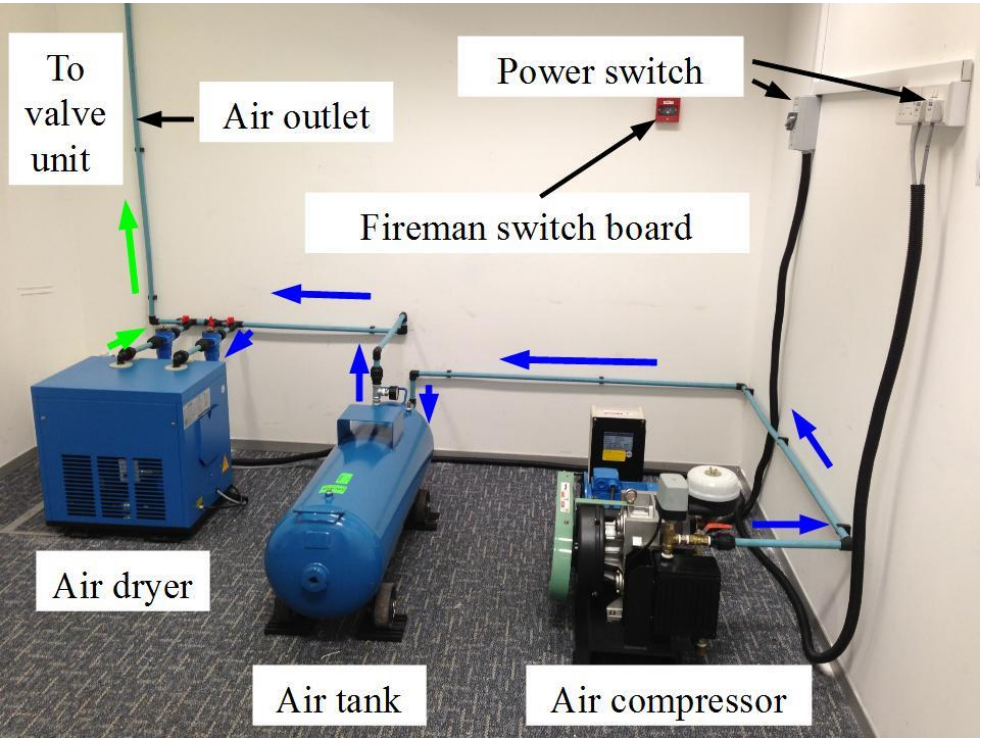


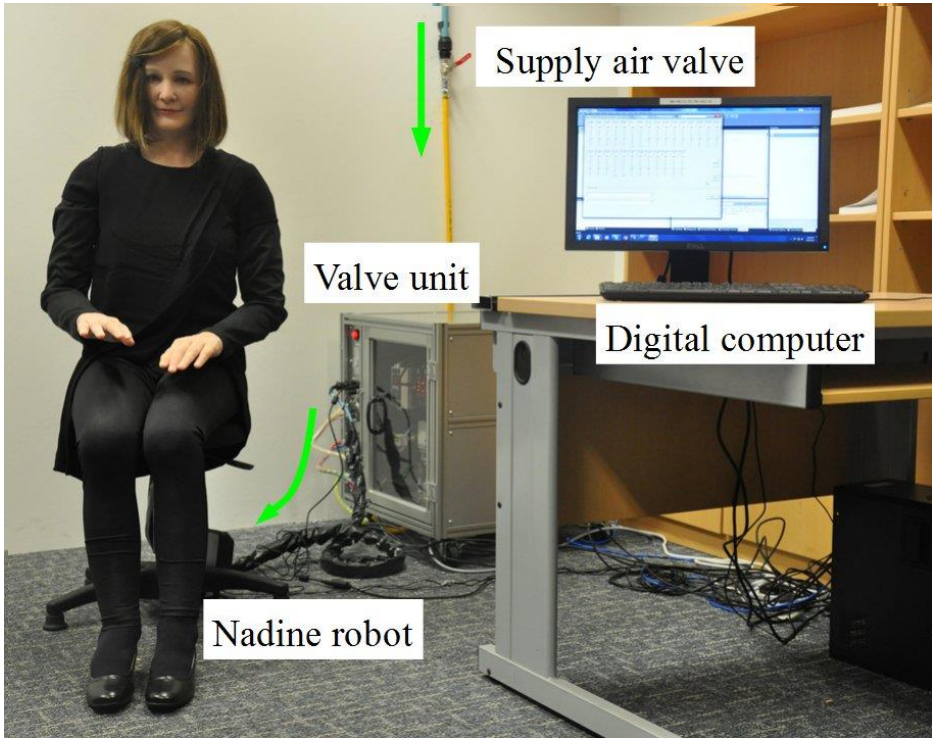
SPECIFICATIONS

MANUFACTURER	IMI, NTU / Kokoro Company Ltd.
YEAR OF CREATION	2013
LOCATION	Singapore
HEIGHT	4.31 ft (131.5 cm)
WEIGHT	77.1 pounds (35 kg)
DEGREES OF FREEDOM	27 DOF
POWER	500W

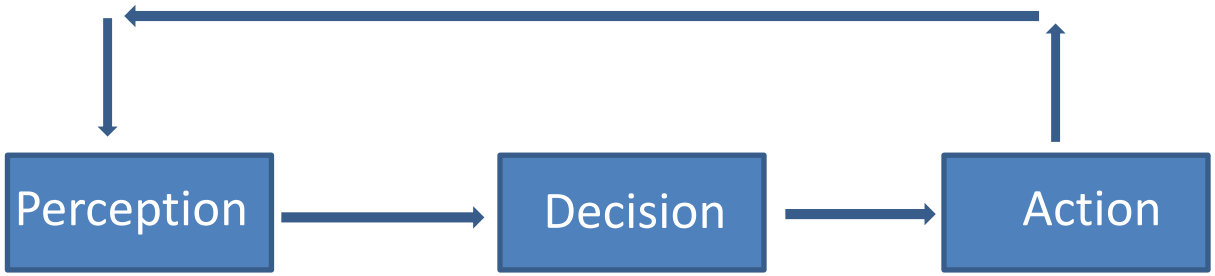

Y
X
Z
{15}
Z
Y
{16}
X
X
Y
Z
{9}
{2}
{10}
{3}
{8}
{1}
{11}
{4}
{12}
{5}
Waist
{13}
{6}
{14}
Forearm
{7}

Hardware


To valve unit
Air outlet
Power switch
Fireman switch board
Air dryer
Air tank
Air compressor


Supply air valve
Valve unit
Digital computer
Nadine robot

Perception/Decision/Action



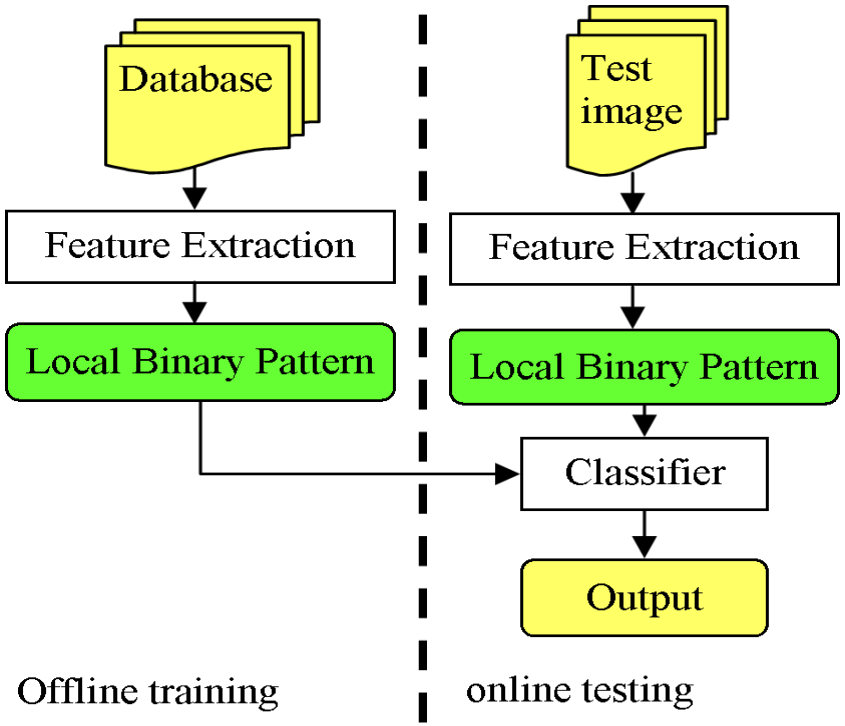
- Microsoft Kinect V2:
  - Face recognition
  - Gestures recognition
  - Understanding of social situations
- Microphone:
  - Speech recognition

- Emotion Model
- Memory Model
- Social Attention
- Chatbot

- Robot controller:
  - emotional and gesture expressions
  - Lips synchronization
  - online gaze generation

Face Recognition using the Kinect

- Novelty: Utilize **extremely fast** Local Binary Patterns features
- Training: 20 seconds
- Recognition: real-time



1. Jianfeng Ren, et al., **Learning Binarized Pixel-Difference Pattern for Scene Recognition,** 2013 IEEE International Conference on Image Processing (ICIP)
2. Jianfeng Ren, et al. **Relaxed Local Ternary Pattern for Face Recognition**, 2013 IEEE International Conference on Image Processing (ICIP)
3. Jianfeng, Ren et al, "**Dynamic Texture Recognition Using Enhanced LBP Features**", ICASSP 2013.

Face Recognition

Body Gesture Recognition

Novelty: sensor fusion of Cyber glove and Kinect
Allows gesture recognition for **any orientation of the hand**

Hand feature extraction
(Cyber-glove)

+

Upper body joints extraction
(Kinect)

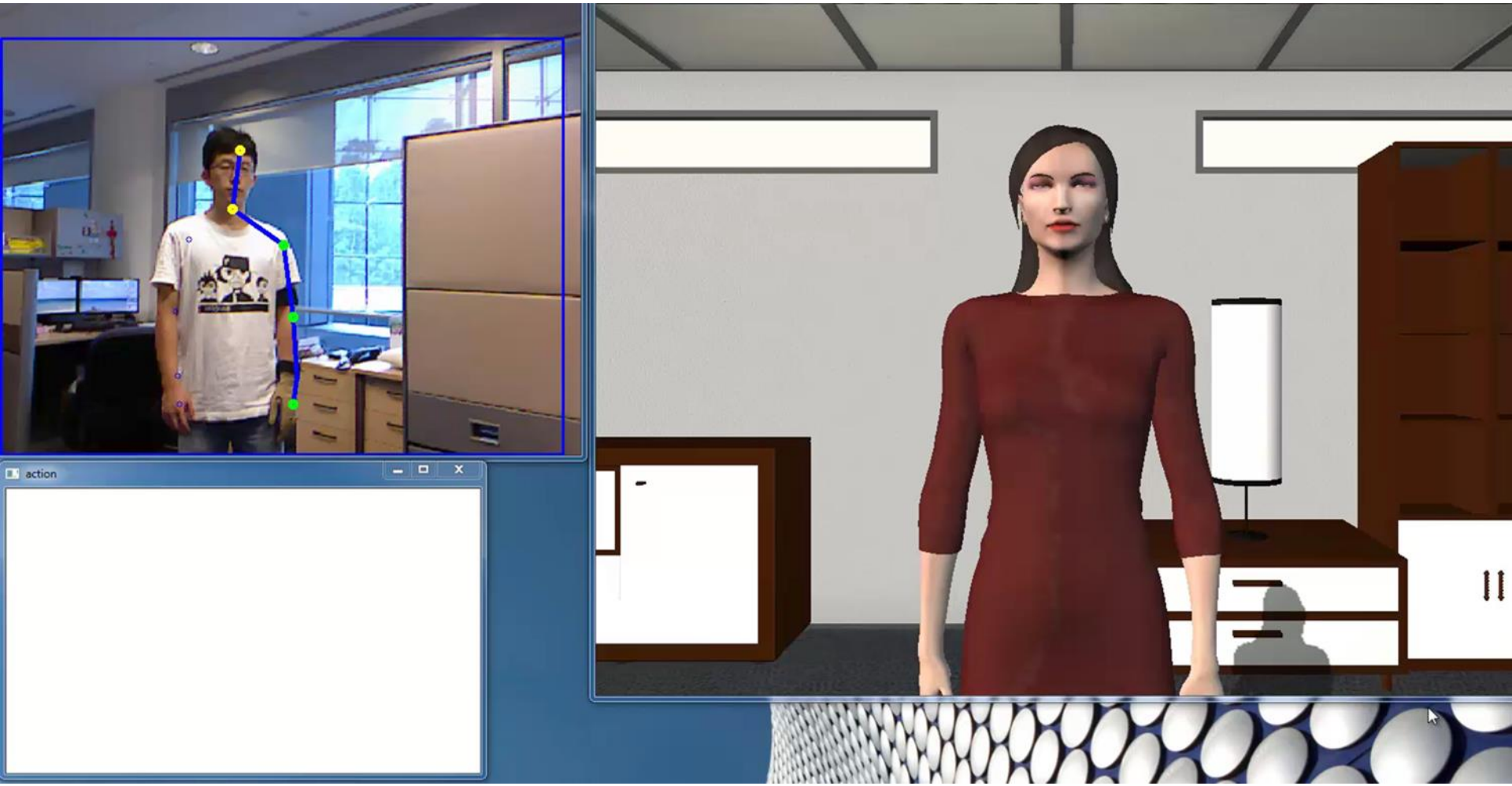
→ Information fusion

↓

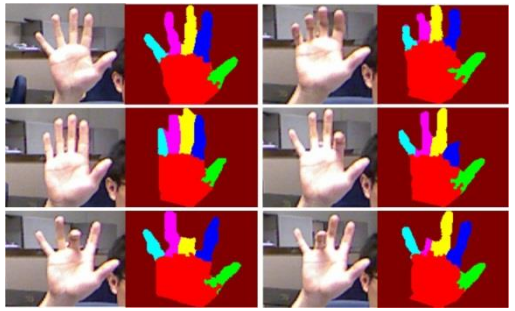
Recognize upper body gesture

Yang Xiao; Junsong Yuan; Daniel Thalmann, **Human-Virtual Human Interaction by Upper Body Gesture Understanding**, Proc. ACM VRST 2013

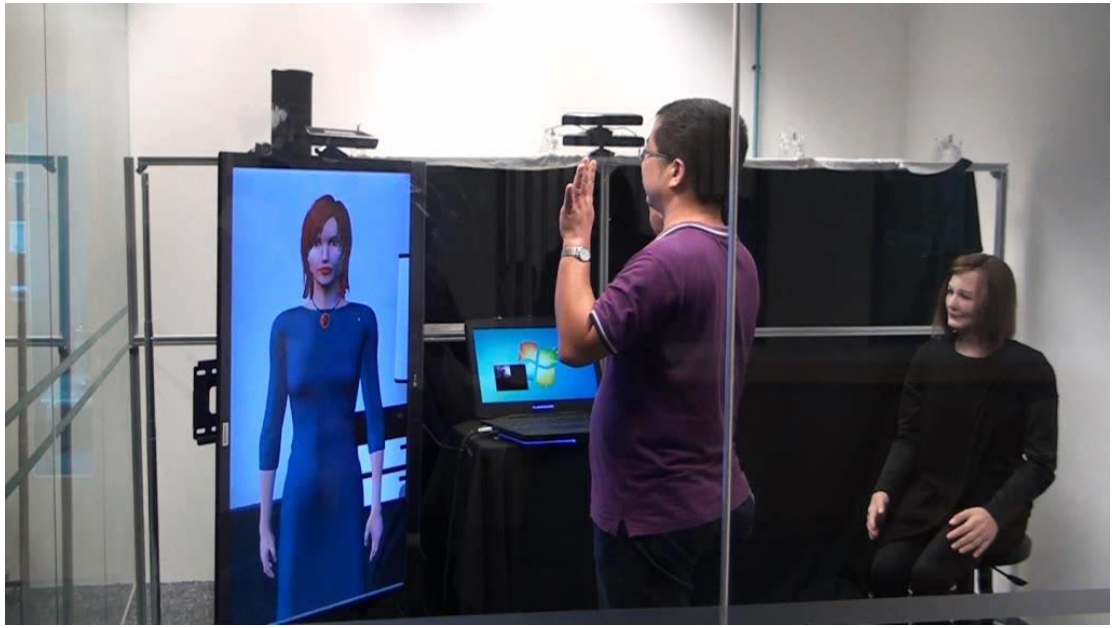
Body Gesture Recognition



Hand Gesture Recognition



- **Novelty:** spatial-temporal feature, which enforces both spatial and temporal constraints in a unified framework for hand parsing and fingertip detection.
- **Result**: more accurate compared to existing methods.

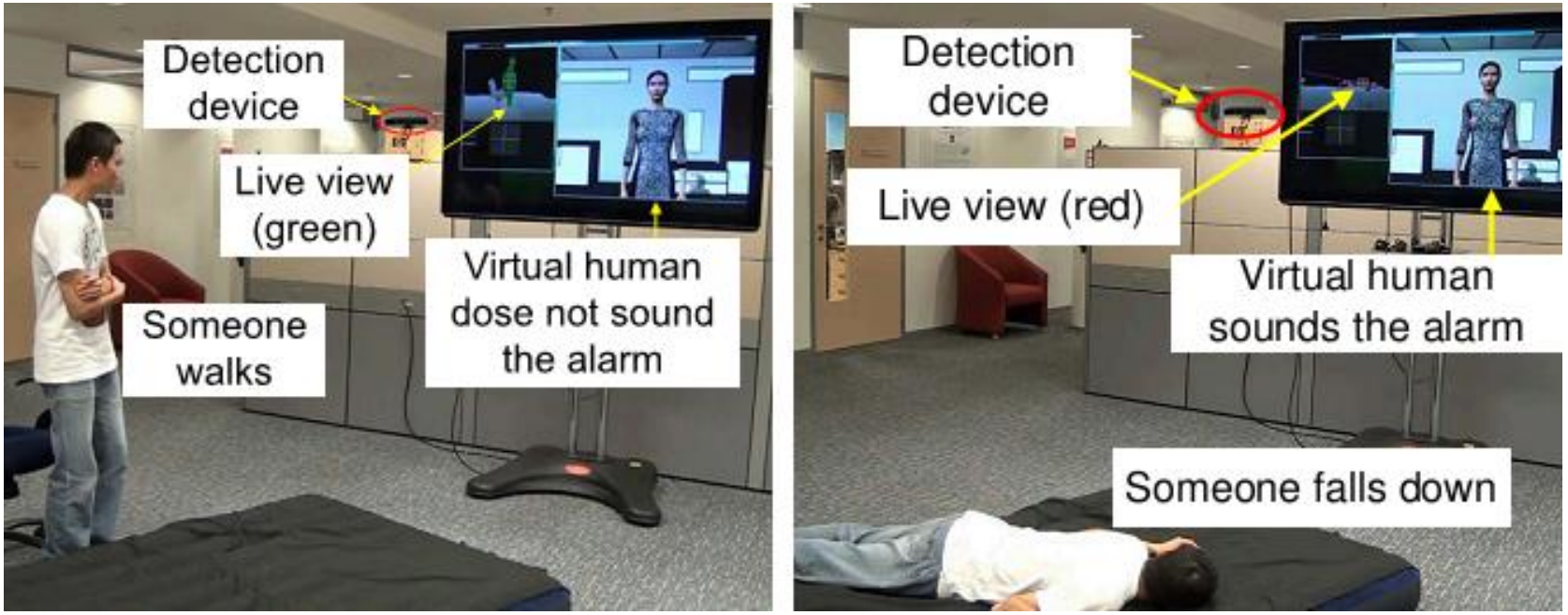


Hui Liang, et al. **3D Fingertip and Palm Tracking in Depth Image Sequences**, ACM International Conference on Multimedia 2012 (MM)
Hui Liang, Junsong Yuan and Daniel Thalmann, "**Parsing the hand in depth images**" IEEE transactions on Multimedia, to appear

Falling down recognition

- Virtual human used to detect falling down

  --a robust fall detection approach by analyzing the tracked key joints of the human body using a single depth camera.

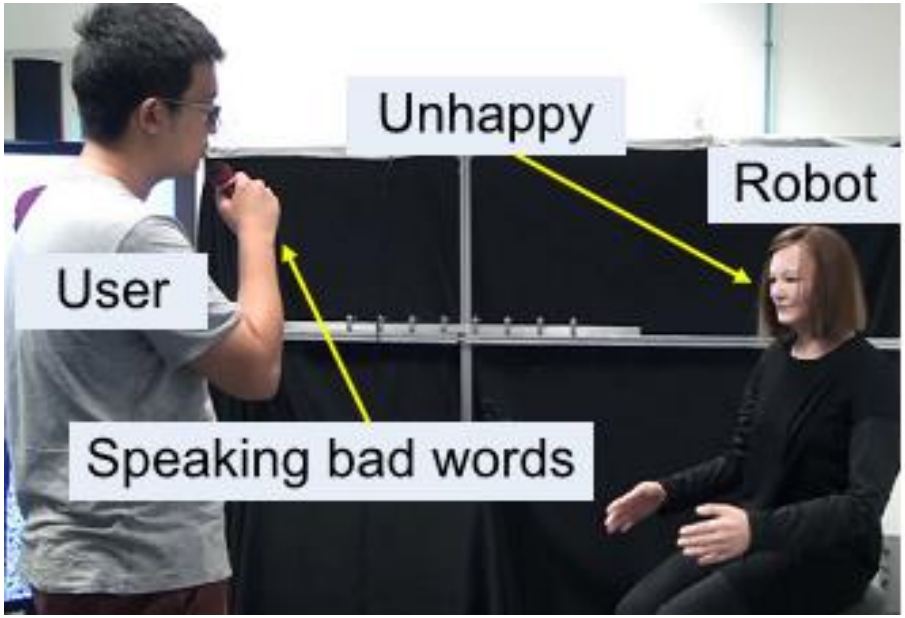


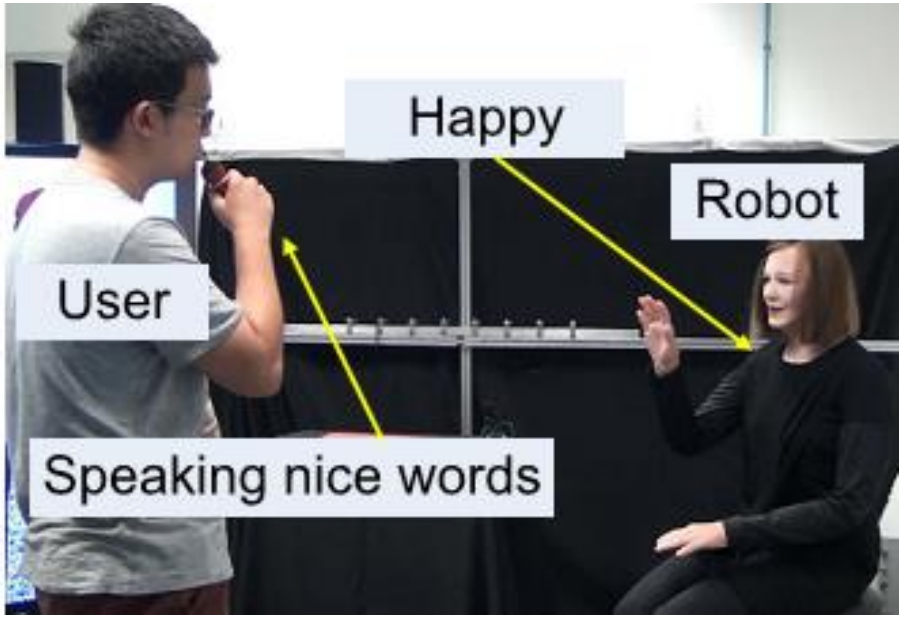
1)Z. P. Bian, L. P. Chau, and N. Magnenat Thalmann, **Human Computer Interface for Quadriplegic People based on Face Position/Gesture Detection**, Proceedings ACM Multimedia 2014

2)Z. P. Bian, J. Hou, L. P. Chau, and N. Magnenat Thalmann, **Fall Detection Based on Body Part Tracking Using a Depth Camera**, IEEE Journal of Biomedical and Health Informatics , 2014

Emotional interaction

--Robot behavior changes with mood dynamics and user inputs.

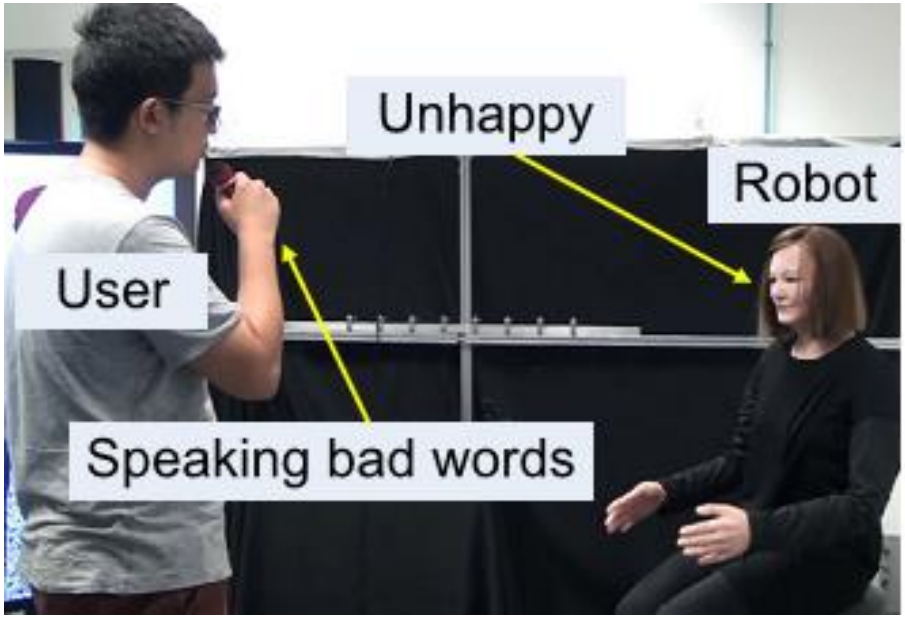


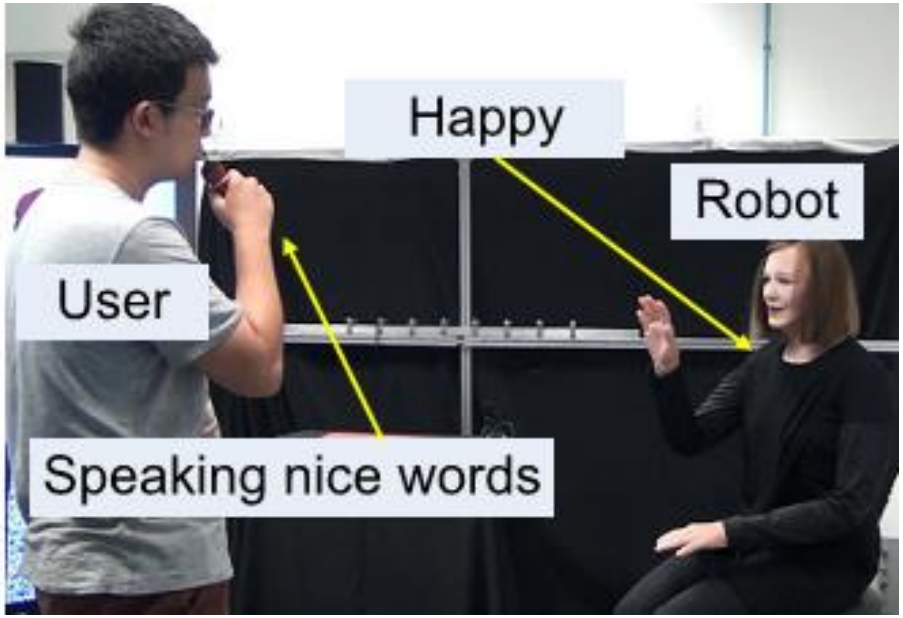


Juzheng Zhang, Jianmin Zheng, Nadia, Magnenat-Thalmann, **Design Affective System for Virtual Human and Social Robots**, CASA 2012, Singapore, May 9-11.

Emotional interaction

--Robot behavior changes with mood dynamics and user inputs.

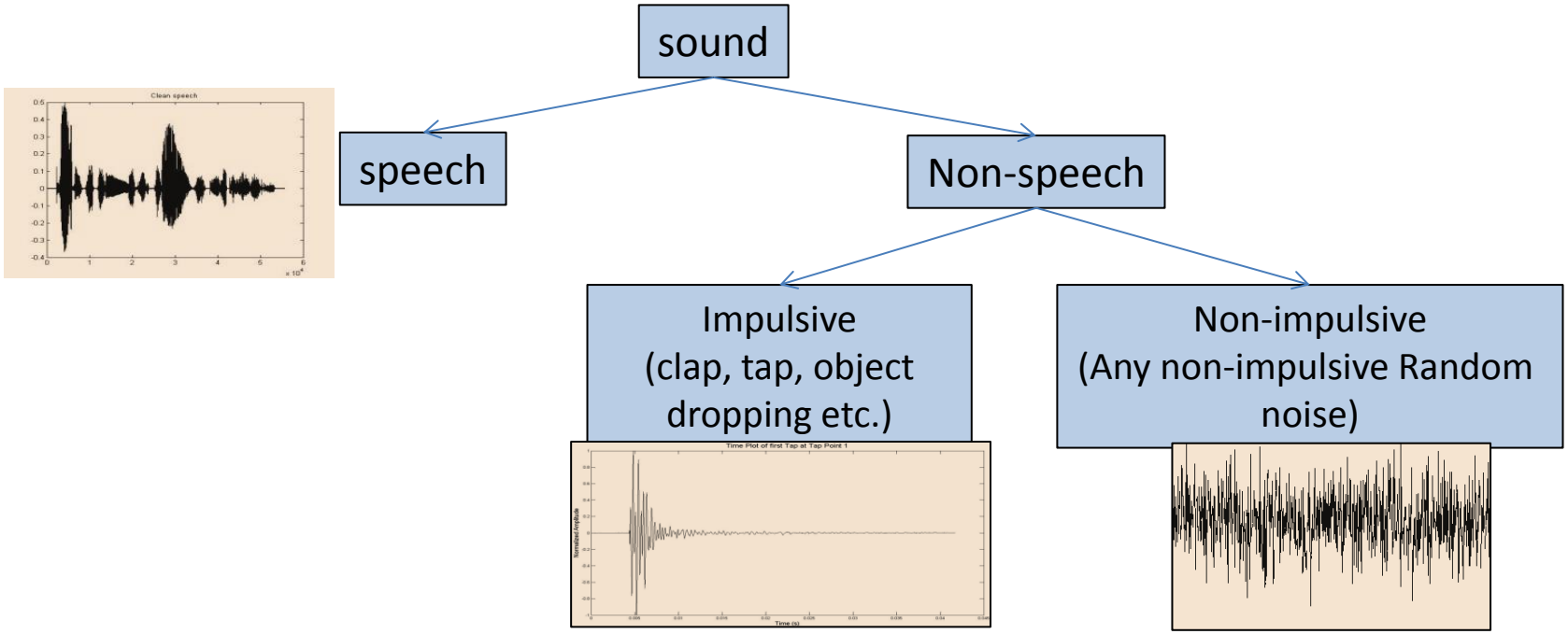




Juzheng Zhang, Jianmin Zheng, Nadia, Magnenat-Thalmann, **Design Affective System for Virtual Human and Social Robots**, CASA 2012, Singapore, May 9-11.

Sound Localization with Classification

- Novelty: a real-time algorithm
  - Can localize the sound by estimating the direction of arrival
  - Can classify the sound into several categories



R. S. Rashobh and A. W. H. Khong, "**A fast frequency domain algorithm for equalizing acoustic impulse responses**,"*IEEE Signal Process. Lett.*, vol. 19, no. 12, pp. 797–800, Dec. 2012

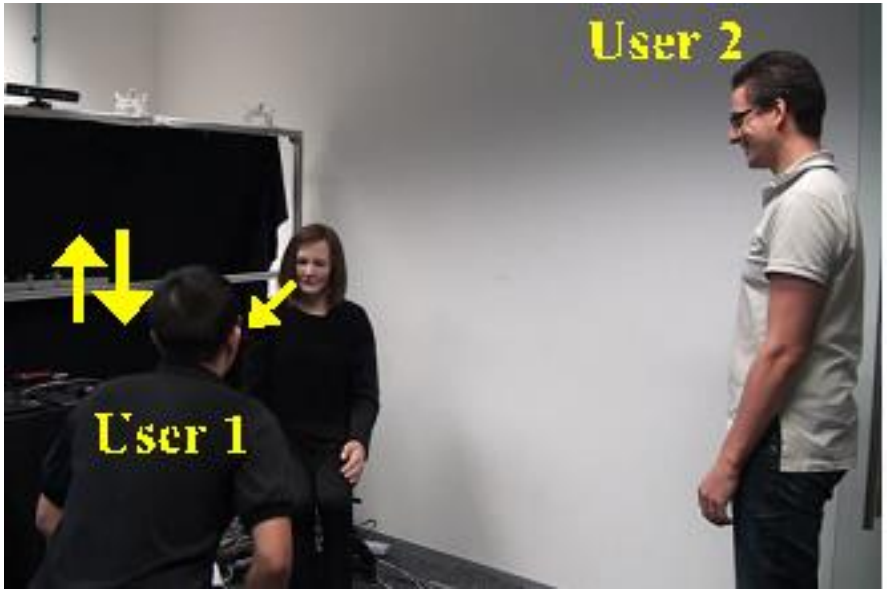
Demo: Sound Localization and Classification

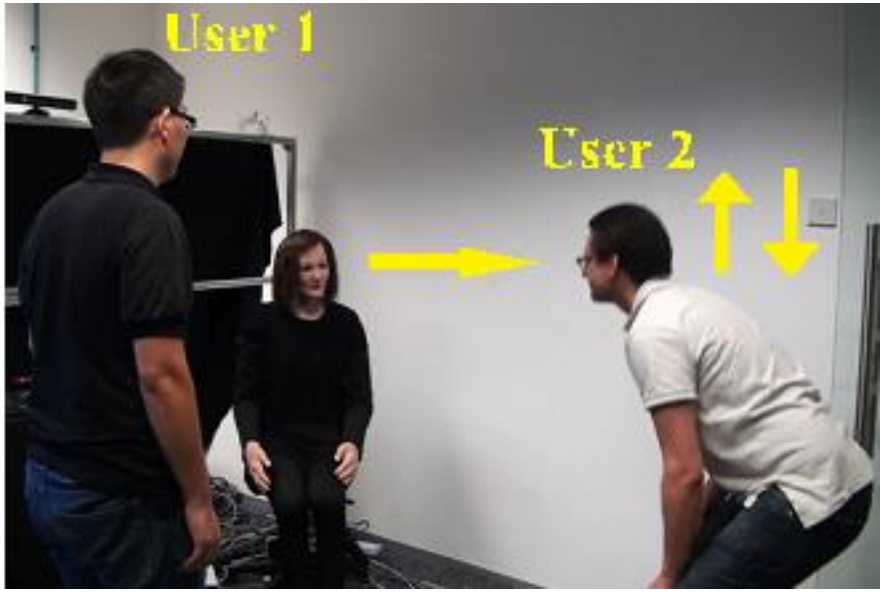
Real-time pointing towards the sound source location

Z. Zhang, A. Beck and N. Magnenat Thalmann, **Human-like Behavior Generation Based on Head-arms Model for Tracking External Targets and Body Parts**, IEEE Transactions on Cybernetics, 2014

Attention System

-- selectively concentrates on specific object in the environment while ignoring other things of the surrounding





Real-time interaction with Nadine in telepresence environment

- Real-time interaction with Nadine in video telepresence

[1] Demo was showed at Swissnex Singapore End of Year Party 2013

Nadine singing at Swissnex Party