PACO-PLUS: Cognition through Learning of Object-Action Complexes

PACO-PLUS **IST-4-27657**

T. Asfour, R. Dillmann, F. Wörgötter

Humanoids 2005, Tsukuba, 05. Dec. 2005

Consortium

University of Karlsruhe
Germany
R. Dillmann & T. Asfour

ATR, Japan
M. Kawato & G. Cheng

Kungliga Tekniska
Högskolan,
Sweden
Eklundh

University of Edinburgh
United Kingdom
M. Steedman

University of
Göttingen
Germany
F. Wörgötter

Leiden University
Netherlands
B. Hommel

Aalborg University
Denmark
N. Krüger & V. Krüger

Consejo Superior de
Investigaciones
Científicas, Spain
C. Torras & J. Cetto

Jozef Stefan Institute
Slovenia
A. Ude

Goal

- PACO-PLUS aims at the design of cognitive robot system capable of

  - developing perceptual, behavioural and cognitive categories in a measurable way and

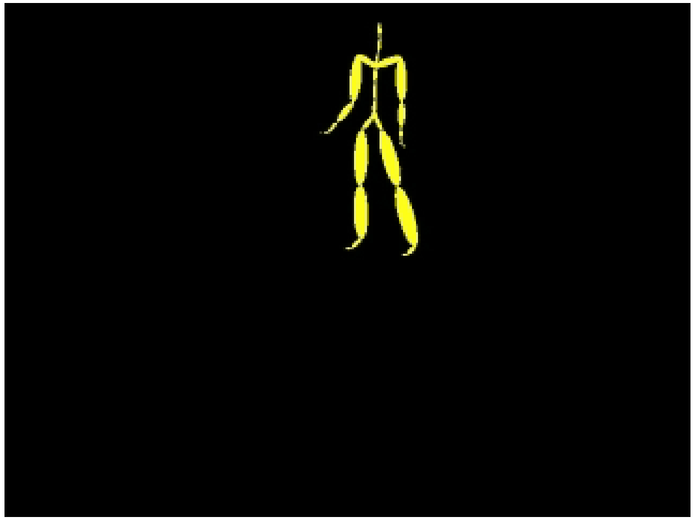
  - communicating and sharing these with humans and other artificial agents

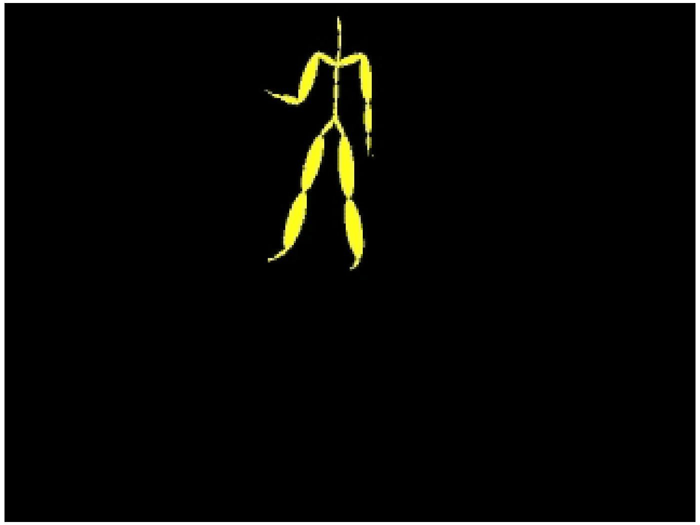
Guiding Principles

- Objects and Actions are inseparably intertwined
  → ***Object-Action Complexes (OACs)***
- The basic assumptions:
  - Cognition is based on recurrent processes involving nested feedback loops operating on object-action complexes
  - Action defines the meaning of an Object
  - and Objects suggest Actions – Affordance Principle

Motivation

- Visually based object recognition fails
- Visual information is too sparse
- Activity involving the object decreases the uncertainty about the object's nature considerably!





CMU Graphics Lab Motion Capture Database
http://mocap.cs.cmu.edu/

Main Project Objectives

- Sensor Representation: Invariant, multi-sensory *representations of objects* through actions performed on them
- Action Representation: Probabilistic, hierarchical representation for actions on objects
- Quantification: Theoretical framework for the design of a conjoint measure for learning- (and action-) success
- Decision making and Planning: Planning system in a *continuous* perception-action space
- Novelty and Creativity: New actions by means of extrapolation within the OAC space along constraints given by motor capabilities and Gestalt statistics
- Language and Communication: Appropriate grammar fragments for representative languages can be machine-learned from exposure to strings paired with plan and action representations
- Realization: Technical cognitive system with basic oculomotor behaviours and sensory-motor primitives

Theoretical Framework

Theory of Event Coding (TEC)
Hommel et al. 2001 BBS

Shared code for perception and action exists.

OACs: Technical version of TEC

Computational Theory of Mind
Putnam, 1961; Dennett 1984, Porr & Wörgötter 2005

Solving the Frame Problem by *embodied, situated, adaptive* agents reacting to disturbances in the environment in a *proactive, anticipatory* way.

Richly integrated robotics platform(s) process a continuum of multiple stimuli while producing meaningful behaviour

Language
Quine 1960, Siskind 1996, Villavicencio 2002 Steedman 2002

Language learning is easy. Developing the concepts that underlie it is hard.

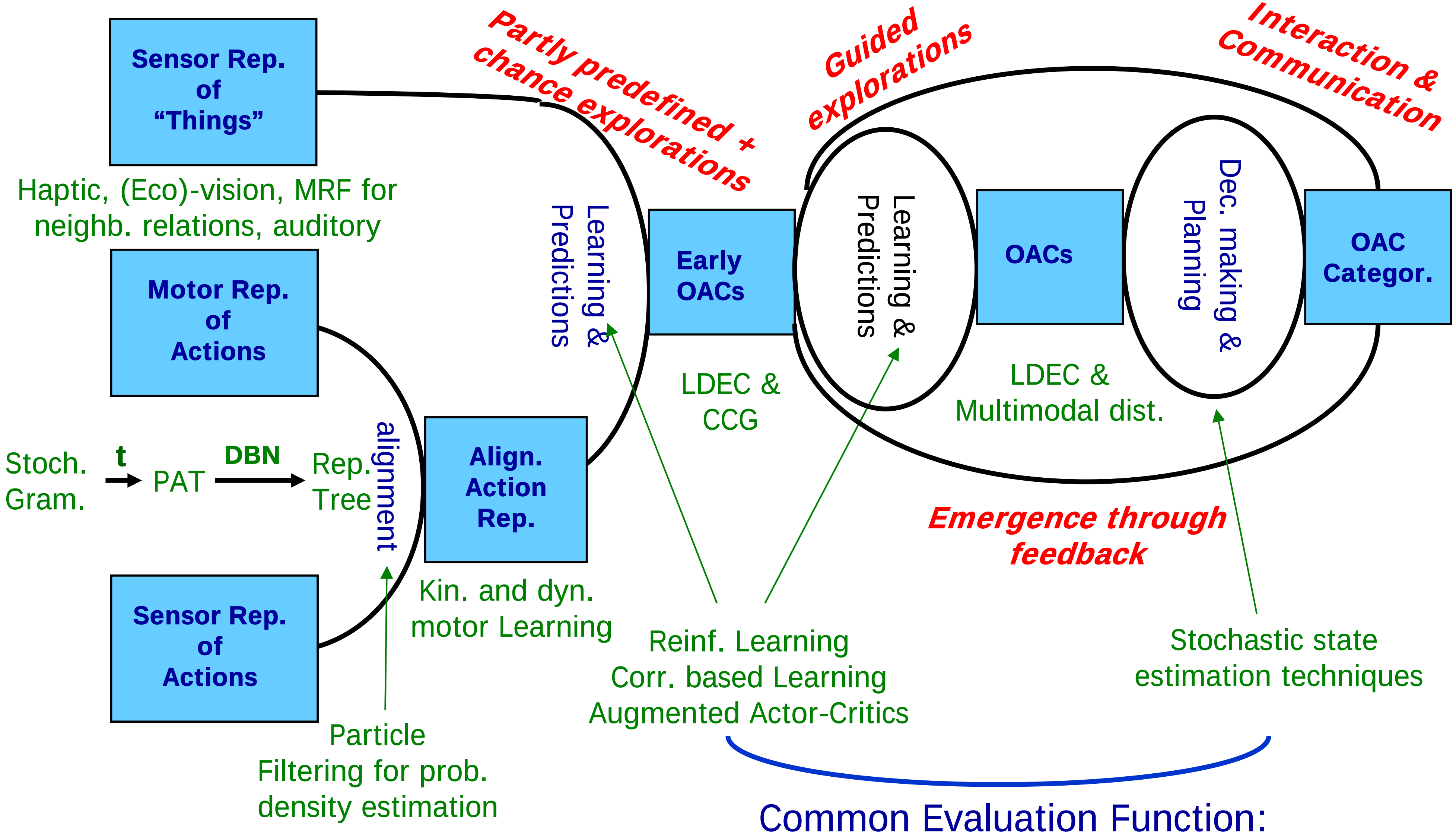
OACs will constitute shared and grounded symbols as the basis of language

Theory of Learning
Camerer 2003 Torras 2002, Porr & Wörgötter 2003

Convergent learning arises from minimization of contingencies together with the maximization of return.

Conjoint evaluation function for success of the system

Algorithmic Framework & Architecture



Common Evaluation Function:
for strategy selection: reward max. vs. contingency min.

Innate Sensorimotor Primitives (Atomic Motor Acts)

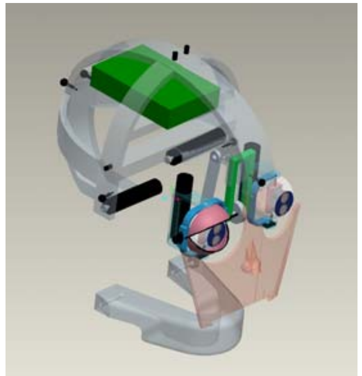
- Closed-loop reaching
- Open-loop reaching
  - Classic (forward-inverse kinematics)
  - Learning (RBF networks, self-organizing maps)
- Foveation control (getting the object into the fovea based on information from peripheral views)
- Smooth pursuit
- Saccades (relating information from before and after the saccade)
- Grasping (Power and Precision grasps)

Beyond Primitive Behaviors

- Imitation of human actions (on objects)
  - Learning what to imitate → which aspect of an object is important
  - Learning how to imitate → observation of human actions on objects
  - Imitation as a process acting on OACs
  - Adaptation of kinematics and dynamics
- Coaching
  - OAC based grammar as the basis to design a robot coaching system in the context of an integrated action-language system.
  - The role of the coach is to specify goals and priorities, to focus attention of the learning agent, and to give feedback such as correction strategies, evaluation and success definition.
  - Communication can take place through direct demonstration and / or verbal instructions.

Humanoid Head with Foveated Vision

- Learning of OACs:
  - identify regions that potentially contain objects of interest (visual search that can benefit from a wide field of view
  - analyze these regions to build higher-level representations (narrower field of view resulting in higher-resolution images)
- Humanoid head with foveated vision able to move at speeds and accelerations similar to human's eyes.
- Foveated vision will be realized using two cameras per eye
  - Related systems: DB at ATR (Japan), Cog at MIT (USA), Infanoid at CRL (Japan).

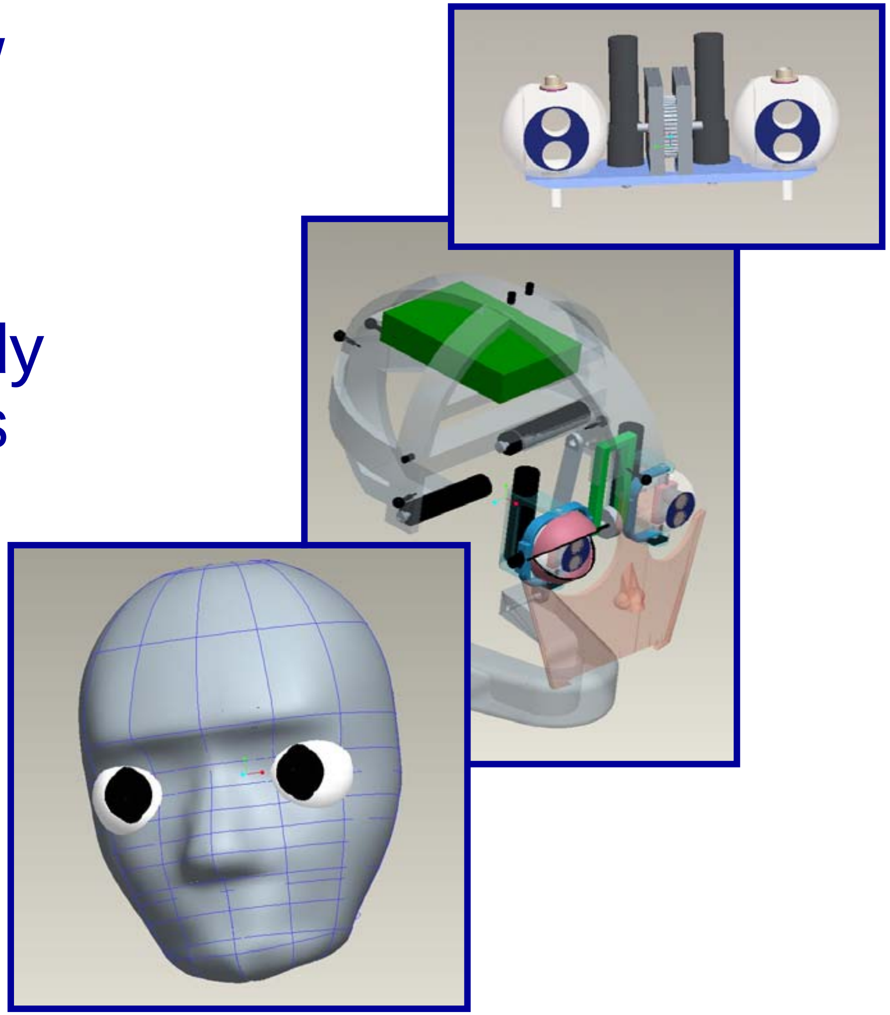


Humanoid Head with Foveated Vision (2)

- The head will be designed as part of an humanoid platform that will allow for the integration of motor control and perception

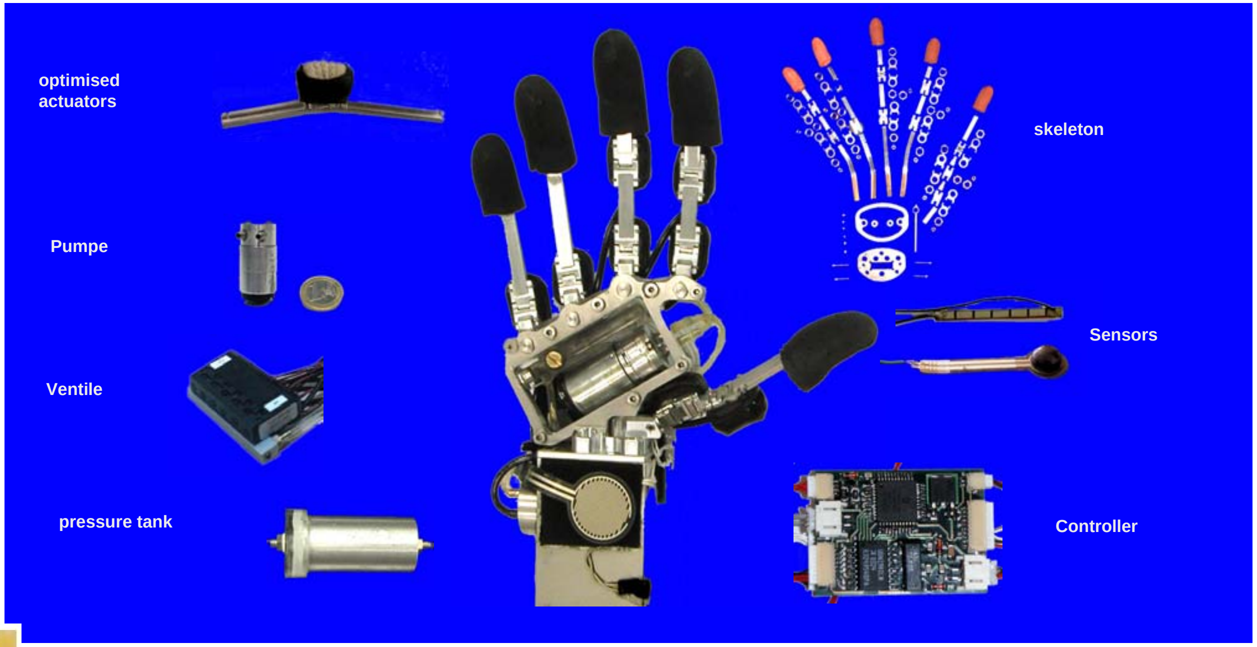
  ➔ Explorative head, hand, and body movements for learning of OACs

- Head Design
  - 2 cameras per eye (3 DOFs),
  - 2 eye lids,
  - color cameras (Dragonfly)
  - 6 Mikrophons
  - Jaw (1 DOF)



Hand with artificial skin (data glove)

- 8 DOF (up to 18)
- Weight: 395 g
- Force:
  - Finger tip: 7 N
  - Holding force: 35 N
- RS 232/Bluetooth



- Dataglove with 8 or 14 tactile sensors
- Bluetooth or serial Interface

Scenarios

- Generation of OACs for things (filling of cups, tool handling)
  - perceive and understand things (e.g. a bottle, cups, pens), i.e. know the actions that can be performed on it and execute these actions if possible.
- Generation of OACs for human movements
  - perceive and understand the postures and movements of a human
  - Imitation of the human movement
  - Configuration of the arms and legs of a human-like doll according to the posture of the observed human

Perceiving Human Motion

- Capture 3d human motion based on the image input from the cameras of the robot's head only
- Approach:
  - Particle Filter framework
  - Localization of hands and head using color segmentation and stereo triangulation
  - Fusion of 3d positions and edge information
- Features
  - Automatic Initialization
  - Close to real-time: 15 fps on a 3 GHz CPU, images captured at 25 Hz
  - Smooth tracking of real 3d motion

[Azad et al. 2004]

PACO PLUS: Key Research Activities

Representation of “things“ and actions

How OACs arise?

How OACs evolve?

Decision Making and Planning

Memorization and Learning

Language, Communication and Interaction

Cognitive Robotics System: Integration and Architecture