# Cognitive Systems 2012

# ALIZ-E

*Adaptive strategies for sustainable long-term social interaction*

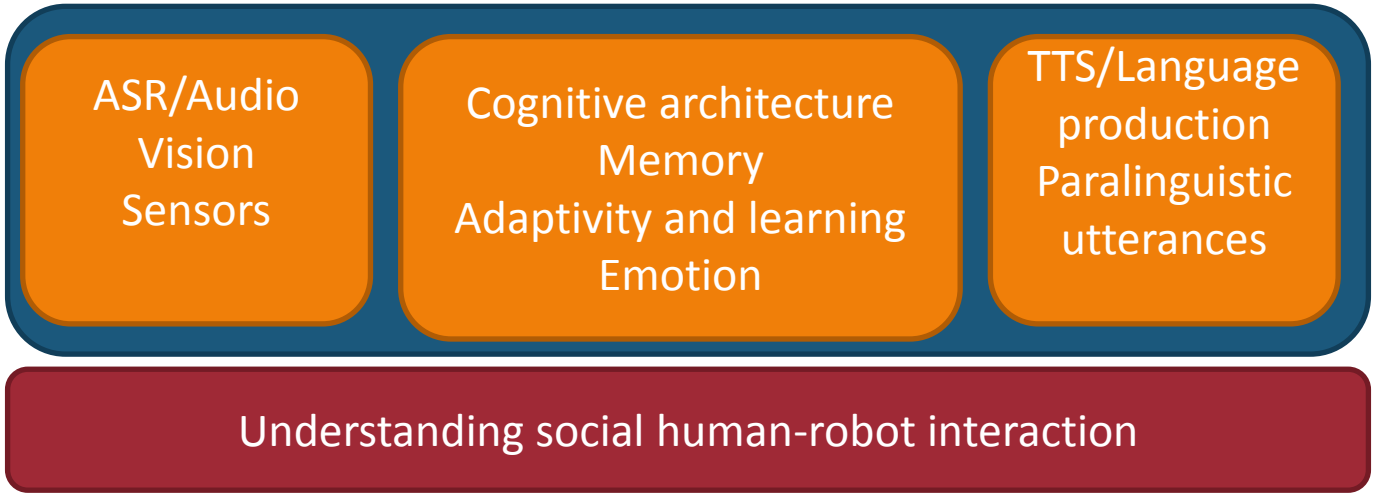
Tony Belpaeme

Plymouth University


ALLIGA

# Overview

- Develop **science and technology** behind cognitive robots which are capable of maintaining **long-term interaction with young users**.
  - Rooted in social behaviour: cognition is inherently and continuously grounded in social interaction.



# Research challenges...

- Human-robot interaction from the range of minutes to the **range of days**.

(Cynthia Breazeal)

(Takanori Shibata)

(Takayuki Kanda)

- Robotic companions in **child-robot interaction**.
  - Different from adult-robot interaction, more promising applications.
- Robust **"any-depth" interaction**.
  - Robustness against low-quality perception and interpretation.
- **Long-term memory** and self-sustained long-term interaction.
  - Key to long-term interaction is having a personalised adaptive memory storing experiences and interaction episodes.

# Research challenges

- Out of the lab into the **real world**: therapeutic settings, **yearly evaluation** with end users.
  - Evaluation in paediatrics department of Hospital San Raffaele.
- Analysis and synthesis of emotion and **affect** in human-robot interaction.
- Pervasive **learning and adaptation**.
- **Speech recognition**.
- **Cloud computing**
  - Extending embedded computational power by using networked processing.

# HRI and healthcare

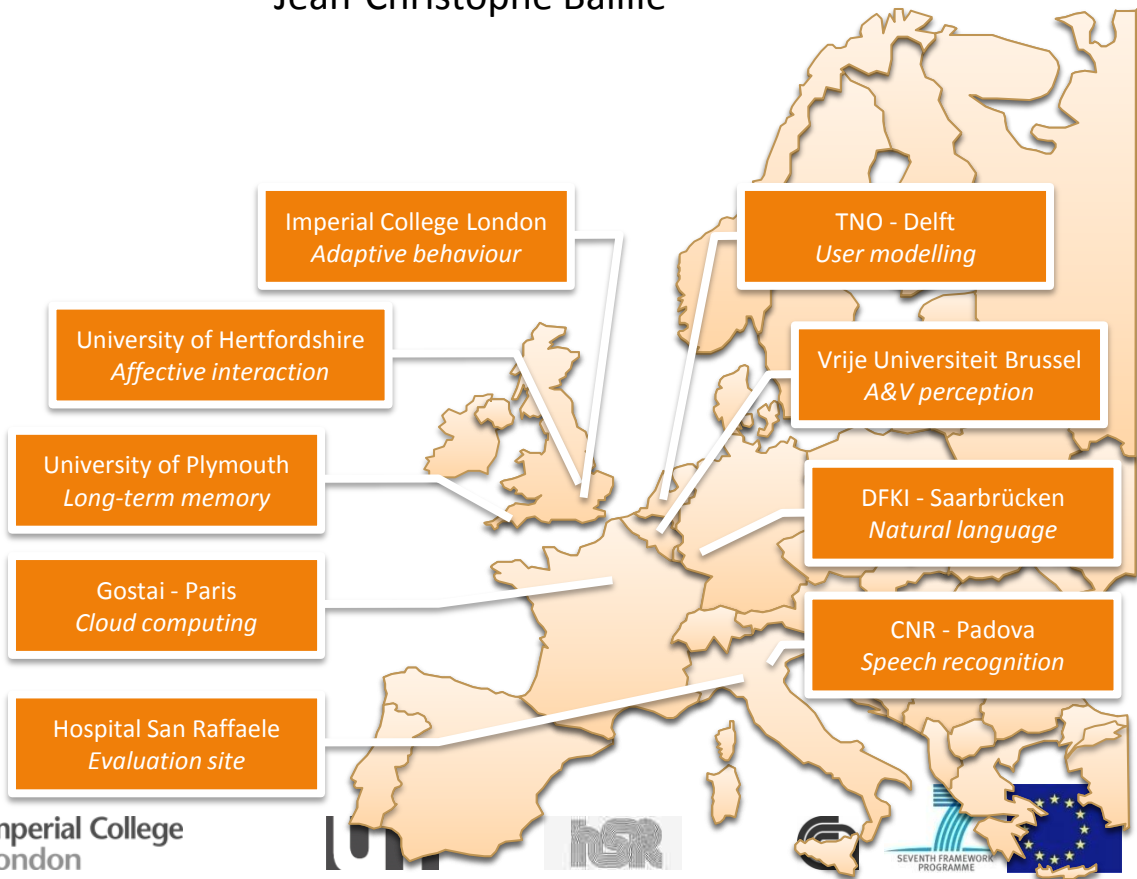
- Variety of potential healthcare applications.
  - Provision of health education
  - Supporting communication between patients and healthcare professionals
  - Providing entertainment for patients.
- Interactions with human users as naturalistic as possible.
- Need for progress in sustaining non-continuous, temporally extended social interactions with users.

(Boston Globe)

San Raffaele hospital, Milan

# Axis of evil

1. **University of Plymouth**
   Tony Belpaeme, Angelo Cangelosi
2. **DFKI**
   Ivana Kruijff-Korbayova, GJ Kruijff
3. **Vrije Universiteit Brussel**
   Hichem Sahli, Werner Verhelst
4. **TNO**
   Mark Neerincx
5. **Imperial College**
   Yiannis Demiris
6. **University of Hertfordshire**
   Lola Cañamero
7. **San Raffaele del Monte Tabor**
   Alberto Sanna, Marco Nalin
8. **National Research Council, Italy**
   Piero Cosi
9. **Gostai , France**
   Jean-Christophe Baillie



# Aldebaran Nao

- Aldebaran Nao humanoid robot.
  - 58cm, 4.3kg, 25 DOF.
  - Sensors and actuators: 2 speakers, 4 microphones, 2 cameras, gyro, accelerometer, range sensors (2 IR, 2 sonar).
  - AMD GEODE 500Mhz, 256Mb, 1Gb flash (new Intel Atom 1.6Ghz, 2Gb, 4Gb flash).
  - Price: about 12k€.
  - Appearance and size: perfect for user base.
  - Functionality: unrivalled at the moment.
  - Project hits ground running.

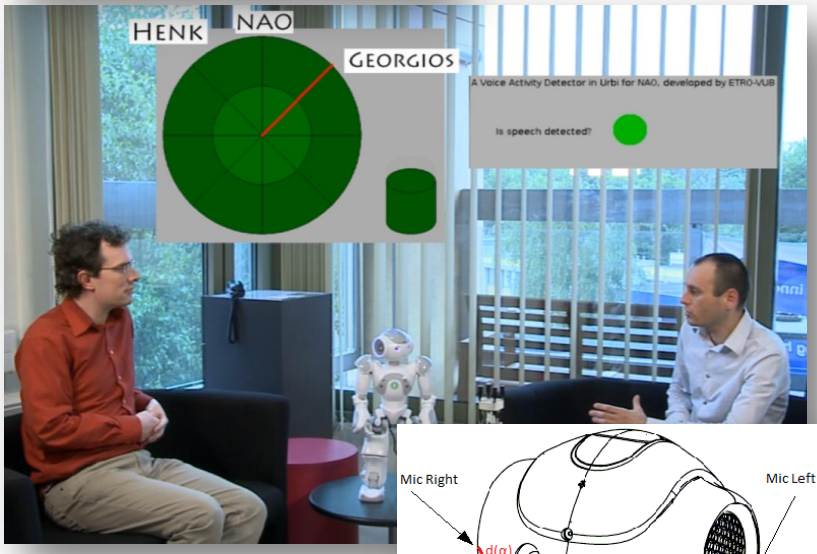
# Audio Processing

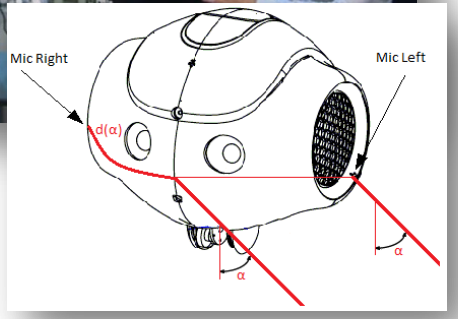
## Voice Activity Detection

- Energy based
- Noise properties are updated
- Real-time processing, integrated on Nao

## Sound Source Localisation

- Calibrated for Nao's geometry to address its shape influence
- High angular resolution
- Outperforms in-built localisation
- Improved accuracy by accounting for the microphones' frequency responses
- Performance assessed under internal and ambient noise conditions
- Speech denoising techniques fused with Sound Source Localization

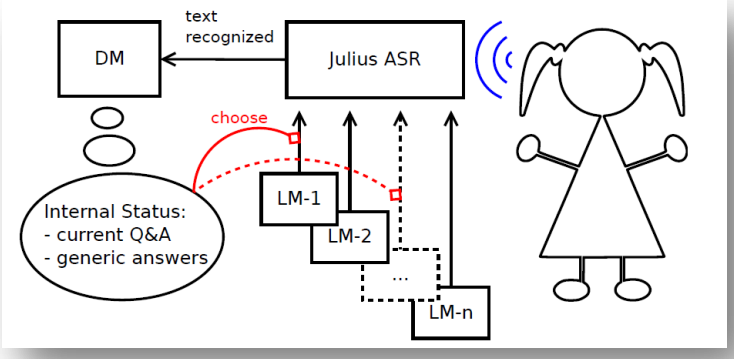


# Child automated speech recognition

- Hard problem
  - ASR is optimised for adult speech, vocal tract length variation, more pronunciation errors.
- Access to code needed: Julius (Sphinx3 abandoned).

Julius
- Core engine C library
- Multi-model recognition
- N-best / Word lattice / Confusion network



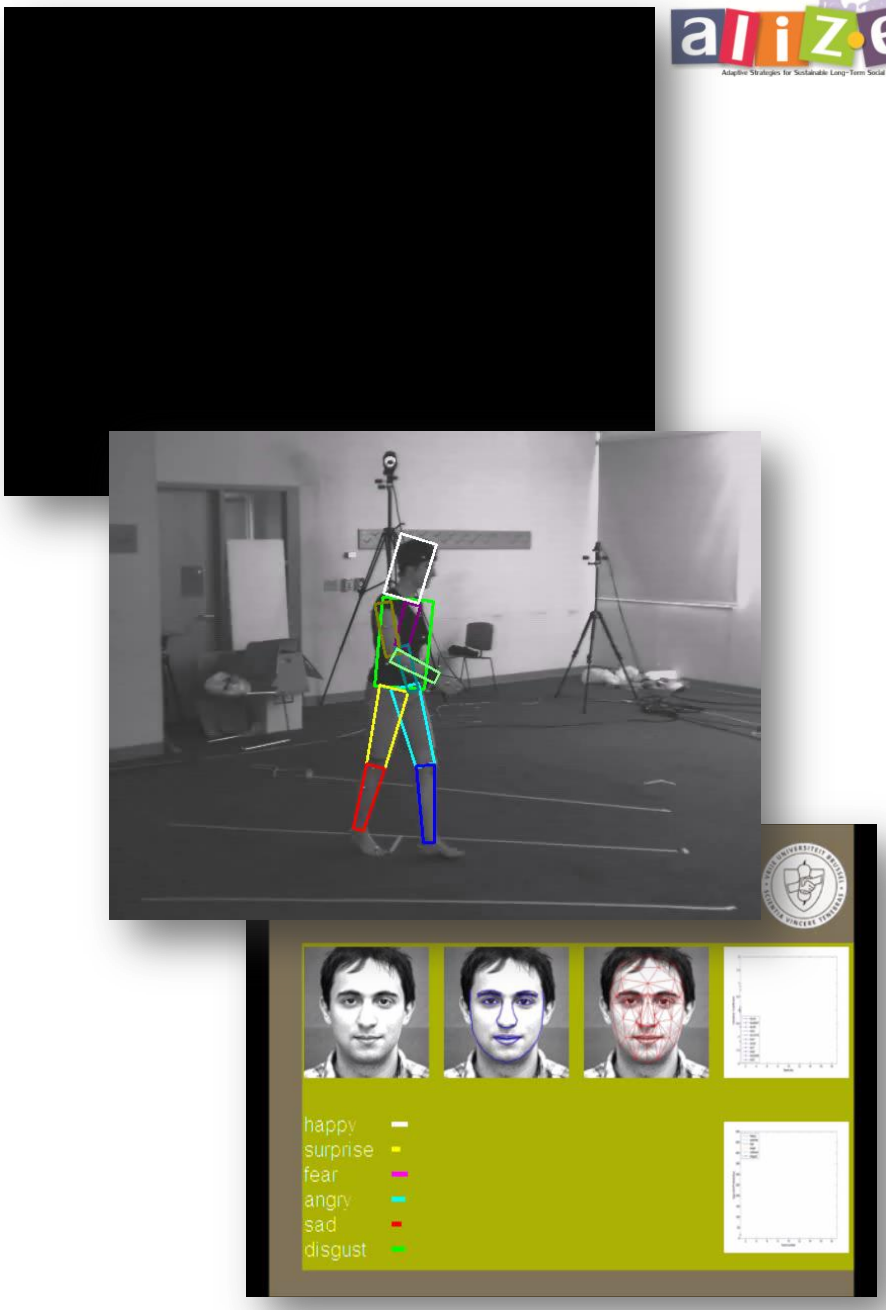

- Acoustic model
  - Trained on Italian adult voices (CLIPS Corpus) and children's voices (Childit Corpus).
  - Additional Italian children's speech (read and spontaneous)

| Type | #Snt | #Wrd | Corr | Ins | **WER** |
|---|---|---|---|---|---|
| ch | 15 | 99 | 81.8% | 12.1% | 18.2% |
| qf | 11 | 61 | 95.1% | 1.6% | 4.9% |
| sf | 10 | 38 | 34.2% | 28.9% | 65.8% |
| dg | 5 | 20 | 35.0% | 55.0% | 65.0% |
| Sum | 41 | 218 | 72.9% | 16.1% | 27.1% |

illustration of recognition performance, single child, using Sphinx3 ASR trained on ChildIt corpus. ch: Childit Corpus, qf: ALIZ-E Quiz game, sf: ALIZ-E Imitation game, dg: Connected Digits.

# Video sensing

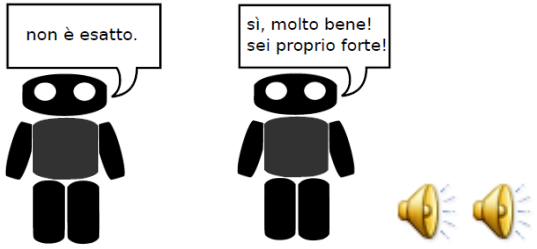
- Head direction estimation and tracking
  - Used for nod and shake recognition.
- Markerless motion capture.
  - For interaction involving physical activities.
- Gesture recognition
- Visual Emotion Recognition
- Optimised using newly available technology
  - GPU computing.
  - Kinect sensors.



# Speech/utterance production

- Text to Speech
  - http://mary.opendfki.de

    MARY Text To Speech
  - Statistical Parametric Synthesis
    - Vocal Tract Scaler to obtain a child-like voice, starting from the female voice
    - Prosody controls to add vocal expressivity




- Non-linguistic utterances
  - How do children interpret "clicks and beeps" ?

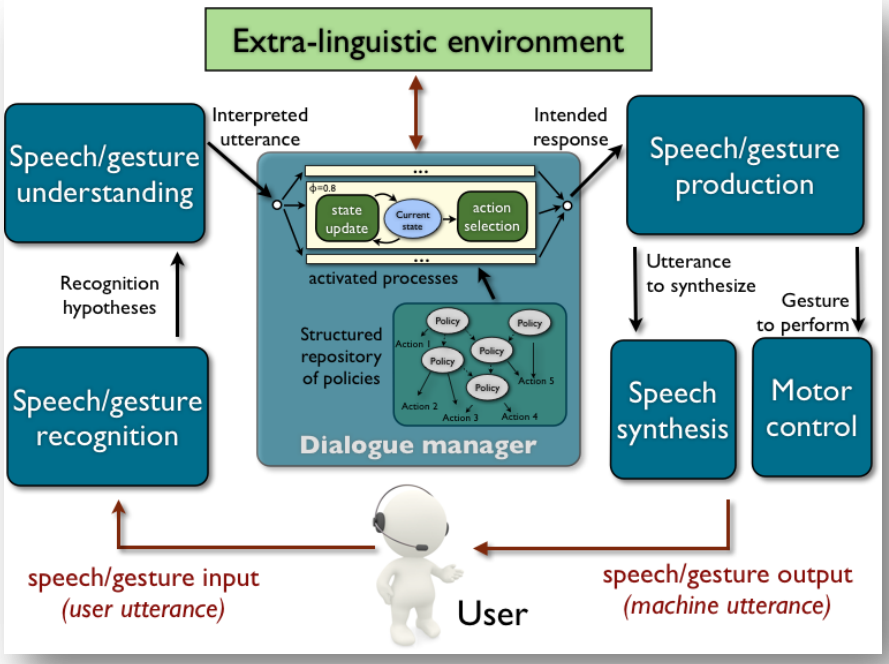
$\alpha$ Values

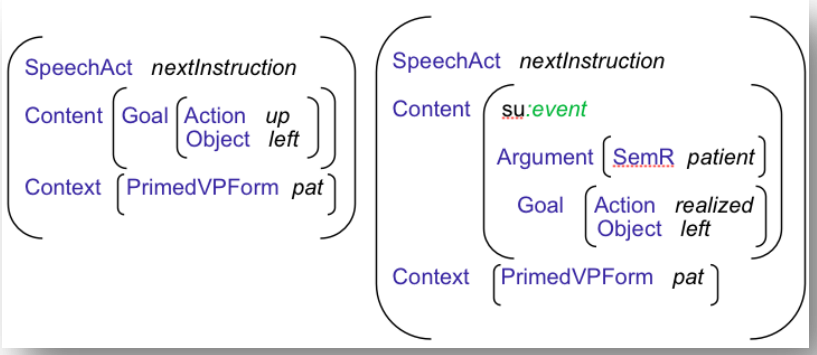
| | | Human | Animal | Tech |
|---|---|---|---|---|
| Nao | Overall | 0.590 | 0.517 | 0.469 |
| | Females | 0.659 | 0.531 | 0.478 |
| | Males | 0.618 | 0.528 | 0.462 |
| Aibo | Overall | 0.612 | 0.520 | 0.470 |
| | Females | 0.671 | 0.522 | 0.478 |
| | Males | 0.652 | 0.547 | 0.465 |

# Natural Language Interaction

- Robustness against ASR errors and incomplete, ungrammatical or out-of-grammar utterances

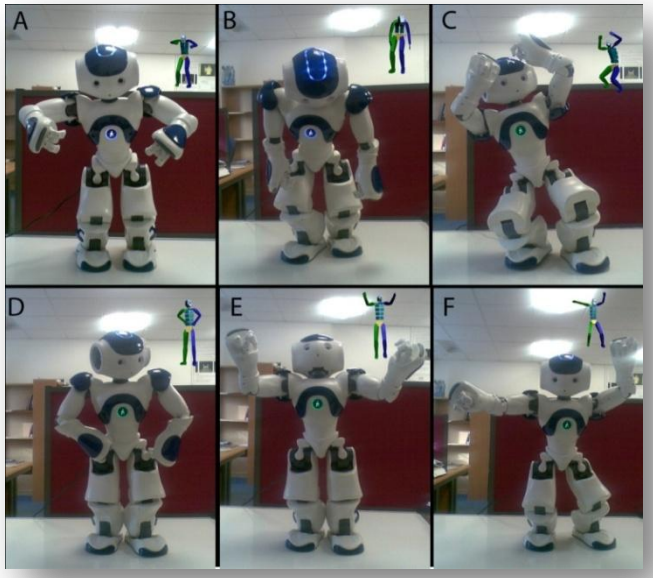
- Natural language processing
  - Using Combinatory Categorial Grammar (OpenCCG)
  - Required building of Italian grammar

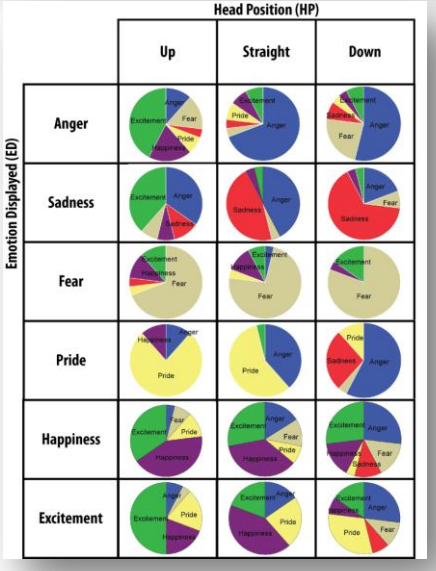


# Emotion expression

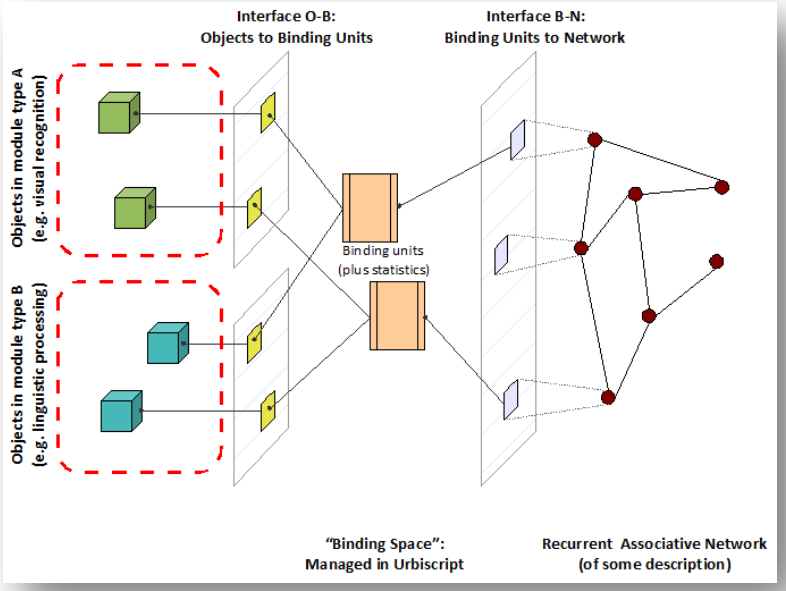
- The Nao has limited capabilities in terms of facial expressions.
- Is it possible to correctly identify emotions displayed by Nao?
- What is the effect of moving the head on the interpretation of an emotion?



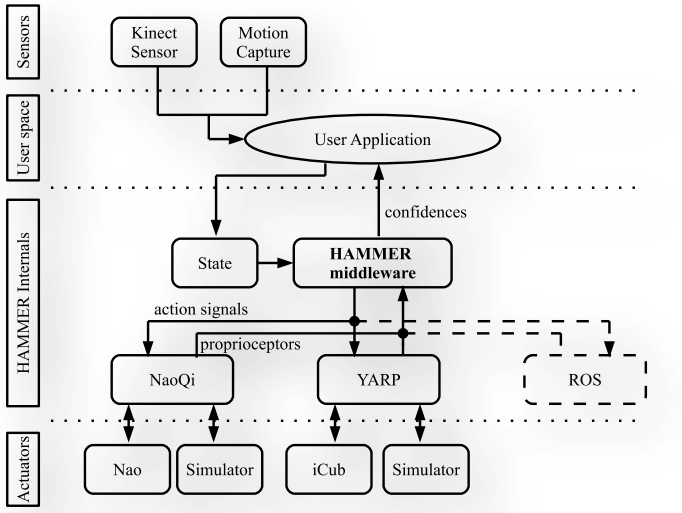
# Memory

- What is the role of adaptive memory in long-term interaction?
  - If the robot does not have a memory, how does this impact on cHRI?
- Argument for memory as a cognitive central system rather than passive data storage.
- Machine learning algorithms for on-line adaptive learning?
- Impact of emotion on creating and using memory.



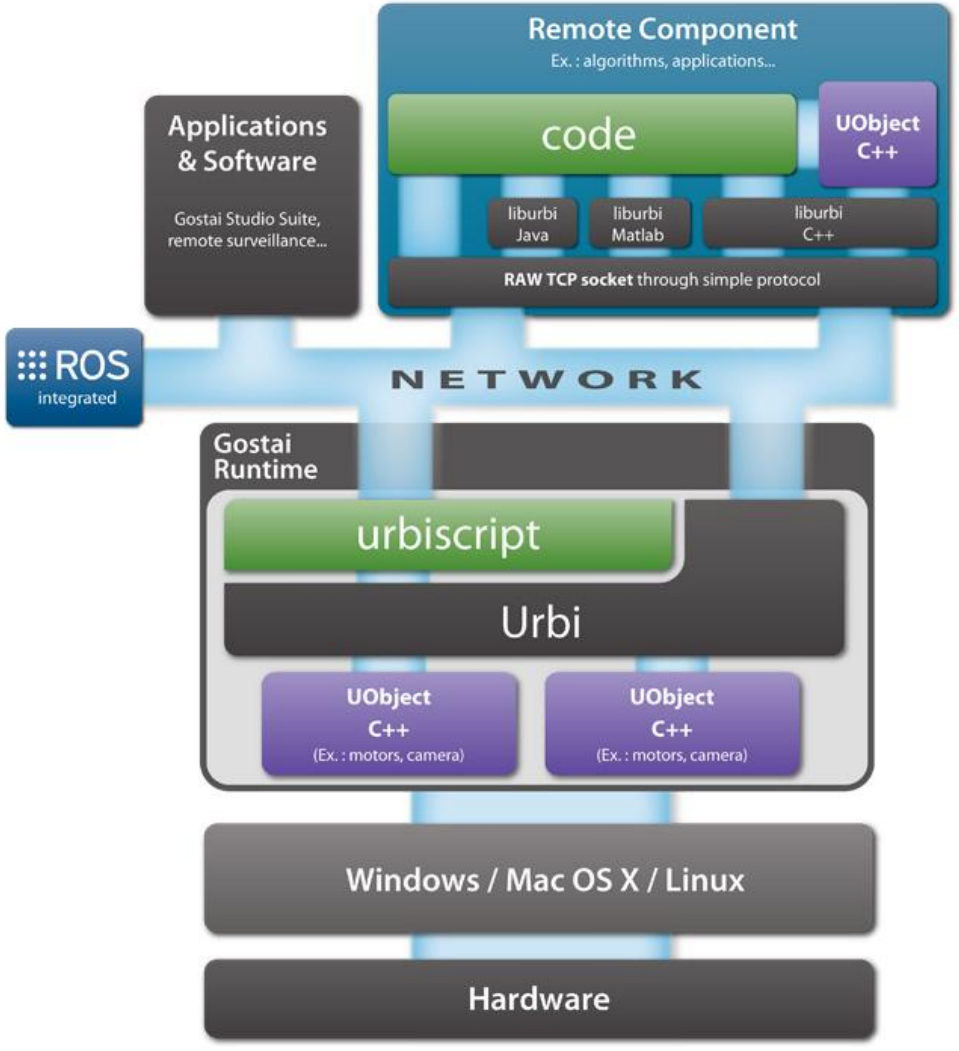
# Learning

- Recognize and predict user activity
  - Actions and intentions.
- Adapt the actions of the robot based on the user's capabilities.
  - During physical activitities the robot adapts to the user's.



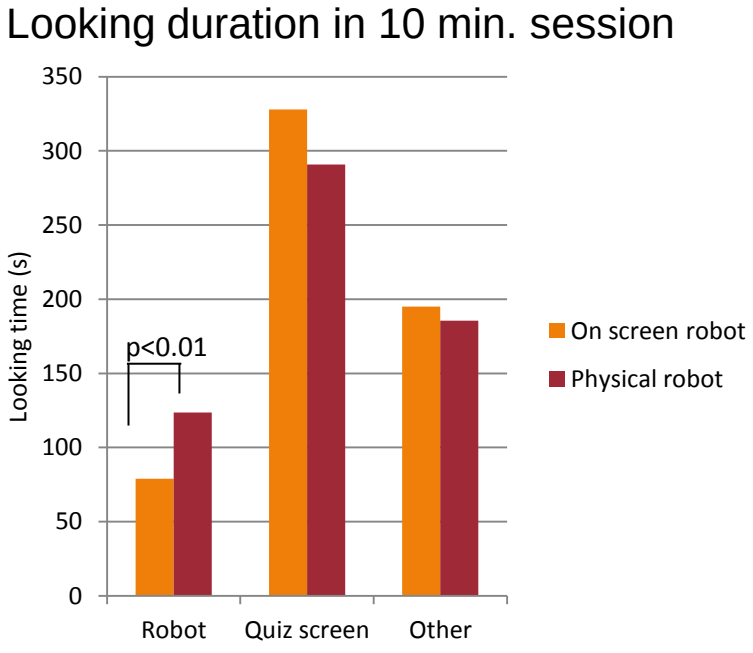

# Urbi middleware

- Technical **integration** of th embodied cognitive syster
  - Fast prototyping due to adoption of common middleware.
  - Urbi by Gostai www.gostai.com
  - Distributed component architecture, parallel and event-driven script languag

# Physical versus on-screen robot

- Study in which children are quizzed.
  - Two conditions: physical robot or on-screen robot.
  - Attention and number of fixations significantly higher.




F(1,8)=57,30

(data and experiments by TNO; Neerincx, Janssen, Looije)

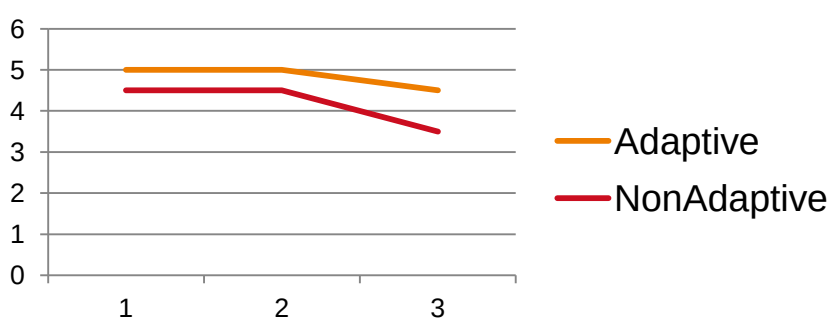
# Adaptivity

Dance

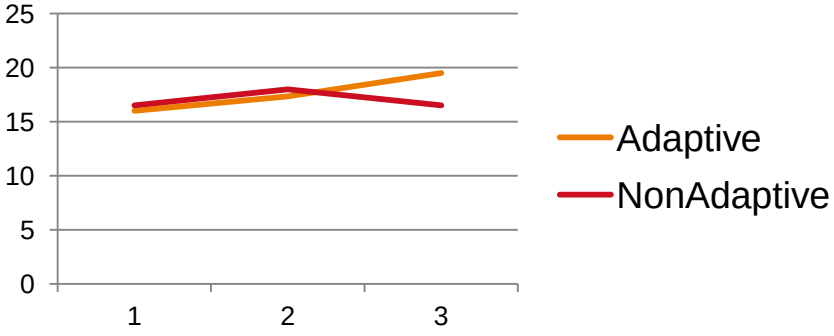
Imitation


Quiz

Does an adaptive robot impact on enjoyment and knowledge retention?




Enjoyment (N=5)




Retention (N=5)

# And so much more...

- “Gibberish” language production for multi-cultural settings.
- Perlin noise modulated by emotion to improve robot’s lifelikeness.
- Ethical and legal issues of human-robot interaction with vulnerable minors.
- Experimental protocols for child HRI.
- Cognitive architecture.
- Reservoir Computing for learning.
- User modelling
- ...


# www.aliz-e.org


alize
Adaptive Strategies for Sustainable Long-Term Social Interaction





TNO

Imperial College
London




GOSTAI
Robotics for everyone

# Hospitalised children

- How can social robots contribute to care of children (8-12) with diabetes?
- Adaptive robot for long-term interaction, capitalising on hospital stay.
  - Experiments in Dutch and Italian hospital.
- Introduction to paediatrics department .
- Learning about diabetes.
- Self-management motivation and habituation.