Action Understanding and Generation

Erhan Oztop
Ozyegin University, Turkey & ATR, Japan

Ozyegin University, Cekmekoy , Istanbul

Advanced Telecommunication Institute, Kyoto

TGMIS
Turkish German Multimodal Interaction Summit
Istanbul, Turkey, 11 November 2014

Research Areas

Goals:

Studying how humans and other biological systems process information and solve problems from a computational viewpoint

Developing paradigms and techniques for science and engineering based on this study

Current Focus

Robotics
- **Human in the loop robot skill synthesis**
- Developmental Robotics
- **Action Recognition/Understanding**

Neuroscience
- **Mirror Neurons**
- Metal State Inference
- Emotion

Cognition and Behavior
- Modeling Human Learning
- Motor Adaptation in full body motion

Mathematical Analysis
- Boolean Functions
- Extremal Combinatorics
- (Inverse)Reinforcement Learning

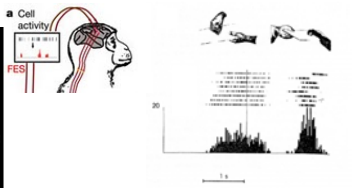
2/47

PART I. Action Production & Perception

Mirror Neurons

Multimodal neurons that respond when a monkey performs an action (e.g. grasping)
AND
When they observe the execution of a similar action performed by others

First found in ventral premotor cortex (F5); later in dorsal premotor cortex (F2) and even in the primary motor cortex (F1) and in the parietal cortex cortex (PF)



Why Mirror Neurons are so 'important'?

- Perception and Action systems are tightly coupled
- Humans are likely to have a Mirror Neuron System that may share the same properties with monkey mirror neurons
- May give clues about Human Cognition
- May give clues about Language Evolution
- For engineering and robotics:
  - Action recognition and generation share common mechanisms. Utilize existing resources

How much is known?

More than two decades has past after the discovery of mirror neurons but still…

Functions: Unknown

Experimental Data: Mostly not quantitative

Developmental Course: Mostly not explored

Computational Models: Not satisfactory

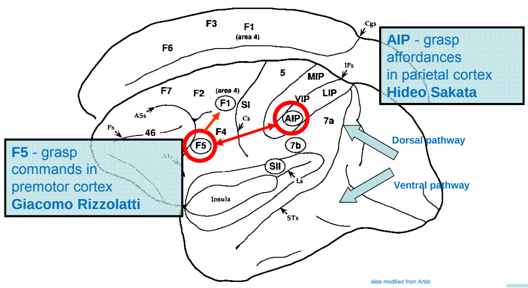
Myths: Many

Mirror Neuron System Model

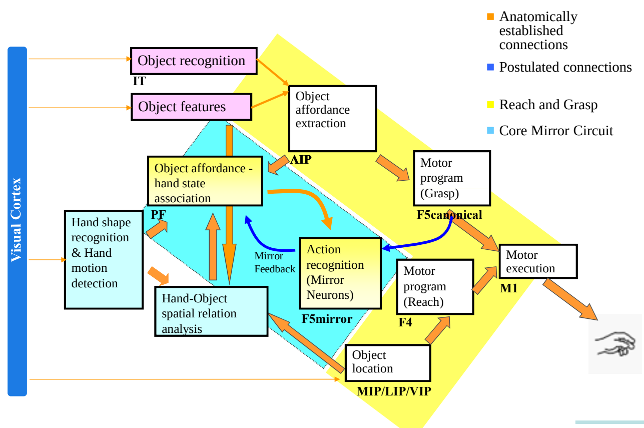
Key Ingredients of MNS model (Oztop & Arbib 2002)

Self-observation mediates formation of mirror neurons: Mirror neurons develop by associating the executed motor (grasp) program with the visual stimuli generated

Object centered representation of action parameters



Visual Analysis Schema



Object centered representation of the hand

Hand State captures the hand and its relation to the target object.

$H(t) = (d(t), v(t), a(t), o_1(t), o_2(t), o_3(t), o_4(t))$

where

d(t): distance to target

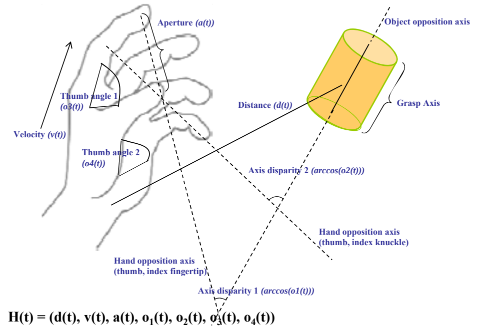
v(t): tangential velocity

a(t): Aperture

$o_1(t)$: Angle between the object axis and the (*index finger tip – thumb tip*) vector

$o_2(t)$: Angle between the object axis and the (*index finger knuckle – thumb tip*) vector

$o_3(t)$, $o_4(t)$: The two angles defining how close the thumb is to the hand as measured relative to the side of the hand and to the inner surface of the palm.



$H(t) = (d(t), v(t), a(t), o_1(t), o_2(t), o_3(t), o_4(t))$

Note that the whole history of H(t) during a grasp is required to represent the grasp.

Summary & Current Focus

The key elements of the model were:

Mirror Neurons are formed by the association of the **neural code for self-executed** grasps with the neural code for the visual stimuli generated

Hand in action is encoded using an **object centered representation** that allows generalization to others' hands.

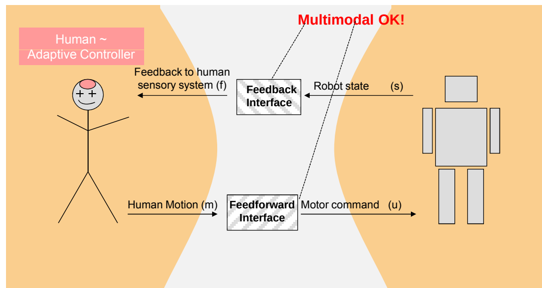
But did not explain why the learning take place (i.e. what are the mirror neurons good for?)

Related papers: Oztop, Wolpert, Kawato (2005); Oztop, Kawato, Arbib (2006); Oztop, Kawato, Arbib (2013)

Current Focus and Funding: *Greek-Turkish Bilateral Project with Raos Vassilis, FORTH) (TUBITAK-GSRT "Neurophysiology and computational modeling of action-observation")* is underway to find the role of mirror neurons in representing control related parameters of observed and executed actions

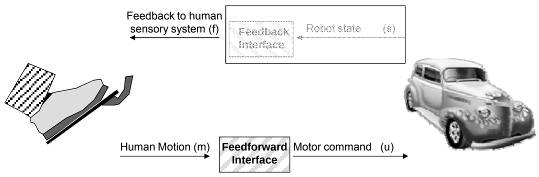
PART II. Robot skill synthesis via human motor learning

- Use human sensorimotor learning ability to obtain robot behaviors
- Include the human in the control loop
- May ask human to do extensive training
- Utilize the human brain as the adaptive controller



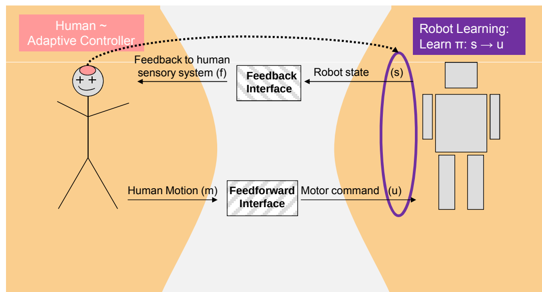
Sensorimotor learning

- Sensorimotor learning is fundamental for adaptive and intelligent behavior
- Driving a car
- Using a pair of chopsticks
- Using a computer mouse
- …
- …
- …



Behavior synthesis for autonomy

For autonomous operation, the key issue is transferring the **control policy** learnt by human to the robot

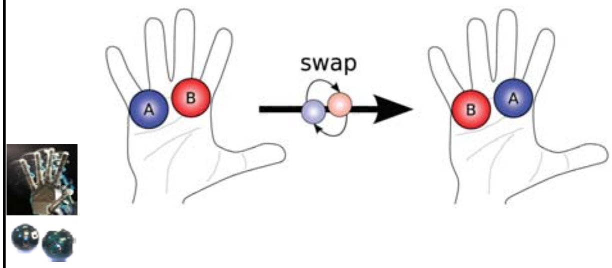


Why should this paradigm work?

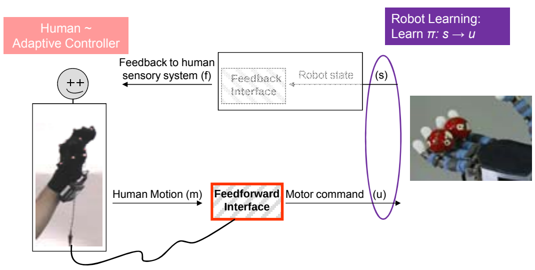
The ability of the brain to learn novel control tasks by forming internal models. The robot can simply be considered as another tool (e.g. as in snowboarding, driving, using chopsticks)

The flexibility of the body schema; extensive training on the human side should modify the body schema so that the robot can be controlled naturally

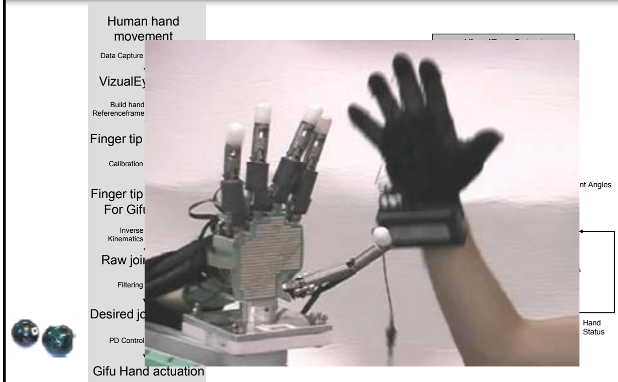
First example: ball swapping



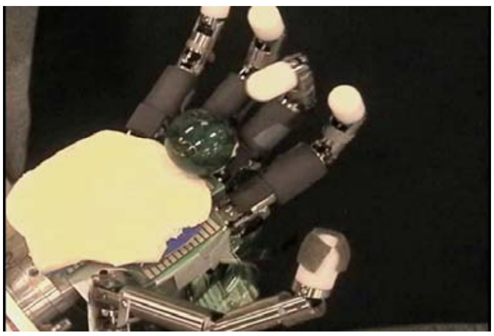
Ball swapping interface



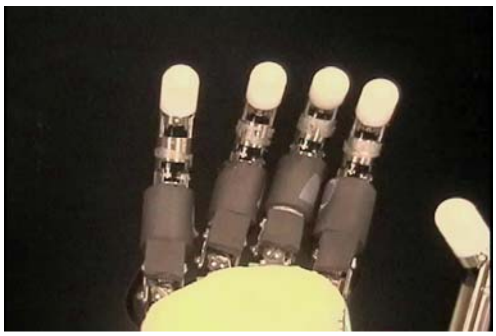
Human control interface



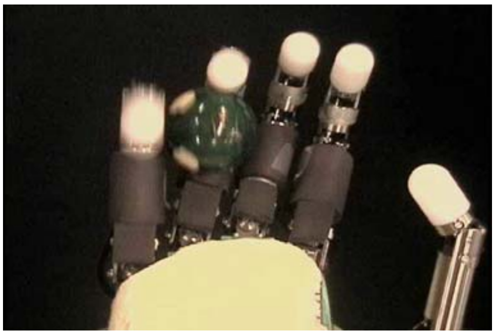
Human sensorimotor learning…



Finally human learns to swap balls



Autonomous ball swapping

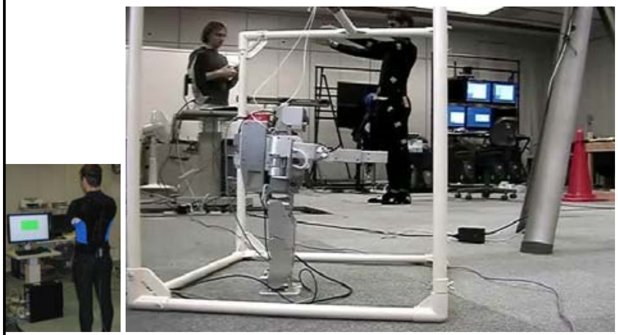


Oztop, Lin, Kawato, Cheng (ICRA 2007)

Open loop control
u=π(time)

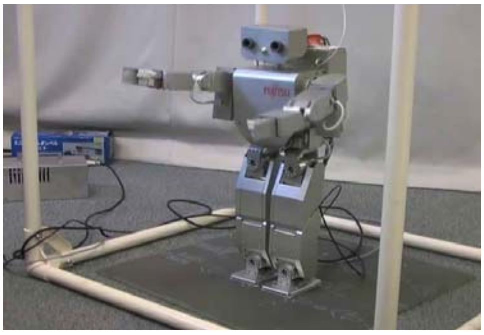
26

Reaching without falling



Babic J, Hale J, Oztop E, (Adaptive Behaviour 2011)

Autonomous stable trajectory tracking



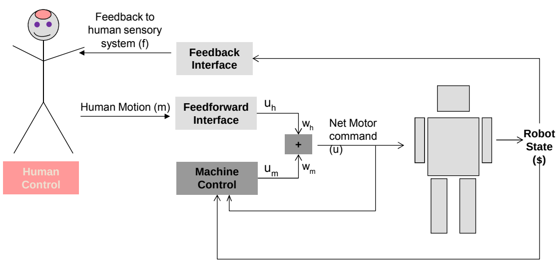
32

Summary & Current focus

Obtaining robot skills via human sensorimotor learning works
- Help built intuitive smart prosthetics
- Can be used to probe human motor control
- Shed light on mechanisms of internal models, agency and body image

Current Focus & Funding:
- *Supported by Marie Curie FP7 Converge (2012-2016) project*
- More intuitive state feedback to the human
- Simultaneous human and machine control
- Transferring force based skills (w/J. Babic, JSI, Slovenia)
- Other tasks, e.g. walking

Simultanous learning



34

Simultaneous Learning of "pole swing-up and pole balance"

✓ Online robot learning with simultaneous human and machine control
✓ State dependent dynamic control sharing

Zamani, Oztop (submitted)

37

PART III. Dynamic Motor Primitives (DMP) for behavior understanding

Work in collaboration with ATR, Japan supported by Ministry of Internal Affairs and Communications (MIC) of Japan. http://www.cns.atr.jp/dbi/en/

The project aims to establish dynamic brain imaging techniques based on non-invasive brain activity measurement for daily life brain-machine interface (BMI) using real-time brain decoding via statistics and large-scale computing

Contributing Organizations:
ATR, NTT, Keio University, Shimadzu Corp. Sekisui House, Ltd.

DMP based segmentation is proposed for automatic tagging/labeling of behaviors for automatic BMI

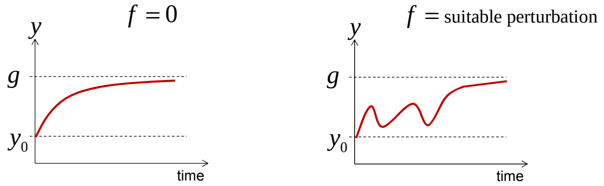
38

Dynamic Motor Primitives

From (Ijspeert et al. 2013):
'The basic idea of our approach is to use an analytically well-understood dynamical system with convenient stability properties and **modulate it** with nonlinear terms such that it achieves a desired attractor behavior (Ijspeert et al., 2003)'

As one of the simplest possible systems, a damped spring model is chosen

$$\tau\ddot{y} = K(g-y) - D\dot{y} + f$$



Dynamic Motor Primitives: learning

When we observe an action, we record goal (g), duration (~ 1/τ) and the position (y) at each time step. Since we assume $\tau\ddot{y} = K(g-y) - D\dot{y} + f$ must hold at each time step we have the following consistency condition

$$\tau\ddot{y}_{obsv,i} - K(g - y_{obsv,i}) + D\dot{y}_{obsv,i} = f(w, t_i)$$

Taking f as a linear function of nonlinear basis functions we have

$$\underbrace{\tau\ddot{y}_{obsv,i} - K(g - y_{obsv,i}) + D\dot{y}_{obsv,i}}_{x_i} = \sum_{j=1,m} w_j \phi(t_i)$$

Which can be written as a linear regression problem:

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \phi_1(t_1) & \phi_2(t_1) & \cdots & \phi_m(t_1) \\ \phi_1(t_2) & \phi_2(t_2) & & \phi_m(t_2) \\ & & & \\ \phi_1(t_n) & \phi_2(t_n) & & \phi_m(t_n) \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{bmatrix}$$

$$\mathbf{X} = \mathbf{\Omega w} \qquad (\Rightarrow \mathbf{w} = \mathbf{\Omega}^{\dagger}\mathbf{X})$$

Application of Dynamic Motor Primitives

Usually used for

- Representing robot trajectories
- Robotic imitation
- Robot learning in policy space

- Gives a compact representation
- Can be made time independent
- Can be scaled
- Can be easily directed to different goals

Can also be used for recognition

- initial work due to Ijspeert et al. but no extensive work

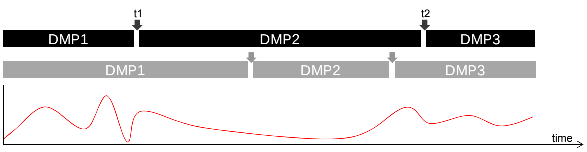
Ijspeert, Nakanishi, Schaal (ICRA 2003)

New IDEA: Automatic segmentation with DMPs

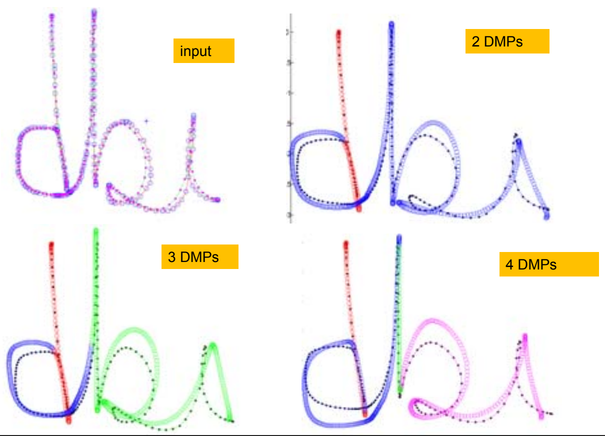
Concept:
A long trajectory **x**, should be better represented with sequentially order ed multiple DMPs that take intro account the complexity of the trajectory they represent

Assumption:
The number of DMPs (e.g. 3) and the total number of number of degrees of freedom (N, # weight parameters) are fixed and given.

Approach:
-Divide the trajectory into equal pieces and determine # of weights for each DMP
OR
- Give equal number of weight parameters for each DMP; but determine their span (i.e. t1 t2 below)



Preliminary results with 2D trajectories



Using Kinect Data from the Smart House for Automatic Segmentation

Preliminary results from kinect skeleton data…

Computer Mouse data

2 coordinates
2 composite DMPs

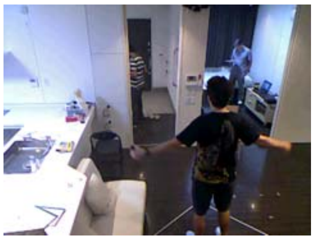
Kinect data:

20 x 3 (20x2) coordinates in Kinect
60 (40 composite DMPs)

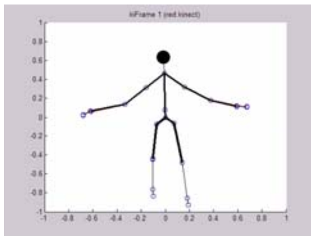
But essentially the same problem

Kinect data representation with a DMP

Original video from Kinect Camera



DMP Reproduction and Kinect Skeleton Data



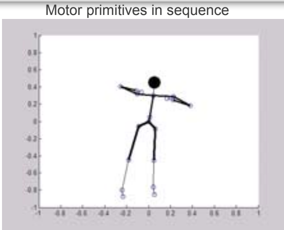
Two DMPs per node, 32 bases functions per DMP; there are 20 nodes in a Kinect Skeleton Data
Therefore we use 2x32x20 = 1280 parameters to specify this movement
Note that we could easily **do dimensionality reduction**!

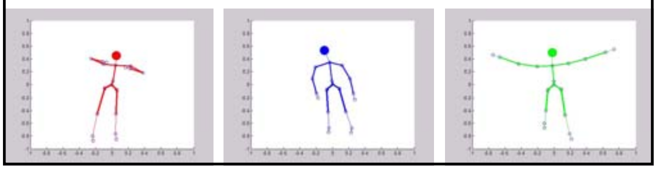
3-4 dimension would suffice ~ 200 parameters

Automatic primitive detection: Example I

Original video from Kinect Camera

Motor primitives in sequence

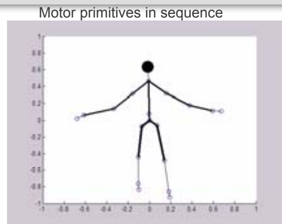


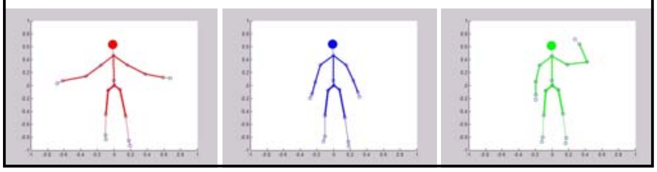


Automatic primitive detection: Example II

Original video from Kinect Camera

Motor primitives in sequence



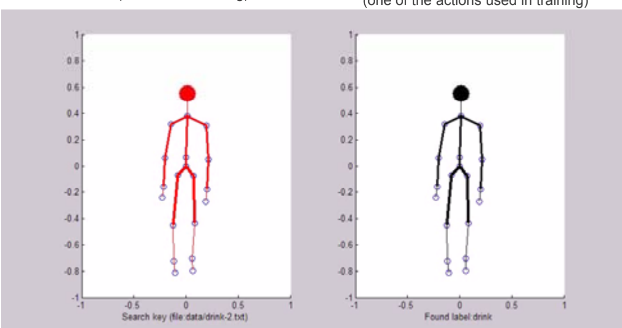


Recognition with Kinect data is also possible

Original video from Kinect Camera

Test action
(not used in training)

Recalled action
(one of the actions used in training)



Summary & Current focus

DMP based recognition is well suited for human action recognition

DMP based segmentation produced encouraging results but more work is needed

Autonomous segmentation has the potential to create fully autonomous daily-life BMI.

Current Focus & Funding:

Funded by a contract in H23 with the Ministry of Internal Affairs and Communications, Japan, entitled 'Novel and innovative R&D making use of brain structures'.

Improving DMP based segmentation

Evaluations for DMP based recognition via comparisons with HMM based recognition

50

THANK YOU FOR YOUR ATTENTION!

51