Robot sensing and manipulation

Lorenzo Natale

IIT School – Robotics week
January 22-25, 2008
Morego (GE)

Acknowledgements

- In robotics collaboration helps…
- The work I am going to present was done together with other people I have been lucky to collaborate with:

Giulio Sandini LIRA-Lab and IIT

Giorgio Metta LIRA-Lab, and IIT

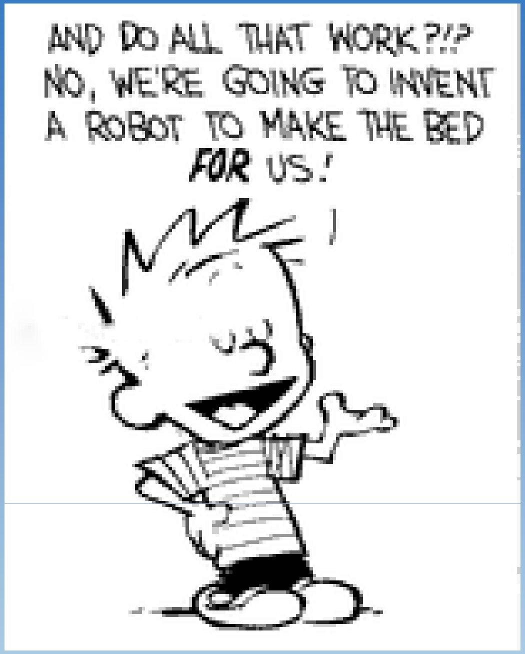
Francesco Orabona, LIRA-Lab

Paul Fitzpatrick MIT and IIT

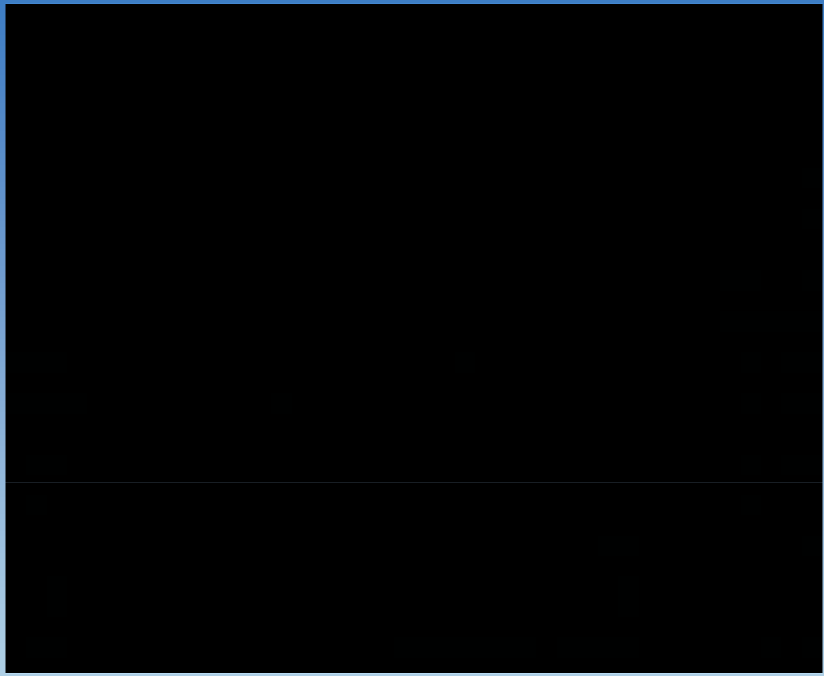
Francesco Nori, LIRA-Lab and IIT

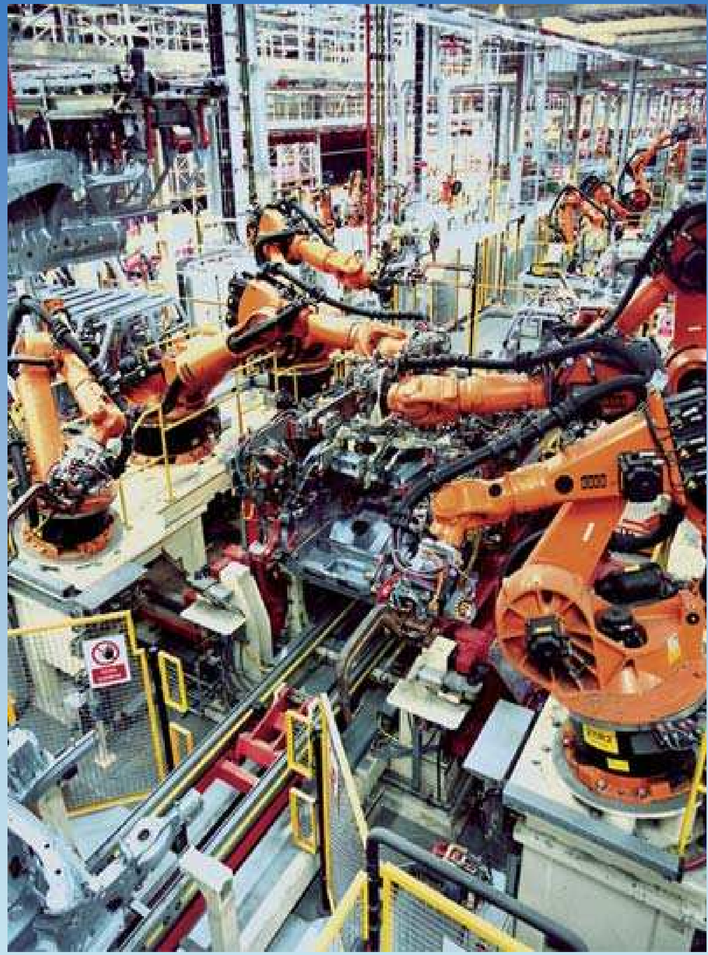
Eduardo Torres-Jara, MIT

Robotics, the dream?



Robotics, now





Industrial robots, assembly line:
Fast, reliable, precise, but highly specific

Robotics, now

Roomba, the robot cleaner, more general but… it does not do much!

…what is missing in current robots?

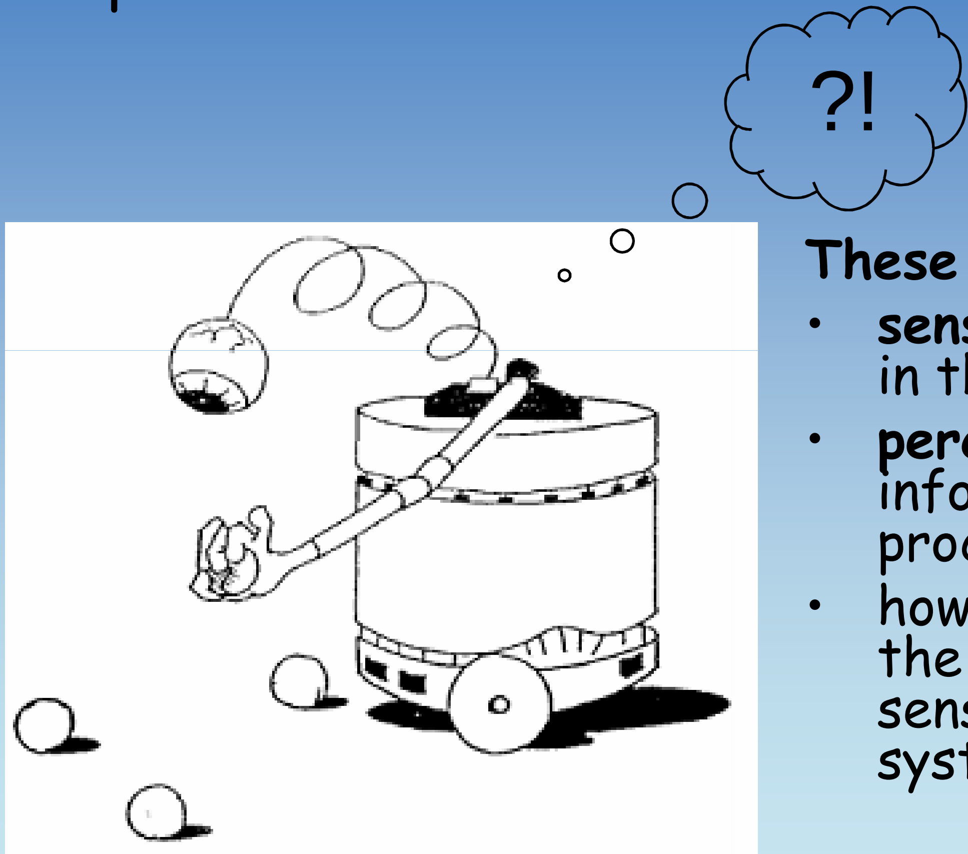
… the challenge for robotics is to realize robots that can deal with *uncertainty*

An extreme example: a robot on a distant planet

picture from: (R. Pfeifer 1996)

- What weapons can we provide the robot to help survive?
- Sensors might be useful: vision, sound, encoders, force, touch, smell, temperature, IR, laser, sonar…

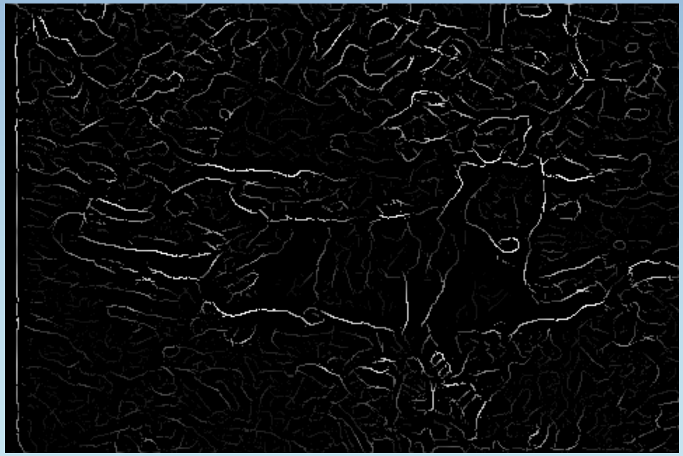
...but it would not be fair if we did not provide some hints about how to *interpret* this information and how to use it to perform the task...



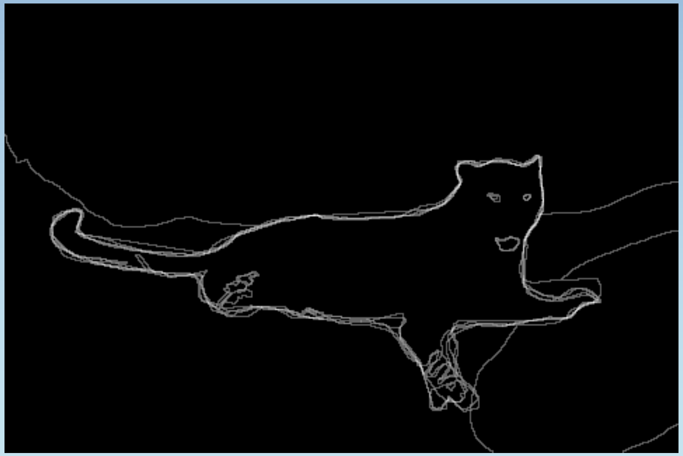
These "hints" concern:

- **sensation**: how to *detect* a stimulus in the environment
- **perception**: how to *interpret* the information that is gathered (and processed) by the senses
- how to build "*internal models*" of the external world, based on the sensation provided by the sensory system

Perception is difficult



What a machine sees



What we see

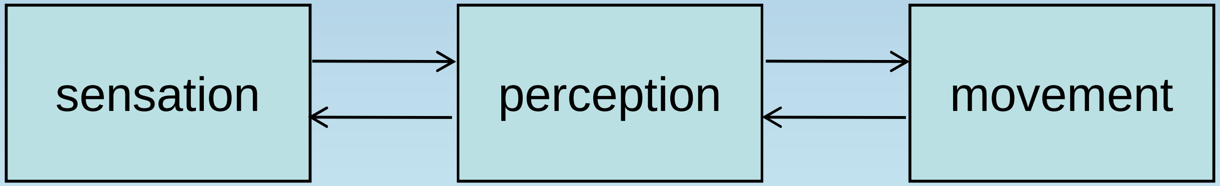
Adapted from: P. Fitzpatrick, et al. 2008, original work from Martin et al. 2004

Some frustrations

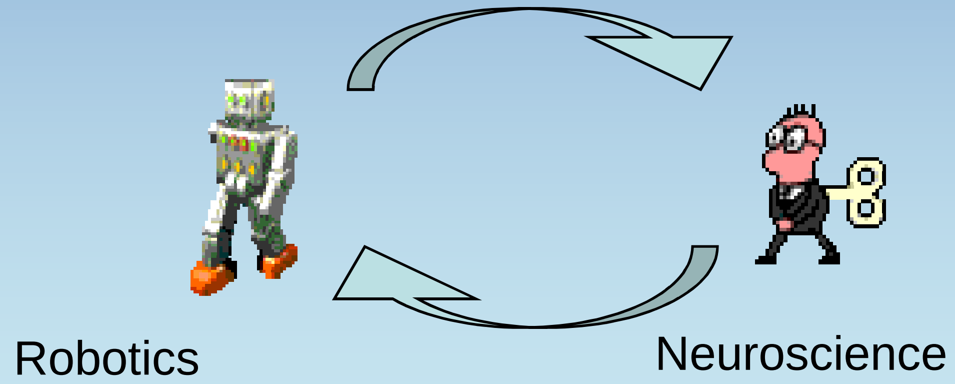
- The amount of information is overwhelming
- Realtime: sensory information is useful if up to date
- Accessibility: information is often hidden
- Reference frame: different sensors are attached to different (moving!) body parts (eyes, head, hand…)
- Noise, variability…

But (advantages)…

- Redundancy: different sensory modalities can provide information about the same events or objects in the environment
- Actions: the robot can be active and perform actions to help perception



- Is there an answer to these questions? Is it possible to build artificial systems that successfully interact with such a challenging environment?
- Biological systems do, so a good starting point could be to learn from them…
- On the other hand, artificial system could be helpful as test platform on which to validate computational models of the brain



So let's talk a little bit about biological systems

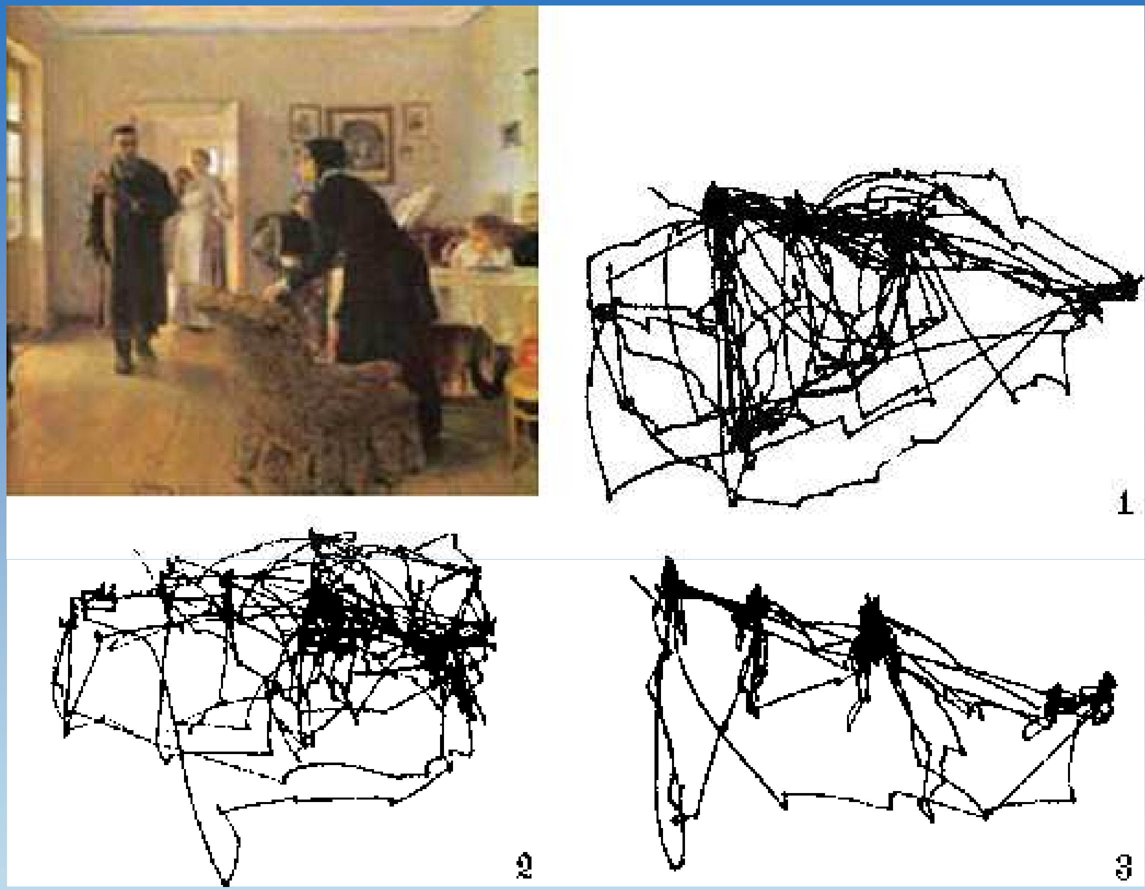
- The brain interprets the environment from the energy intercepted by the senses

  Examples:

  - light and sound → carry information about objects or events that might be at considerable distance from us
  - pressure on the skin → information about objects we touch

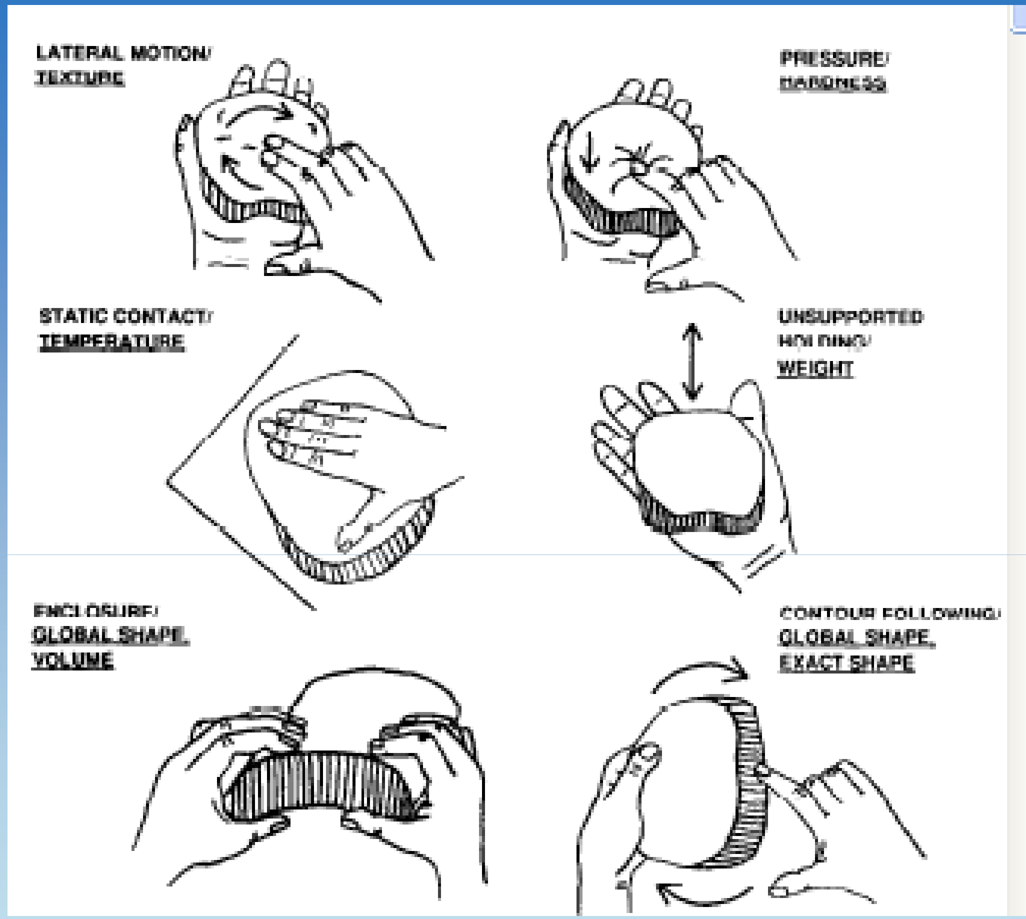
Perception is an active process

- This energy is of no use unless it is channeled to our receptors (light must be focused to the retina, sound channeled to the inner ear)
- We always have an active role in this process → move the eyes or the neck to look at something, explore objects with the hand to determined their shape, consistency or texture



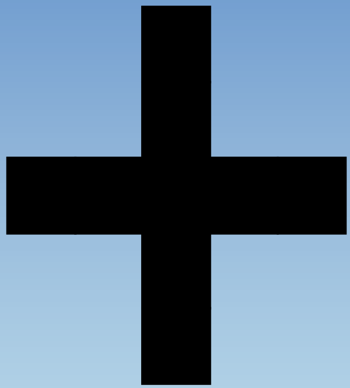
(from Yarbus 1967)

Subjects use specific (optimal) hand movements to evaluate different object properties



Adapted from Lederman & Klatzky, 1993

Experience affects perception



a cross

0	0	1	0	0
0	0	1	0	0
1	1	1	1	1
0	0	1	0	0
0	0	1	0	0

a binary cross

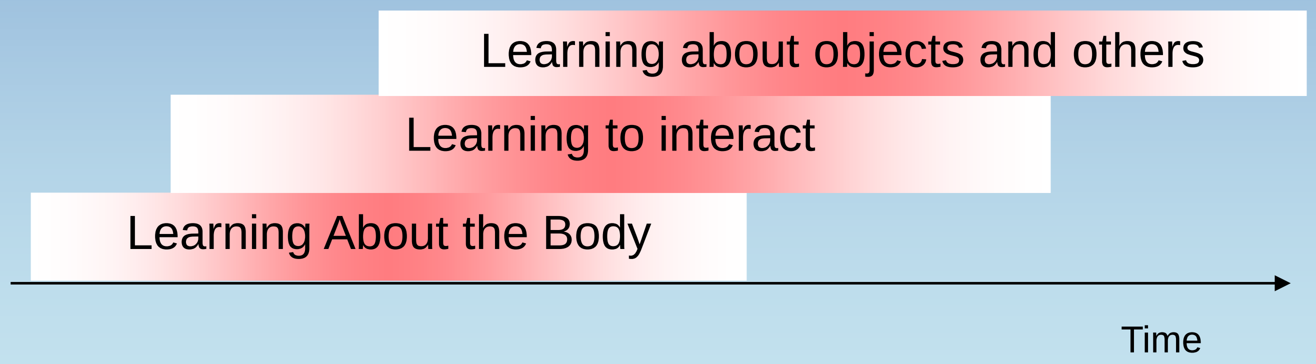
4	12	17	4	15
9	21	3	10	25
5	23	11	37	13
8	18	7	42	6
27	46	31	32	50

?

Developmental Robotics

1. Take inspiration from infant development
2. What we know has been learnt with great efforts

1. Focus on **adaptation**, rather than **performance**
2. Can the robot learn **everything** at the same time? constraints dimensionality, incremental learning → study **the process of building** a complex system



Goal of the talk

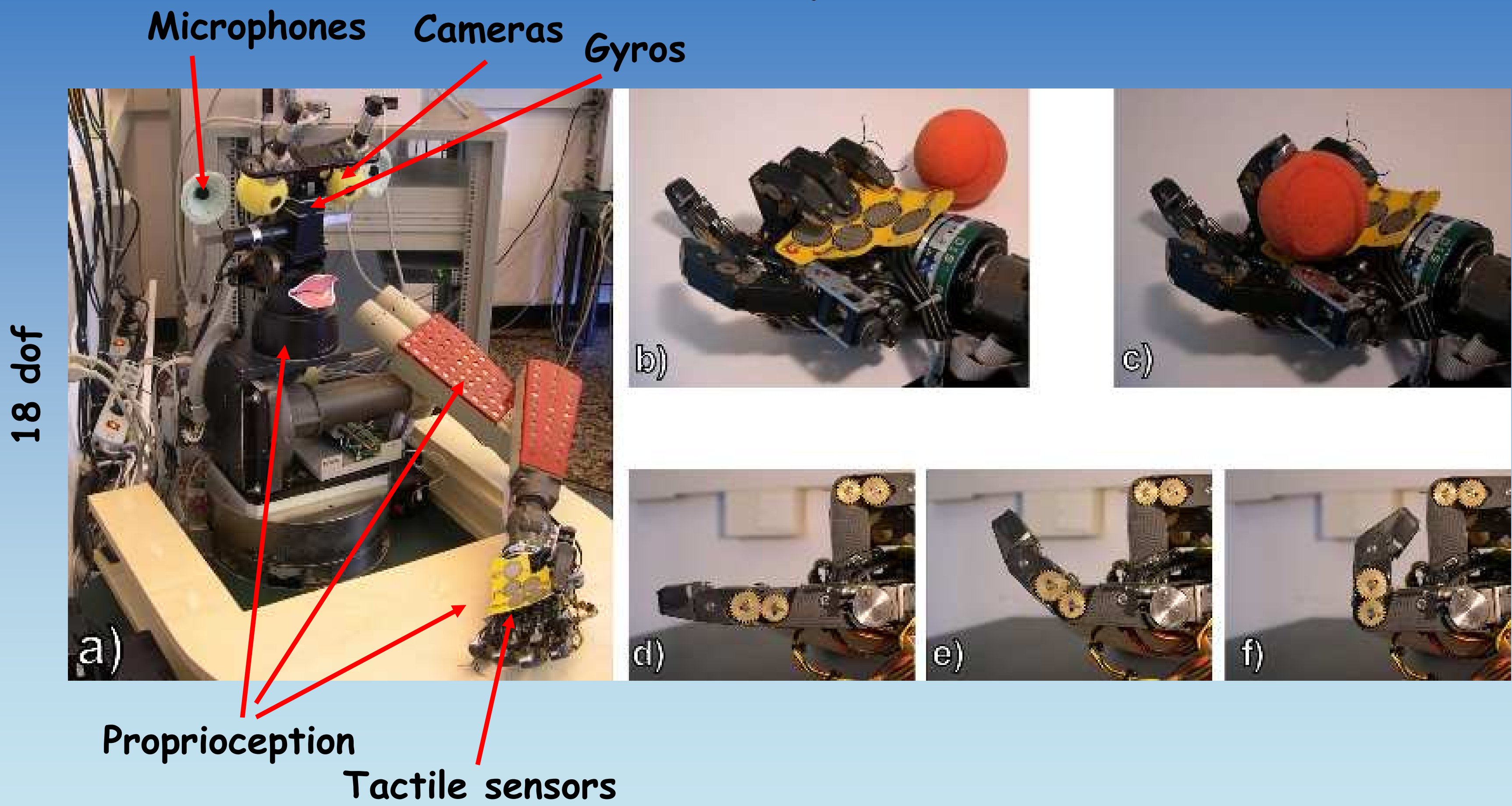
- In the rest of the talk I'll present some "robotics" examples
- I'll try to prove:

  - the importance of the body as a "processing" device

  - the role of actions to solve perceptual tasks

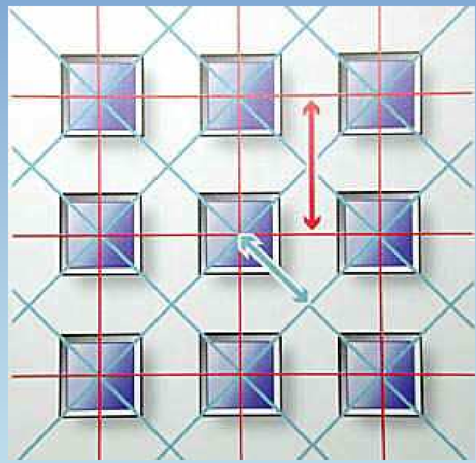
  - and how this can improve actions, that is the way the robot interacts with the environment

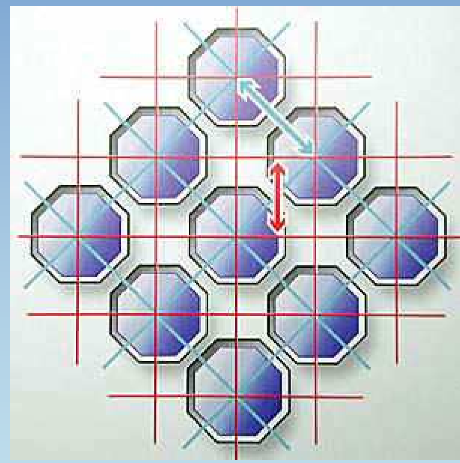
Robotic platform 1: Babybot



Back to sensing: camera sensors

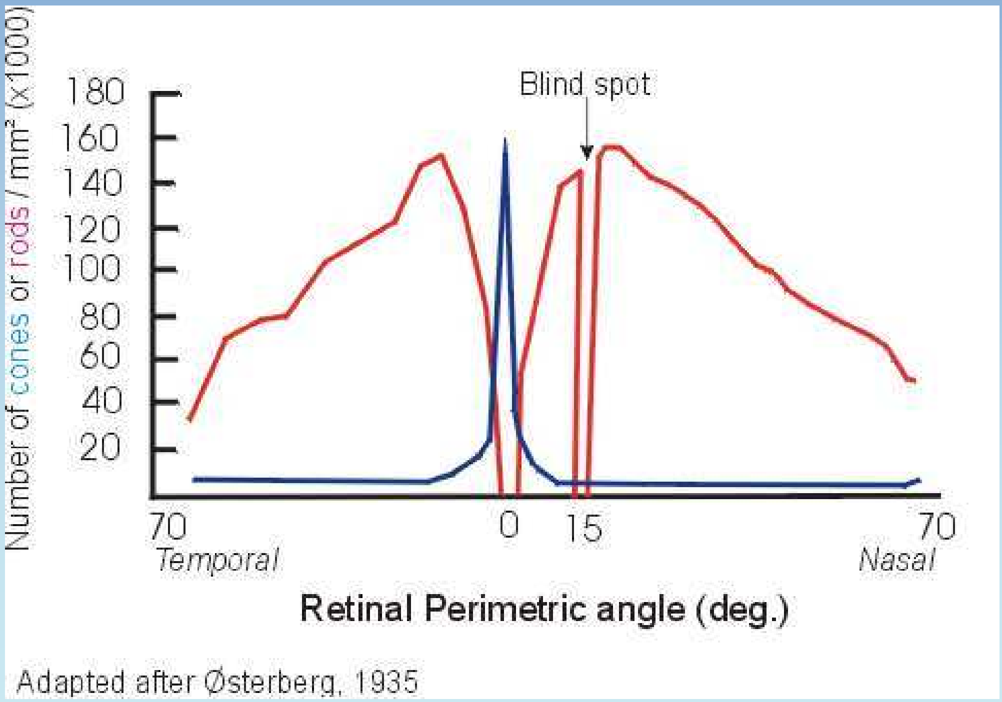
- A digital image is made up of tiny elements called *pixels*
- Photosites on the sensor capture the *brightness* of a single pixel
- The typical layout is a rectangular grid

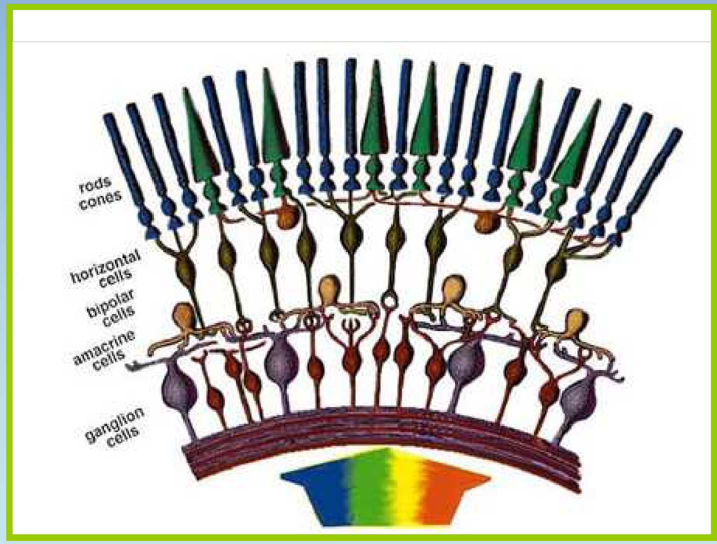


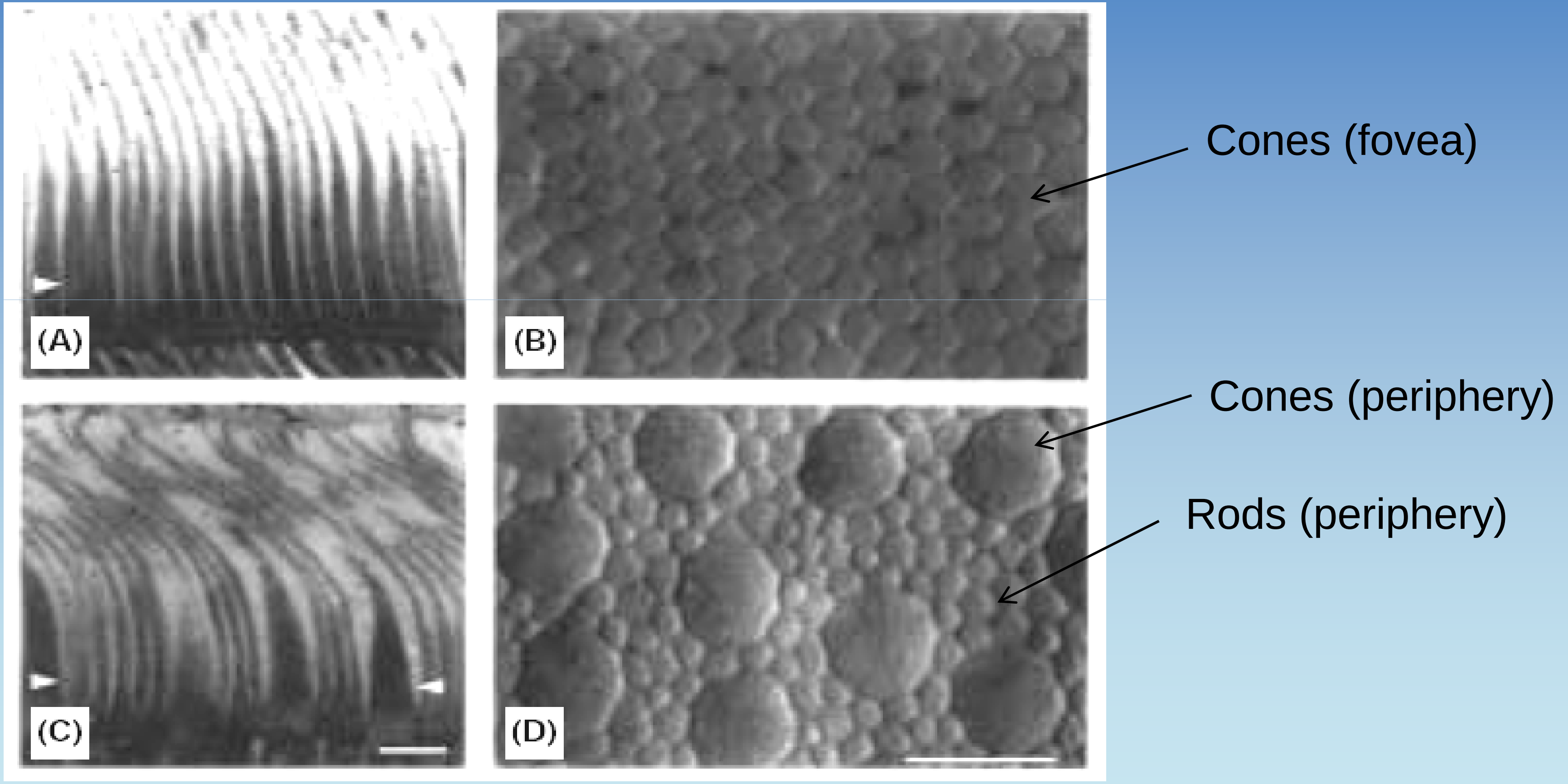


A different solution: the retina

- Two types of receptors in the human retina: cones and rods
- Receptor have a spatial arrangement






(A)
(B)
(C)
(D)
Cones (fovea)
Cones (periphery)
Rods (periphery)

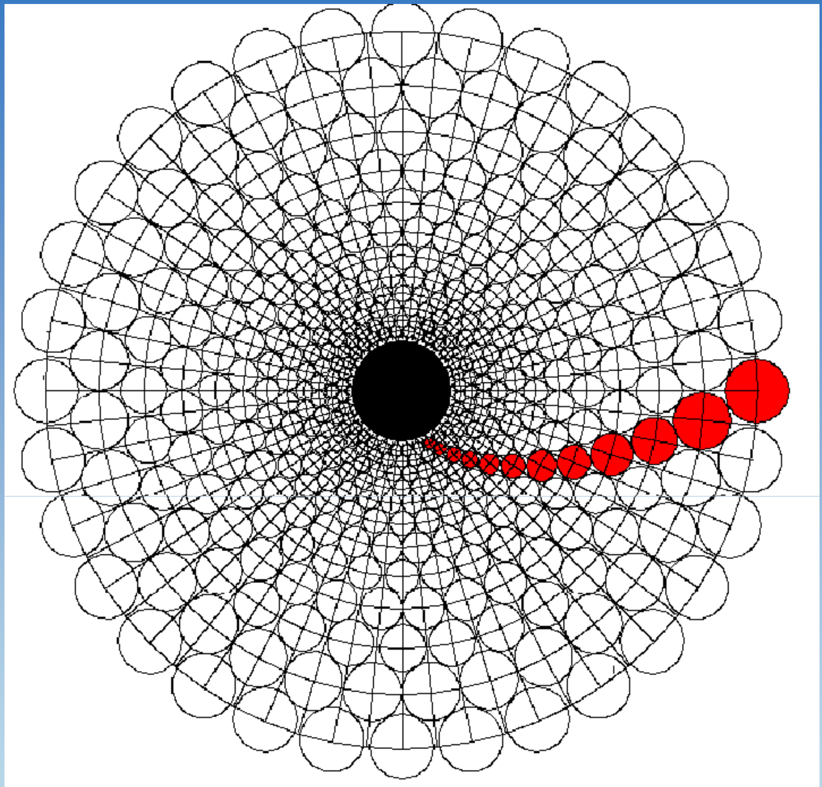
If the retina was designed as a camera

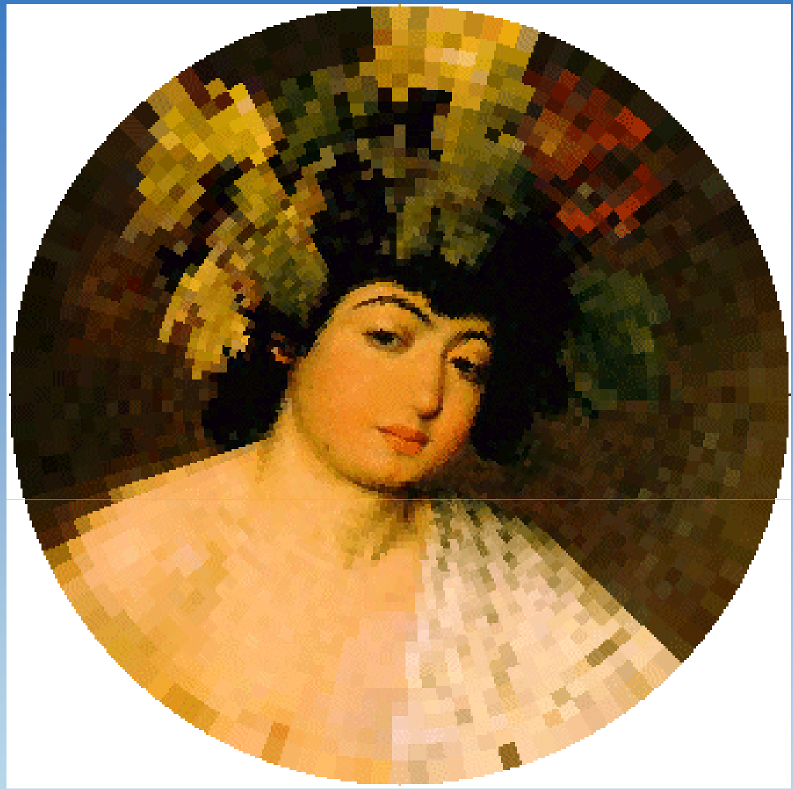
Visual Field: about 160 deg
Maximum Resolution: about 1/60 deg

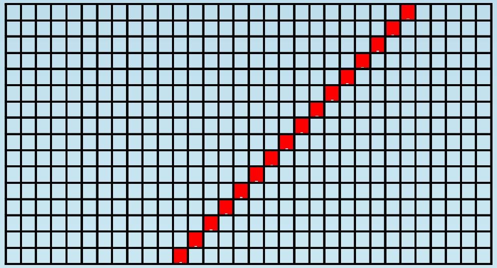
According to K. Nakayama and E. Schwartz the saving is from 5,000 to 30,000 times.

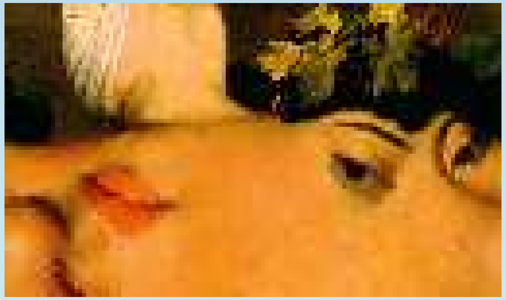
Optic nerve: diameter 4 cm
Brain weight: from about 3 to 20 tons
Amount of food: ?
Processing time: ?

Log-polar images

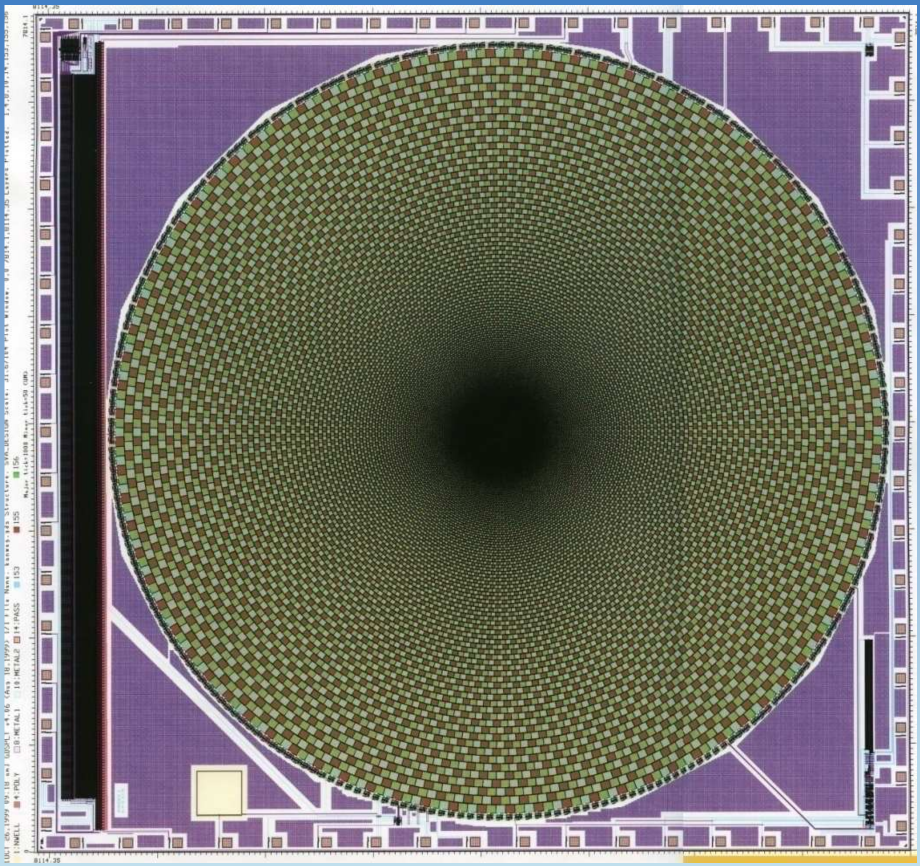






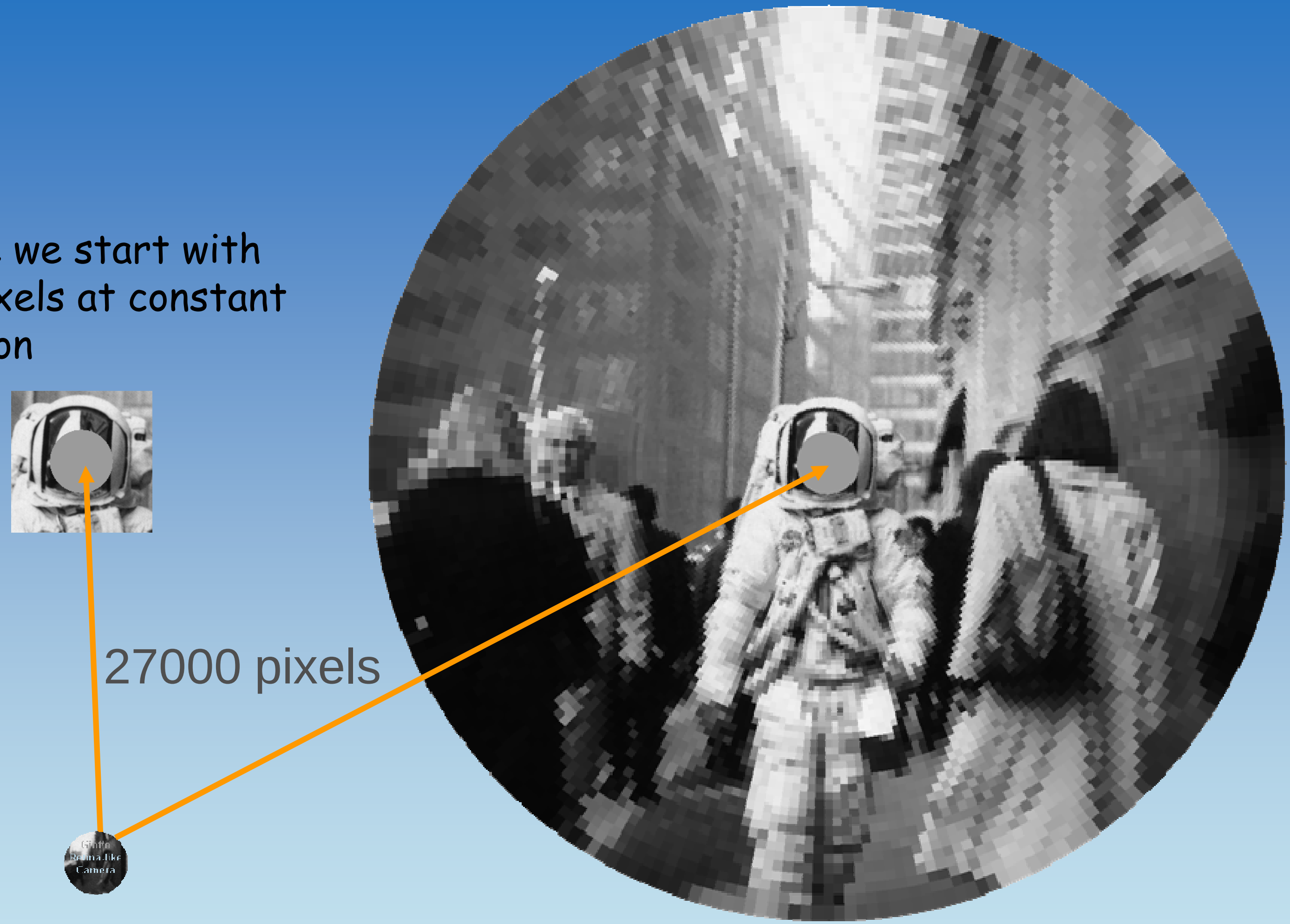


1998 – CMOS Color



Same Layout

Suppose we start with
5000 pixels at constant
resolution



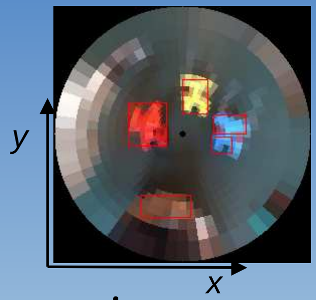
5000 pixels

The same 5000 pixels
plus 27000 retina-like pixels

Controlling gaze, example sound localization

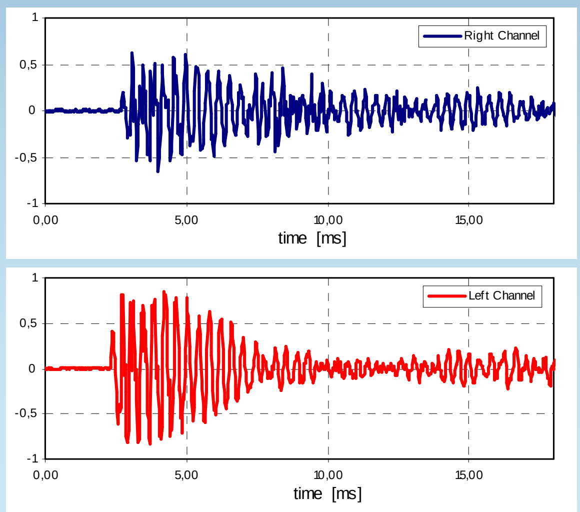
What is sound localization ?

- visual information is spatially organized



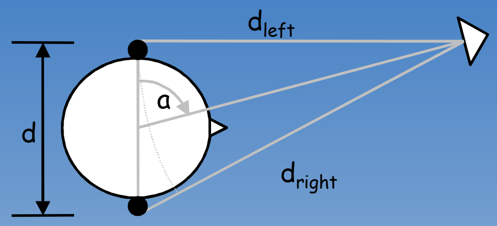
$$s = f(x, y)$$

- we need some sort of computational process to extract spatial information from the sound signal



$$s = f(t) \longrightarrow ? \longrightarrow (x, y)$$

Example: Sound localization

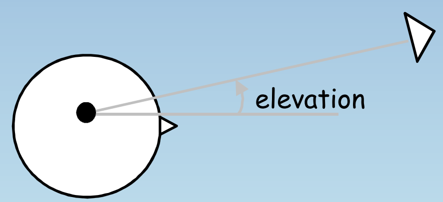


1) Horizontal Component

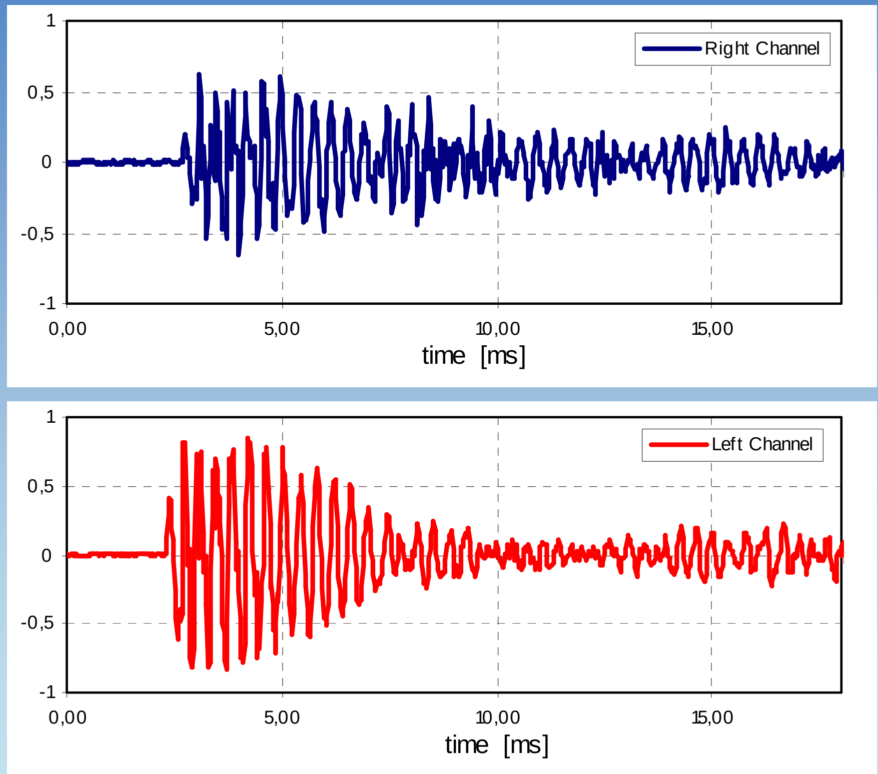
→ Interaural Timing Difference (ITD)

→ Interaural Level Difference (ILD)

2) Vertical Component



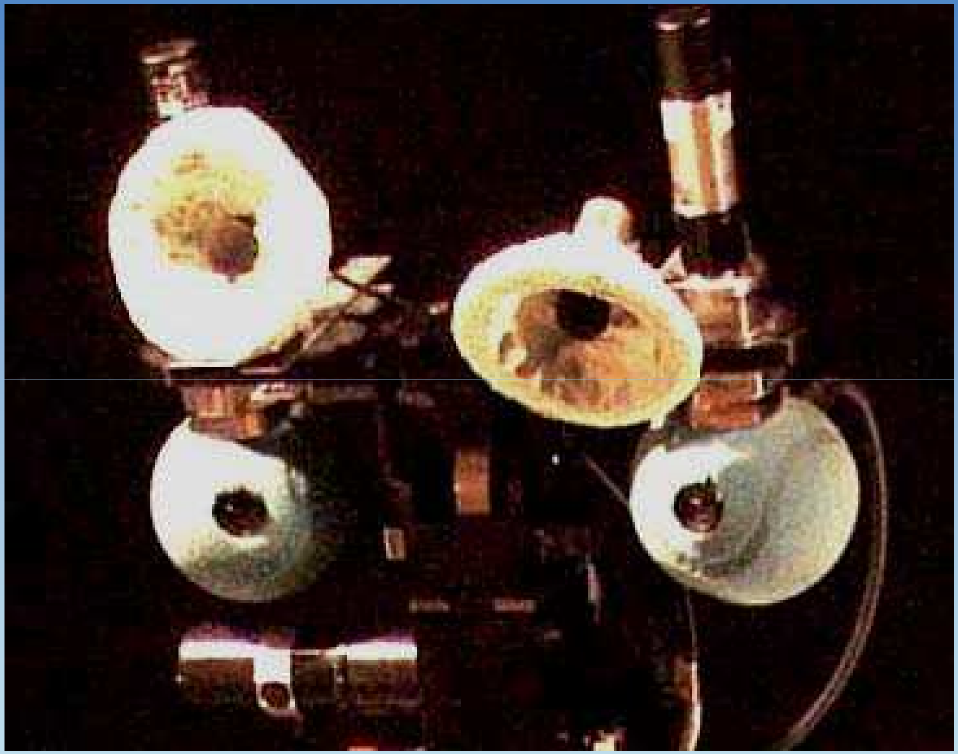
→ Interaural Level Difference (ILD) at high frequency


1
0,5
0
-0,5
-1
0,00
5,00
10,00
15,00
time [ms]
Right Channel
1
0,5
0
-0,5
-1
0,00
5,00
10,00
15,00
time [ms]
Left Channel

A trick…

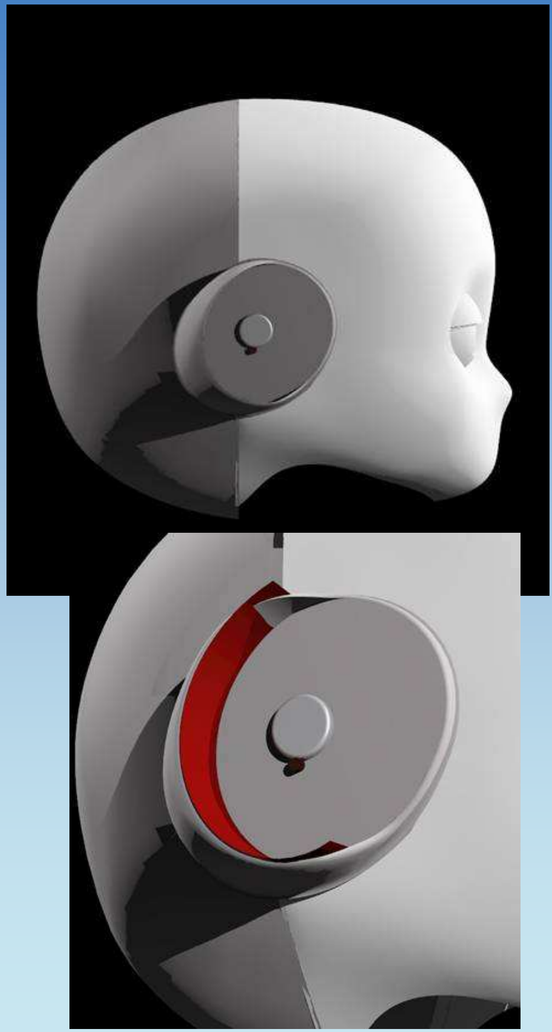
Estimation of the shift between the signals – ITD, horizontal position

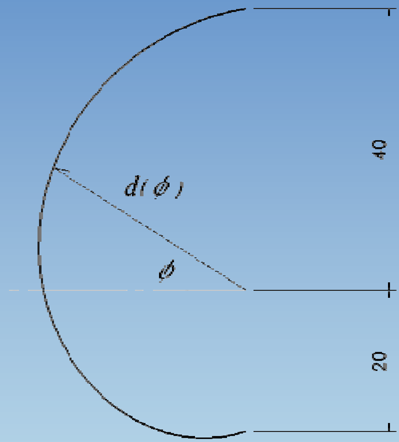
Asymmetric external ears – ILD is "strictly" related to the elevation of the sound source

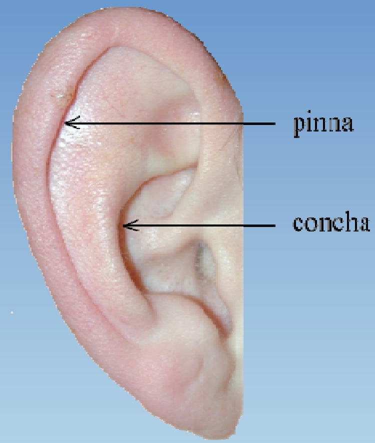


L. Natale, G. Metta, and G. Sandini,
Robotics and Autonomous Systems, 2002

A more accurate method

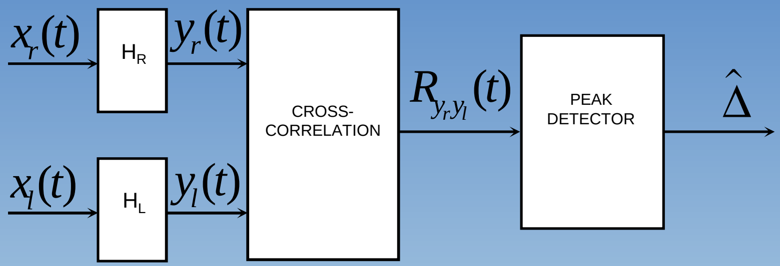




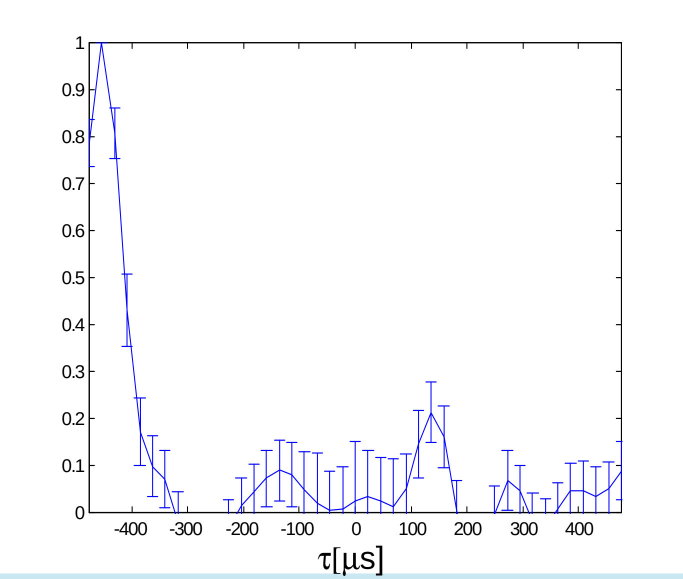


Computation of the ITD

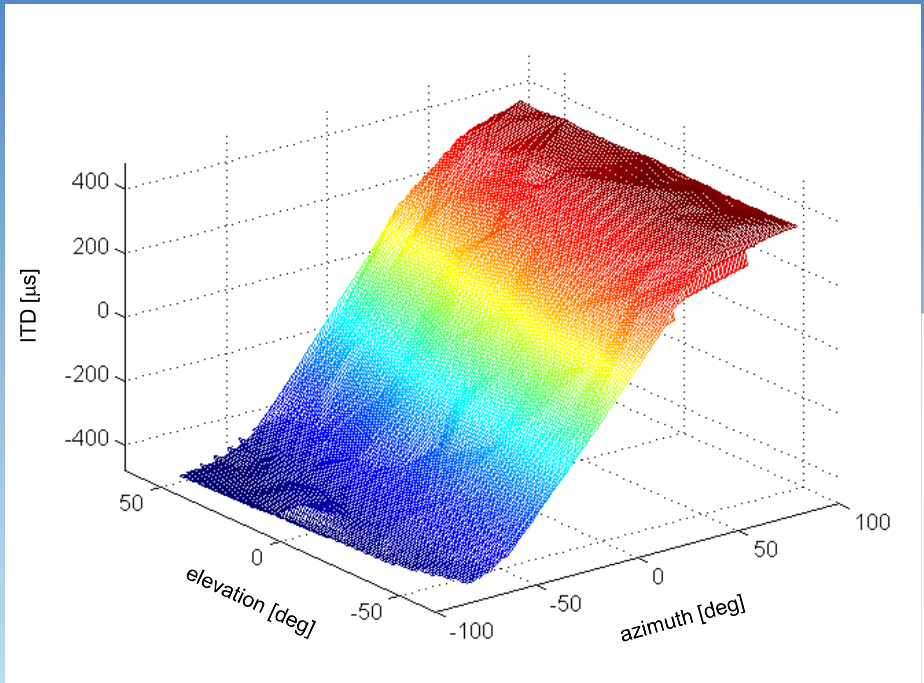
- Generalized correlation method (Knapp 1976)

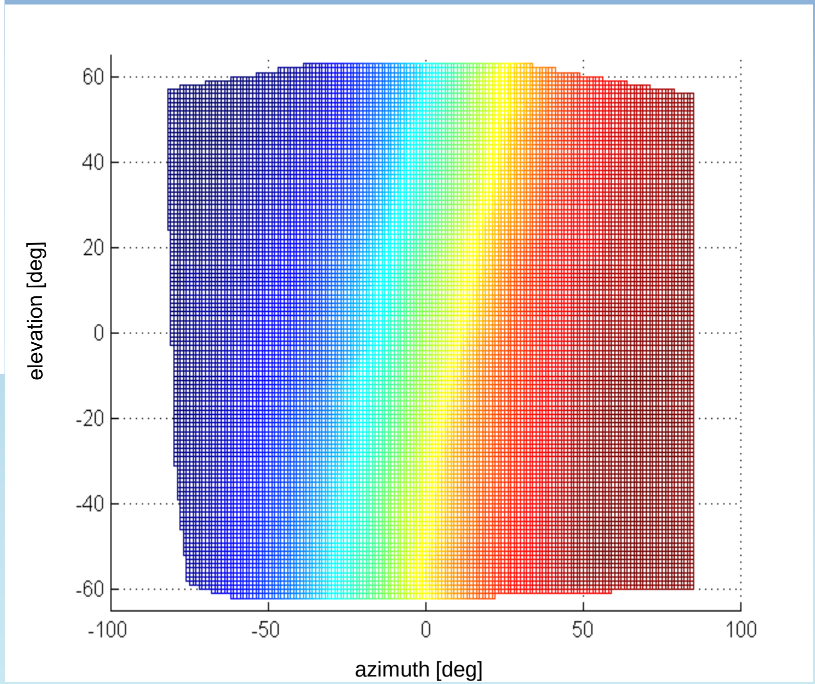


$$R_{y_r y_l}(\tau) = \frac{1}{T-\tau}\int_{-T/2}^{T/2-\tau} y_r(t+\tau)\, y_l(t)\, dt$$



Spatial variation of the ITD



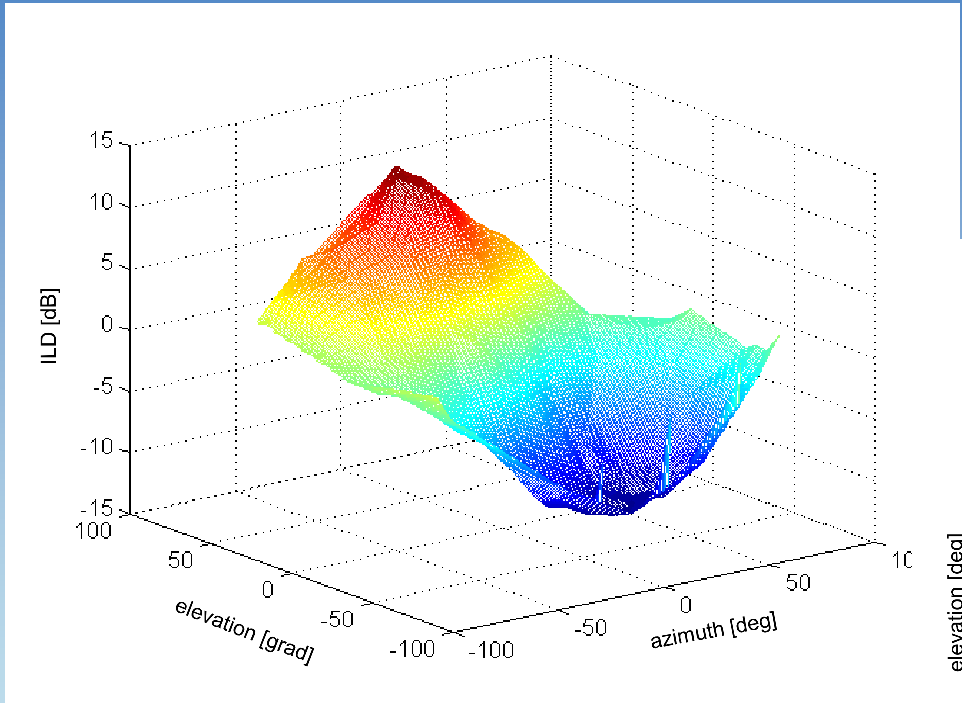


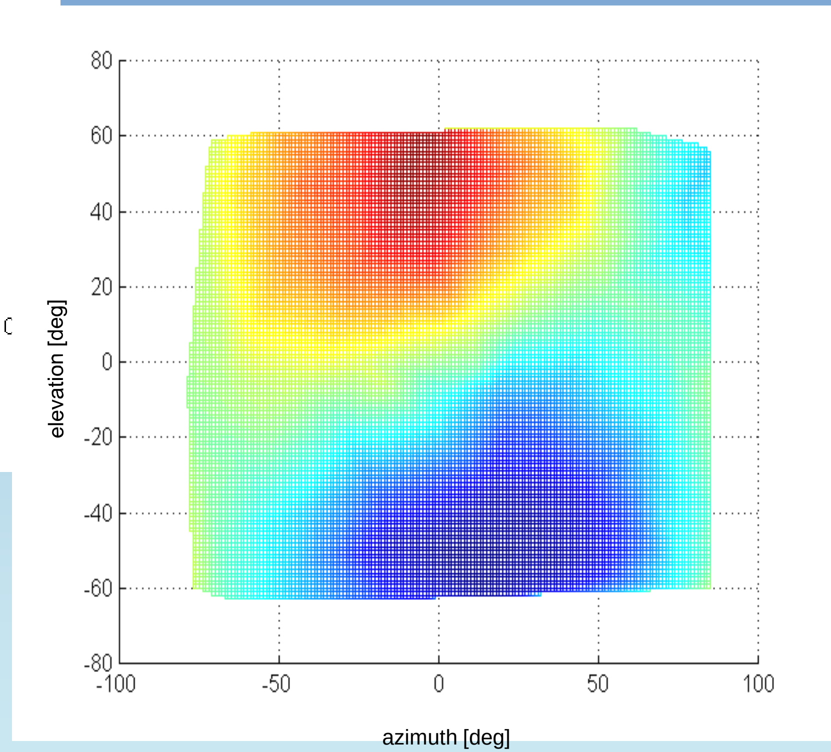
Computation of the ILD

- Ear lobes – directionally dependent response
- High pass filter (>3 kHz)

$$ILD = 10 \cdot \log \frac{\int S_r(f)df}{\int S_l(f)df}$$

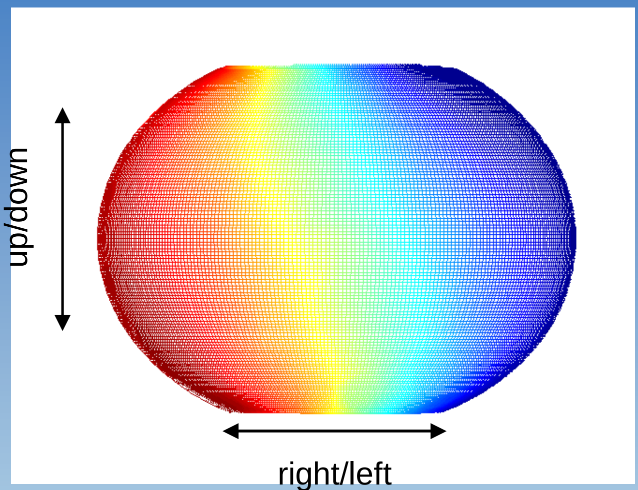
Spatial variation of the ILD



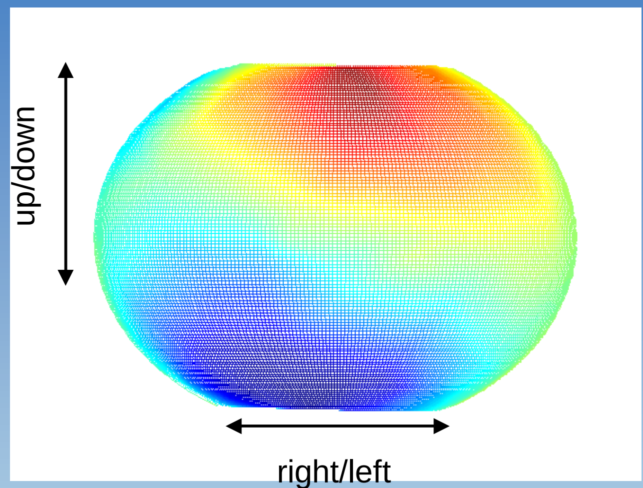


Babybot vs Barn Owl

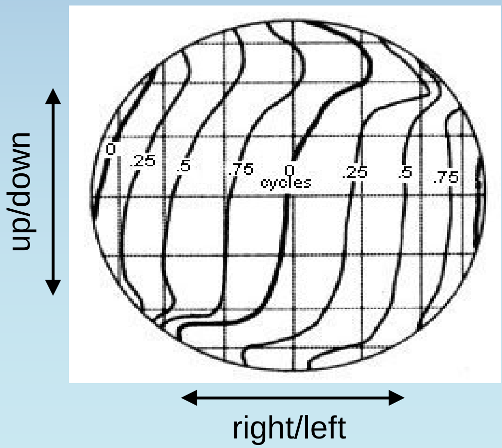
ITD (babybot, white noise)



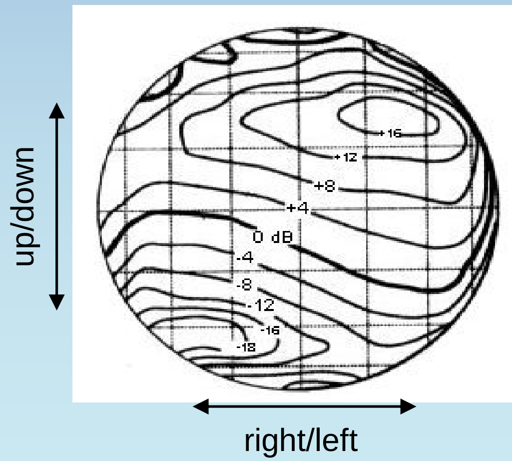
ILD (babybot, white noise)



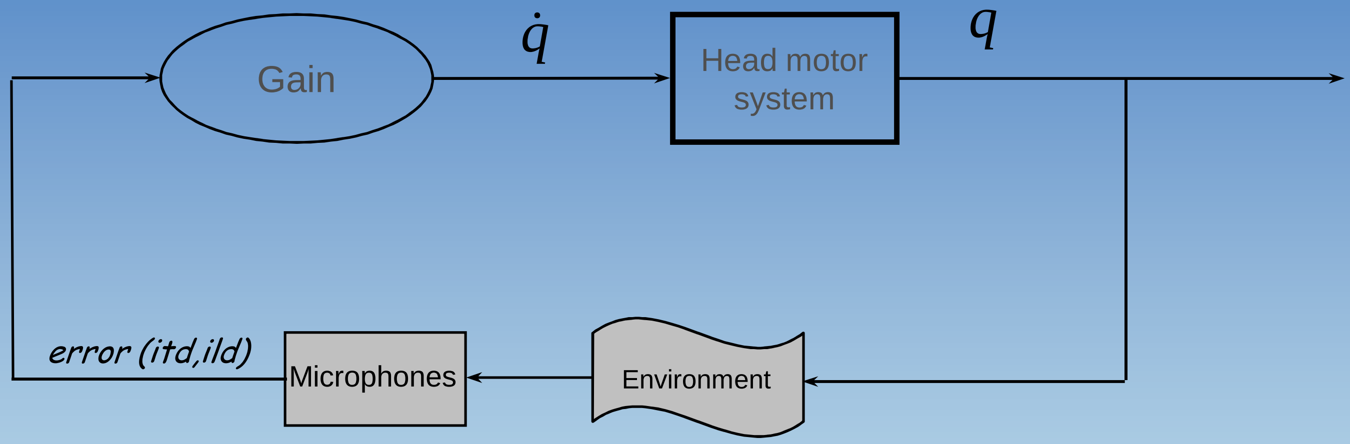
IPD (barn owl, at 6 kHz)

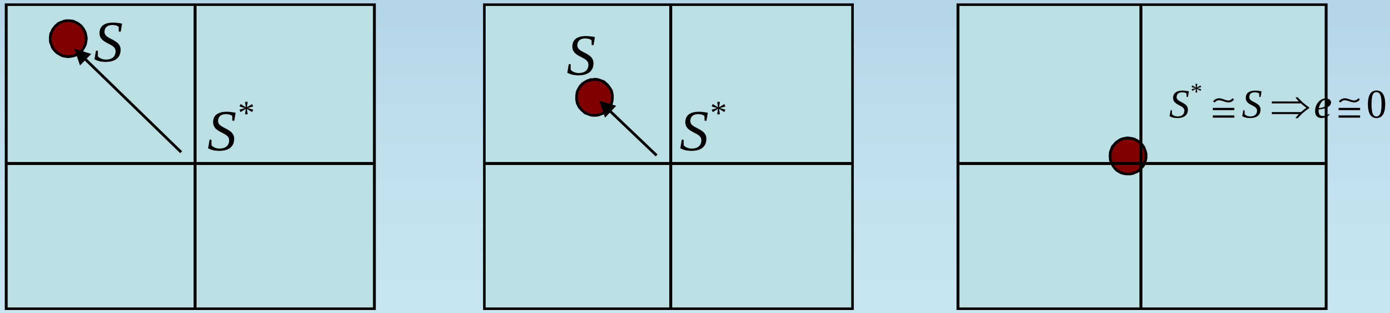


ILD (barn owl, at 6 kHz)

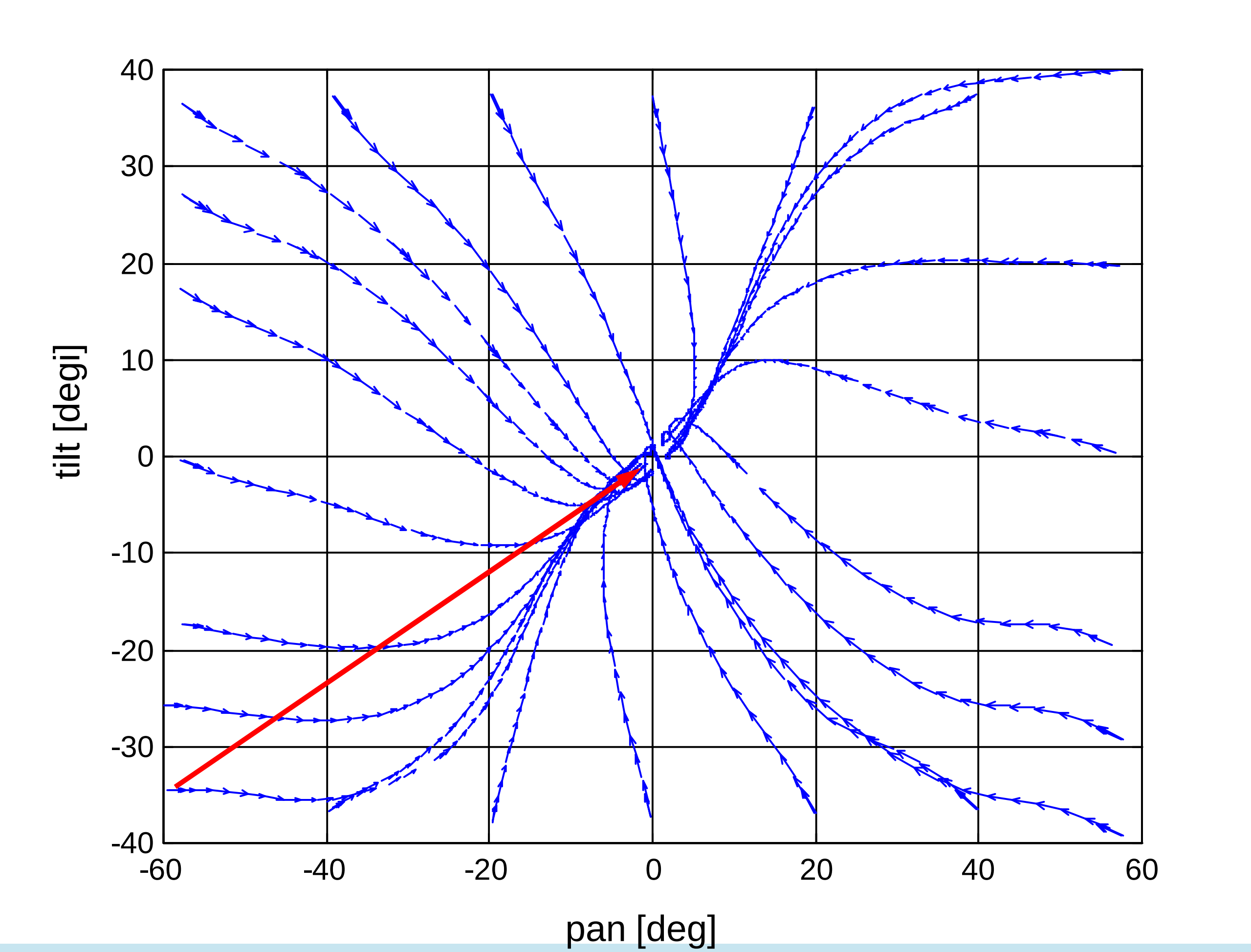


Control schema: closed loop



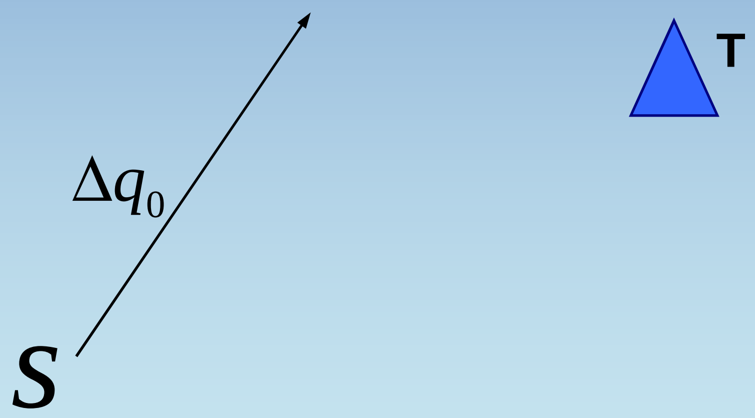


Closed loop trajectories

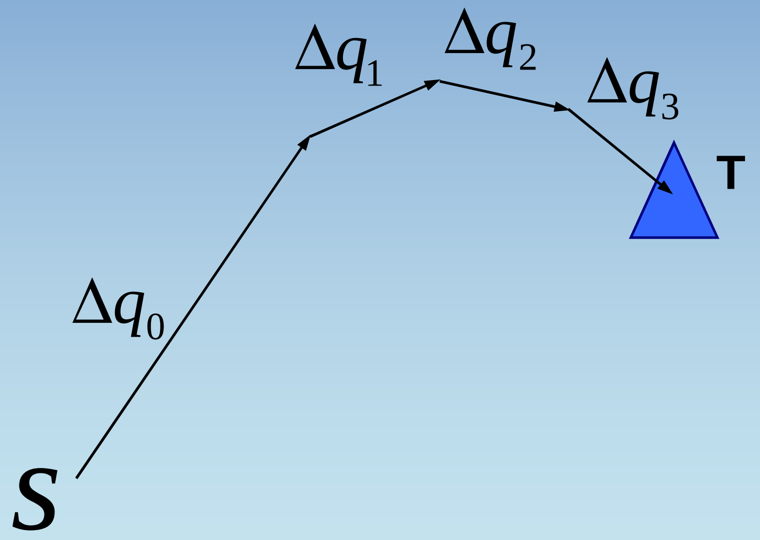


Suppose we have a target T, whose position in sensory space is S

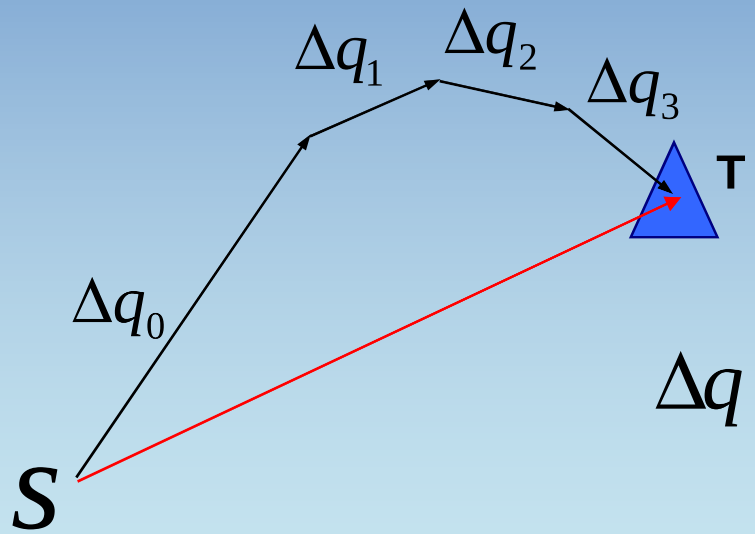
We don't know how to go to T directly, but we know how to move *closer*



We finish the movement using the closed loop control

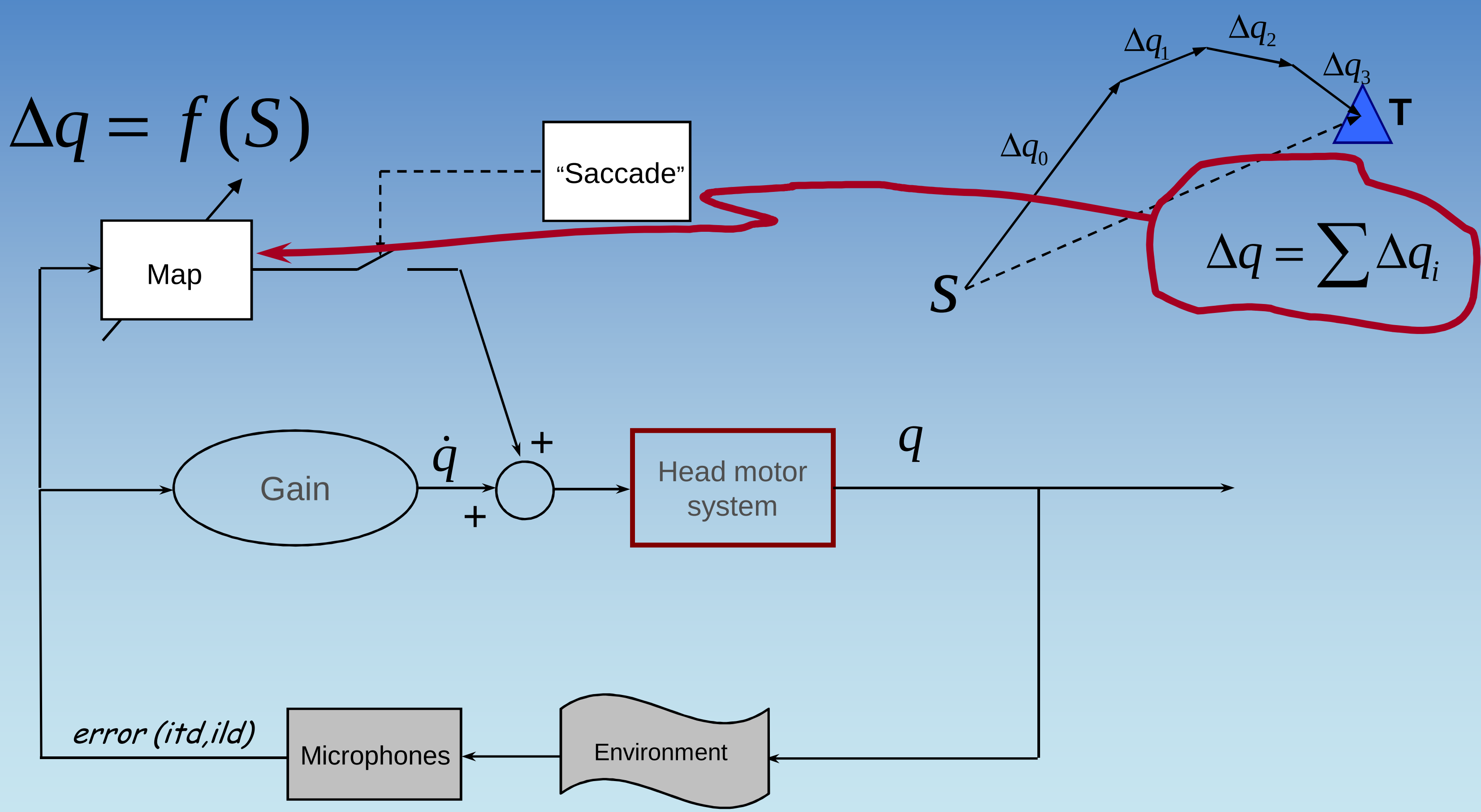


Once we reach the target, we integrate the trajectory to get a command that we store and use in the future, if we are presented a target in the same location

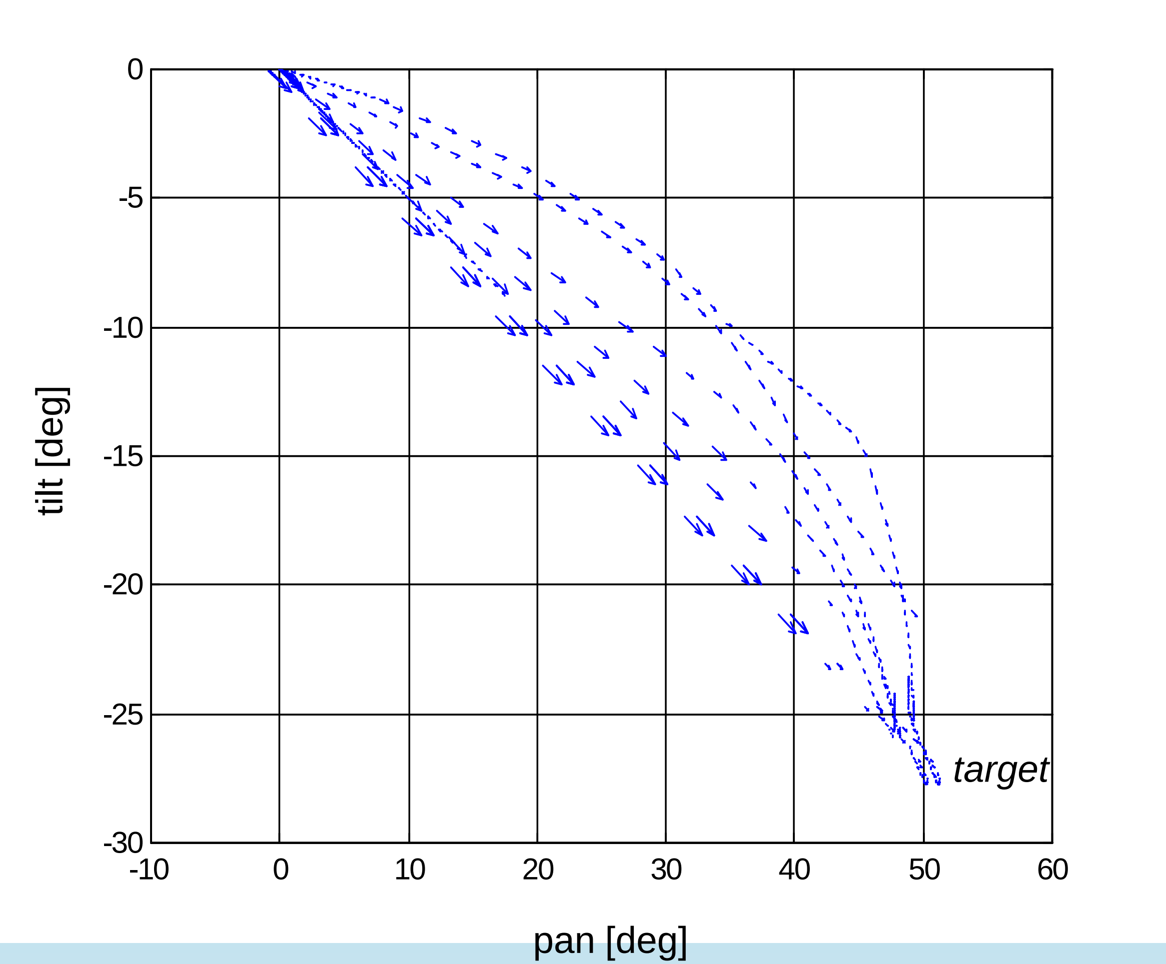


$$\Delta q = \sum \Delta q_i$$

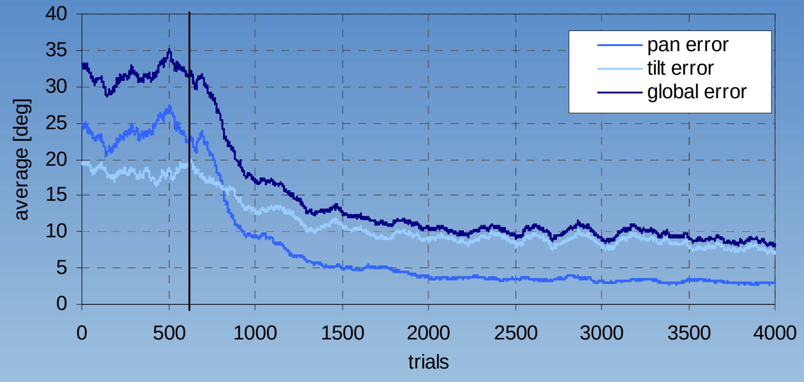
Control schema (2)

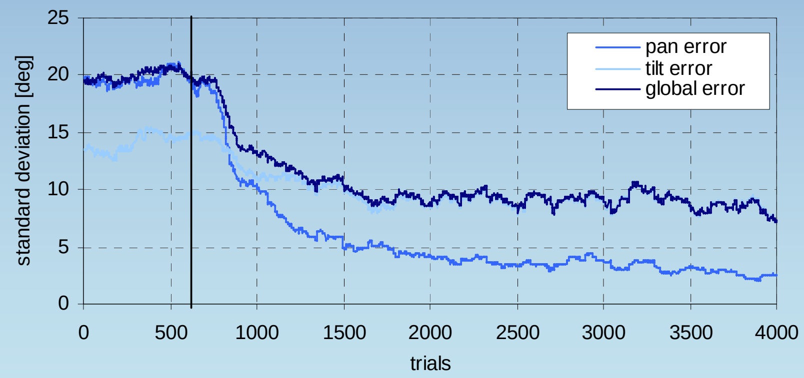


Learning (1)

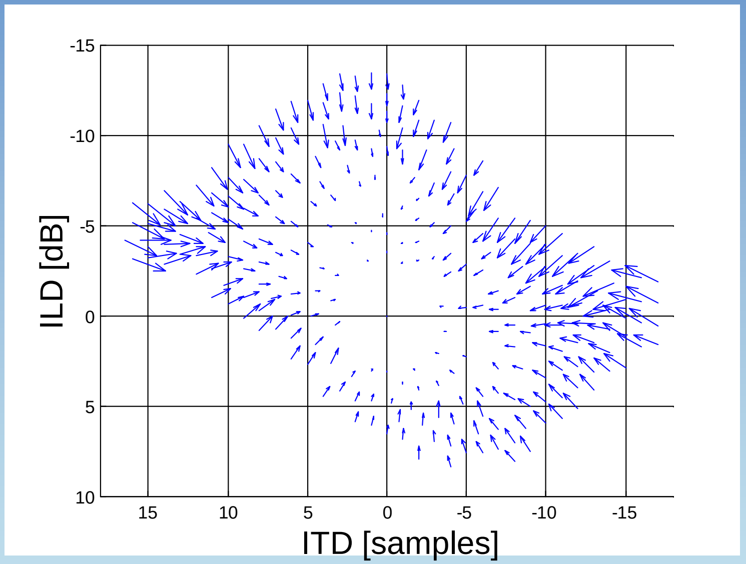


Learning (2)

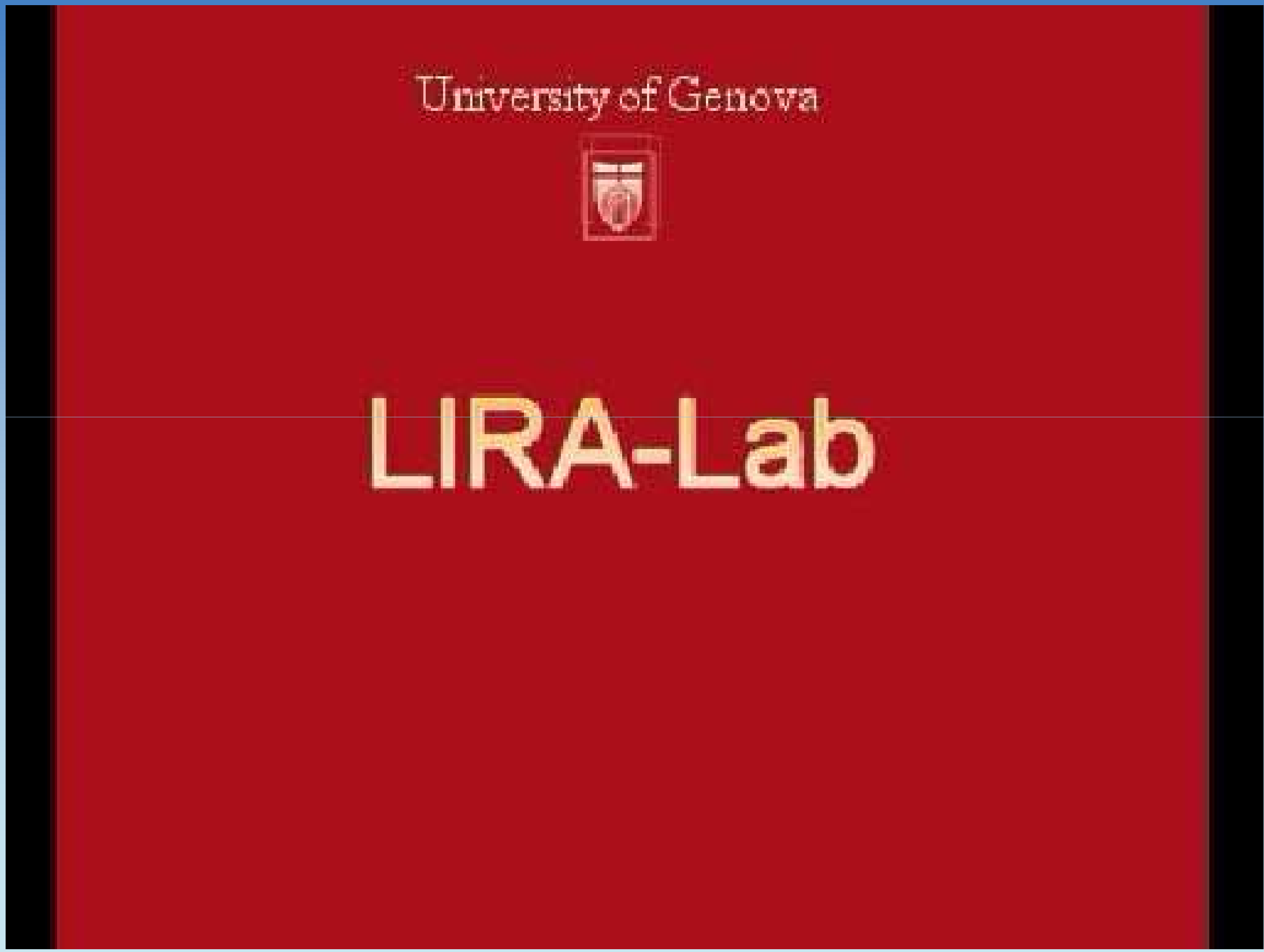


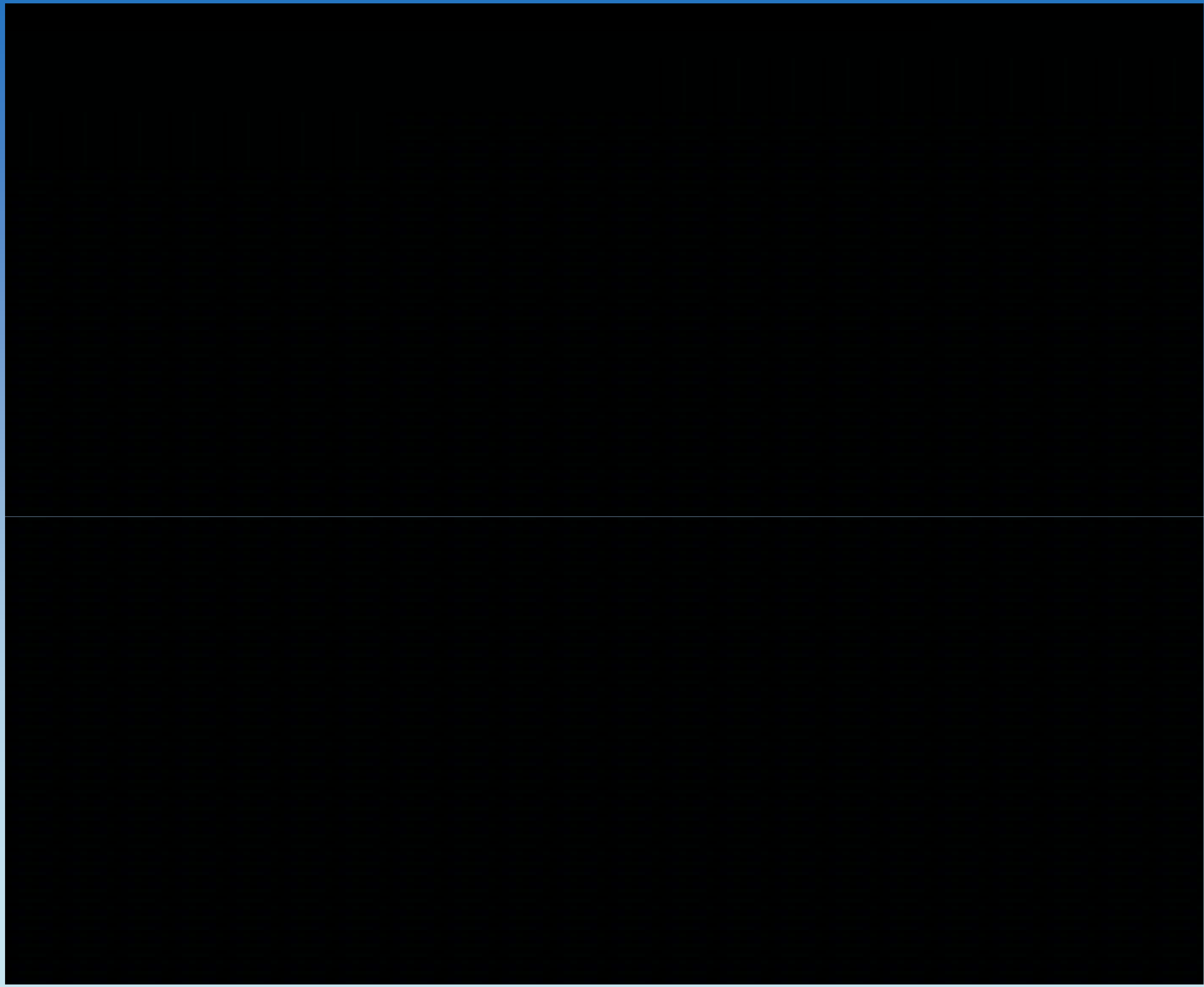


Map
(lookup table)



Sound localization clip





To recap so far:

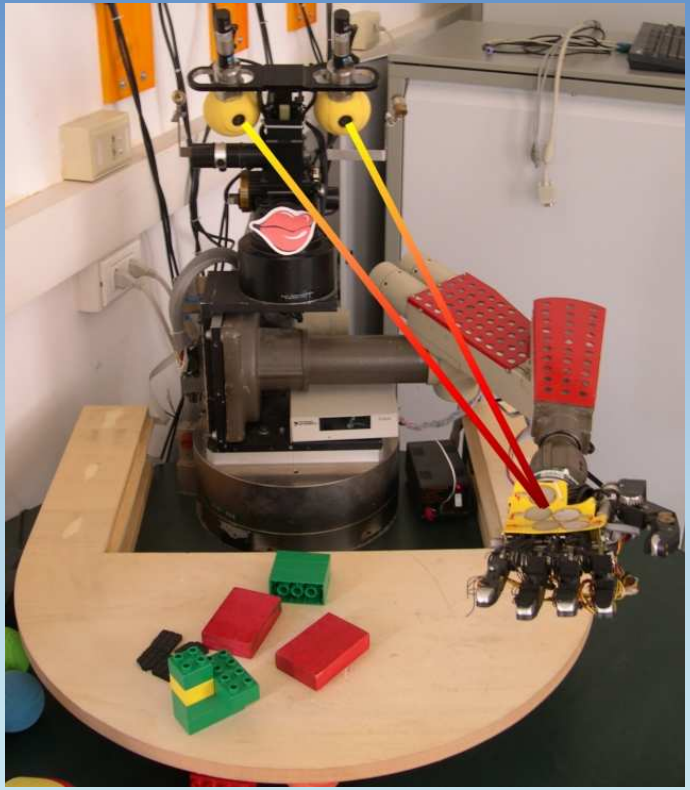
- Biological systems can suggest “smart” solutions
- Even inaccurate signals can be useful to “close the loop”
- A simple controller can bootstrap learning and improve performances

Learning about the body

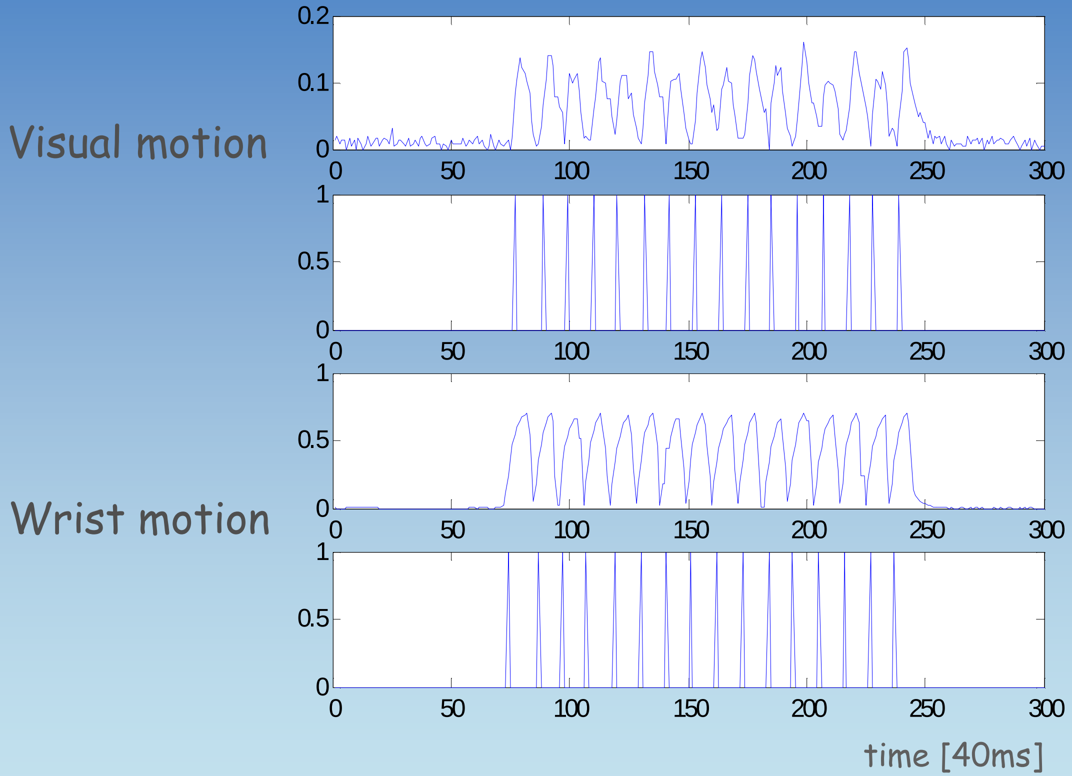
Body Map: Hand localization

General question: how can the robot discriminate between its body and the environment ?

- the body is something the robot **can control**
- link to infant development (Rochat and Striano, 2000):
  - combined double touch
  - multimodal perception → vision and proprioception

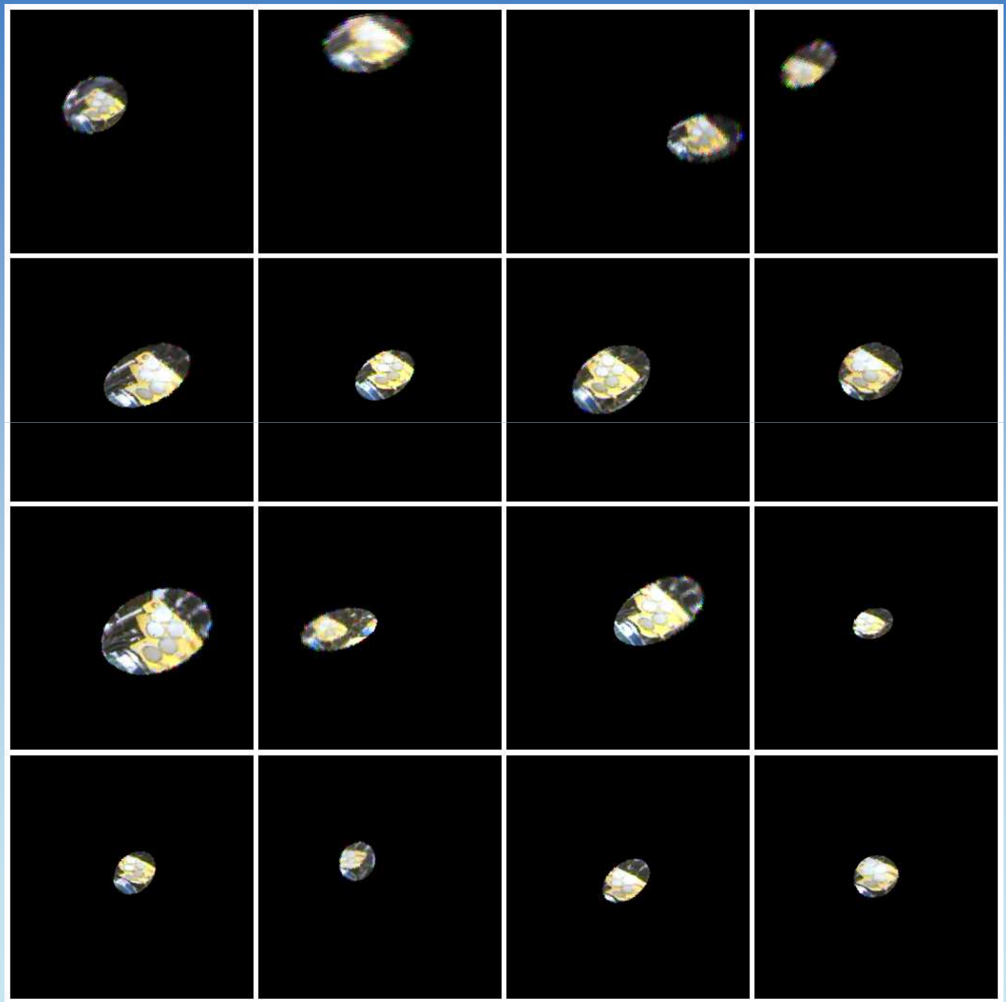


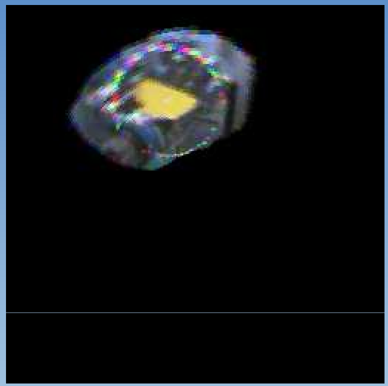
Periodic motion example:



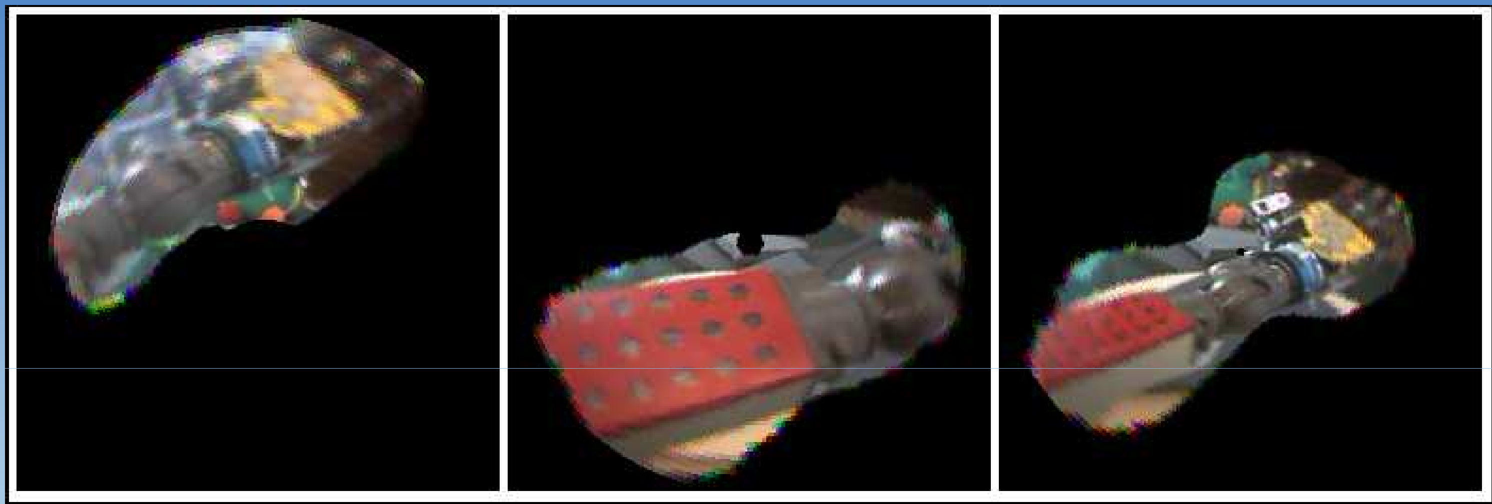
L. Natale, F. Orabona, G. Metta, G. Sandini,
Progress in Brain Research, 2007

Hand segmentation: examples





Other body parts

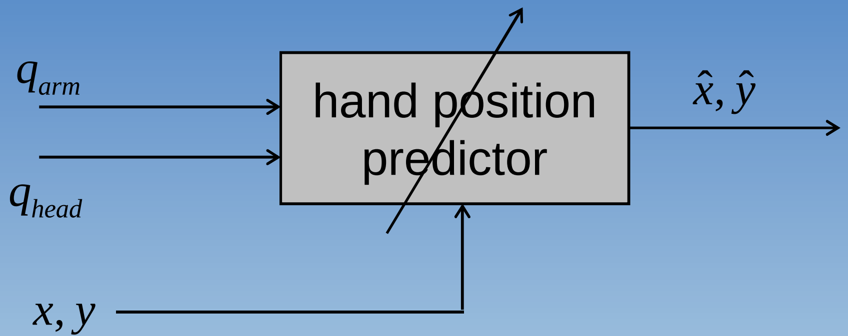


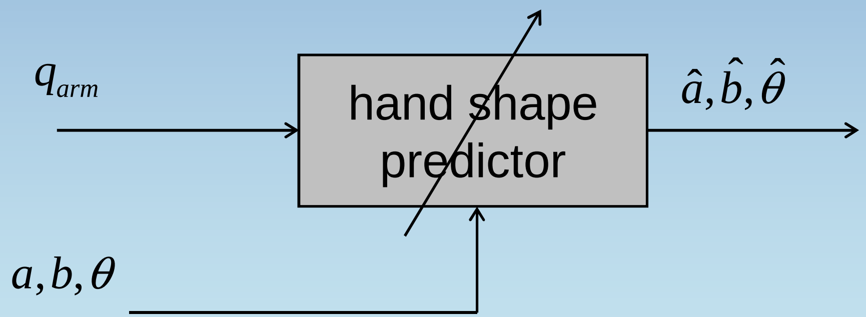
The same algorithm can be used to segment any body part that is visible when it moves

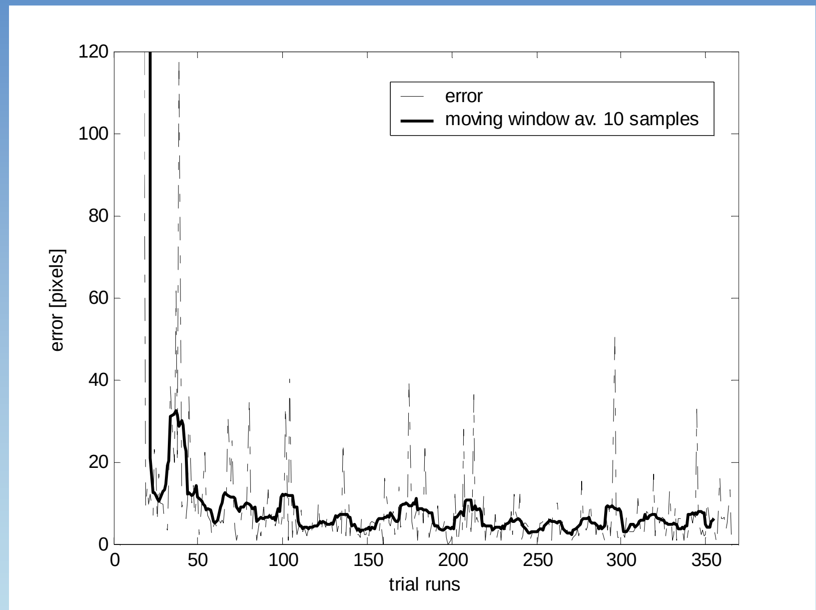
Building a model of the hand

- this algorithm cannot be used to track the hand of the robot or to localize it during a grasping action
- however it is a good starting point to build more complex models of the hand:
  - ellipse fitting, train a neural network to compute position and shape of the hand in the image plane based on the current arm configuration
  - color histogram

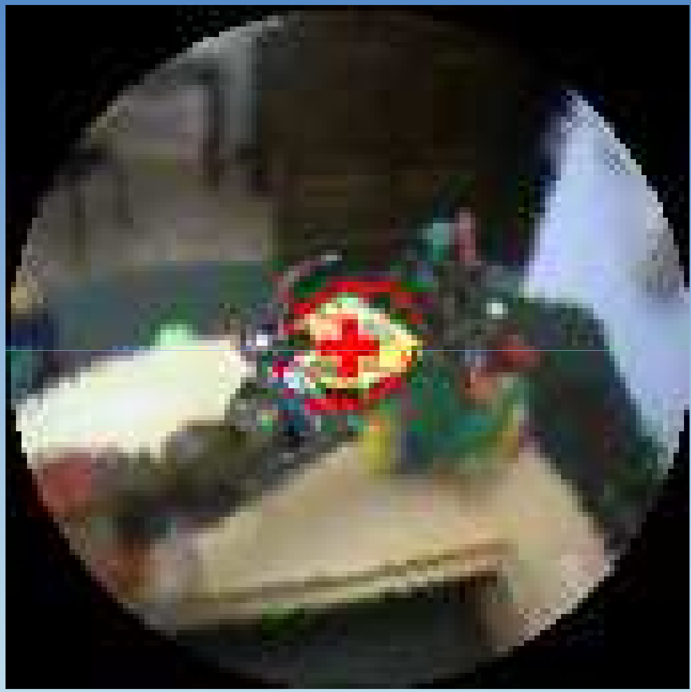
Hand localization: forward models



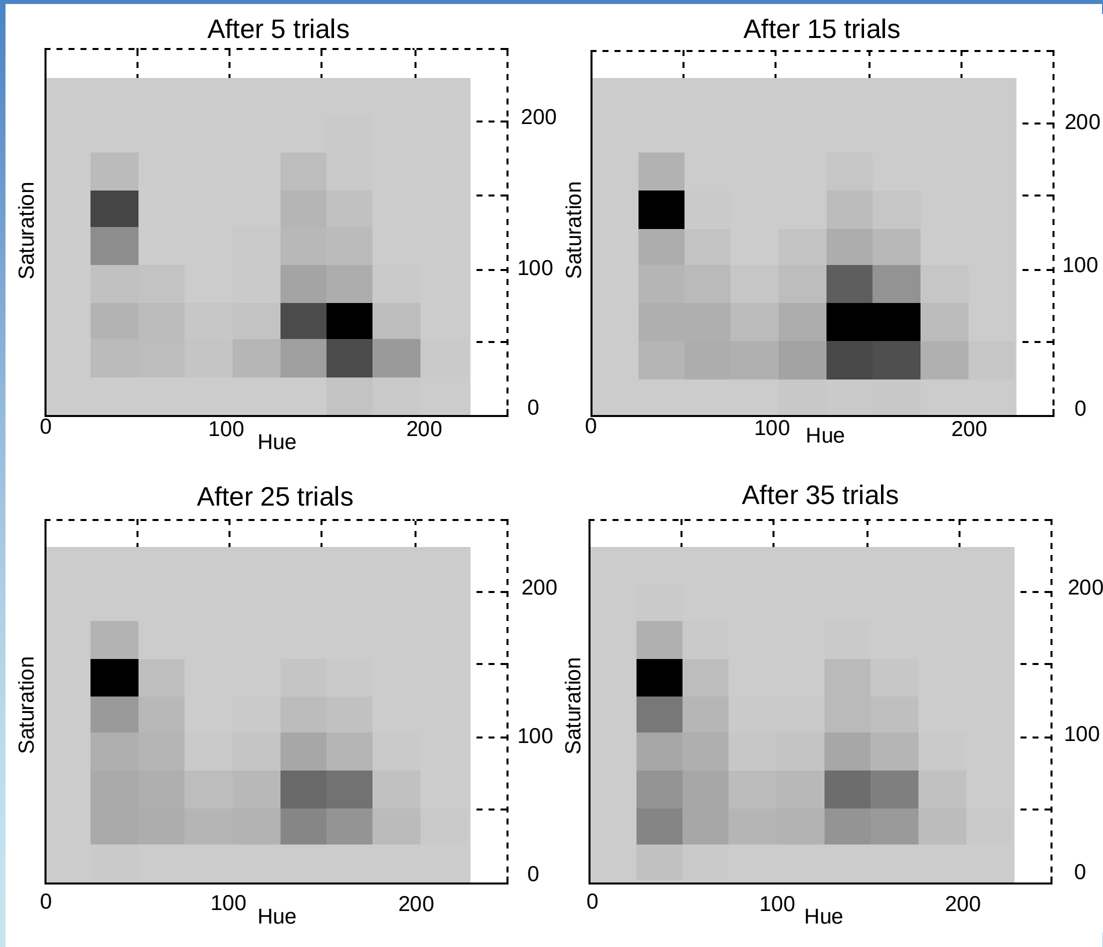




Hand localization: clips



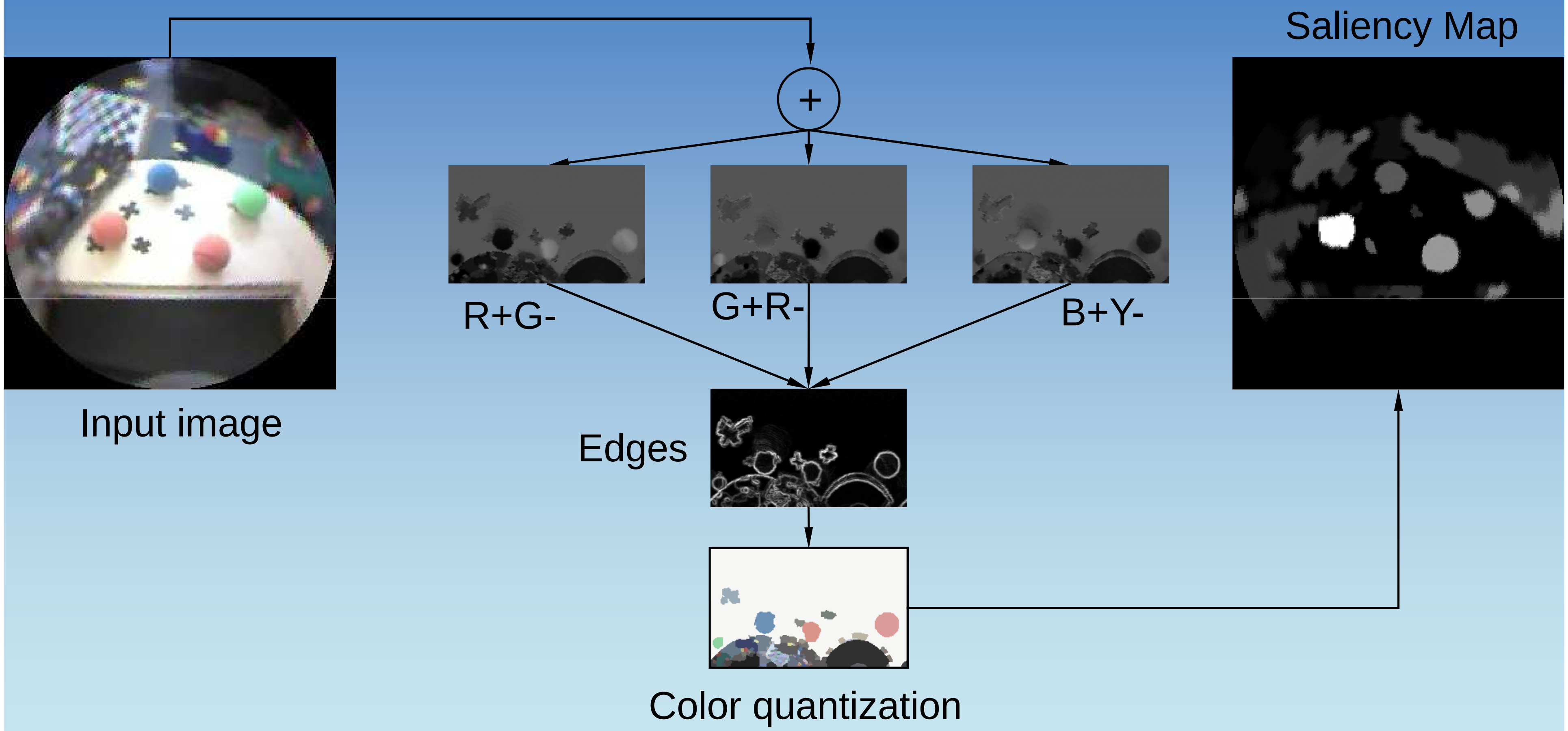
Hand localization: color histogram



Hand localization: color

Learning about objects

Start with a simple blob detector



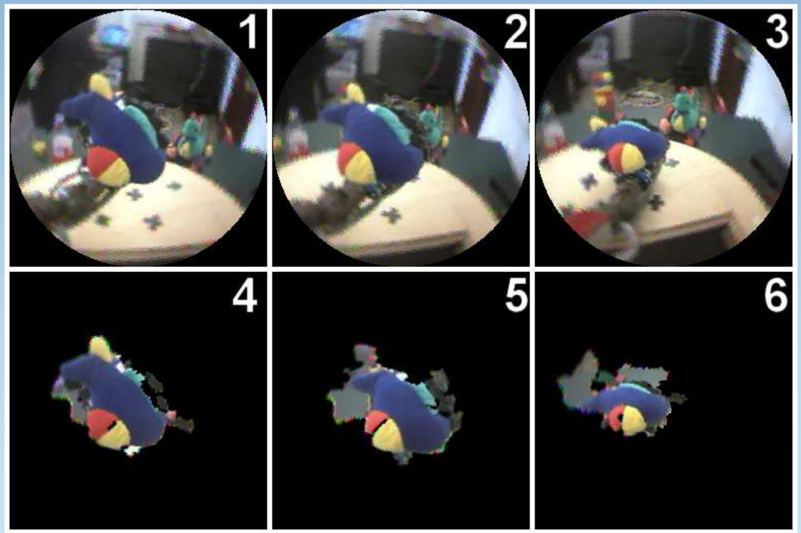
L. Natale, F. Orabona, G. Metta, G. Sandini,
Progress in Brain Research, From Action to Cognition, 2007

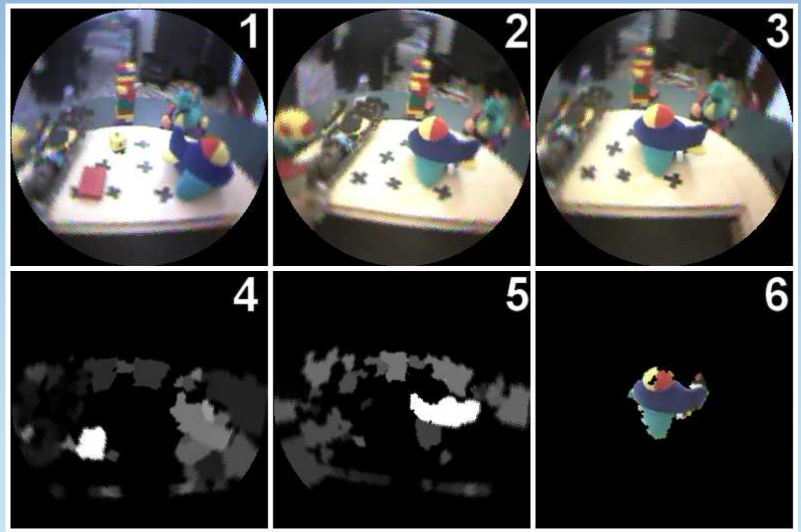
Learning about objects

Extending the concept of object through interaction

- Watching the hand holding the object
- Hypothesis: central blob $\in$ object
- Estimation:

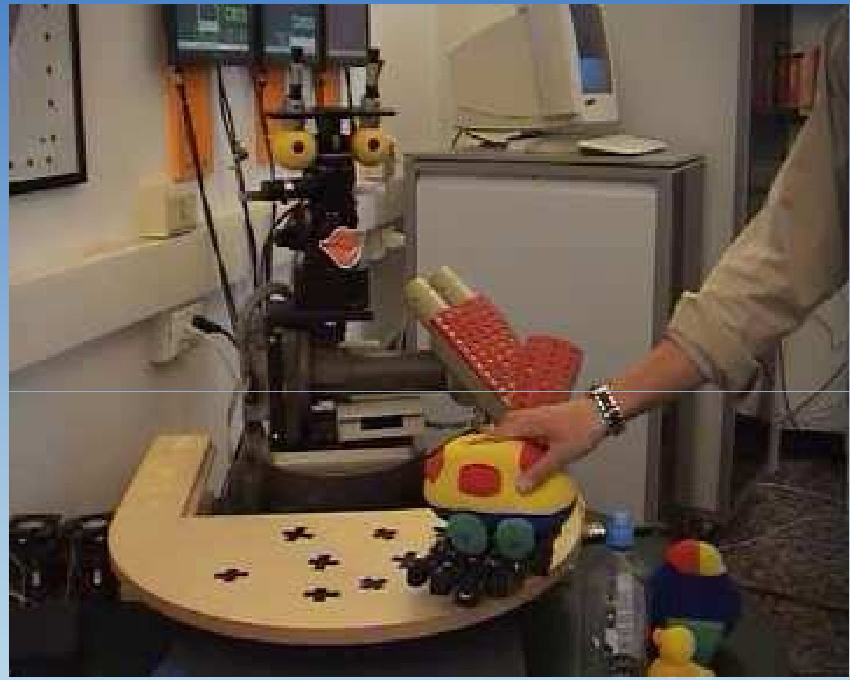
$$P(blob_i \in object | fixating\ object)$$





A couple of examples

Grasping a toy airplane



Grasping a toy car

To recap:

- Redundancy (periodicity) can be useful
- Knowledge about the body is important (reference point)
- Exploit actions to produce sensory feedback and bootstrap learning

The role of haptic feedback during manipulation

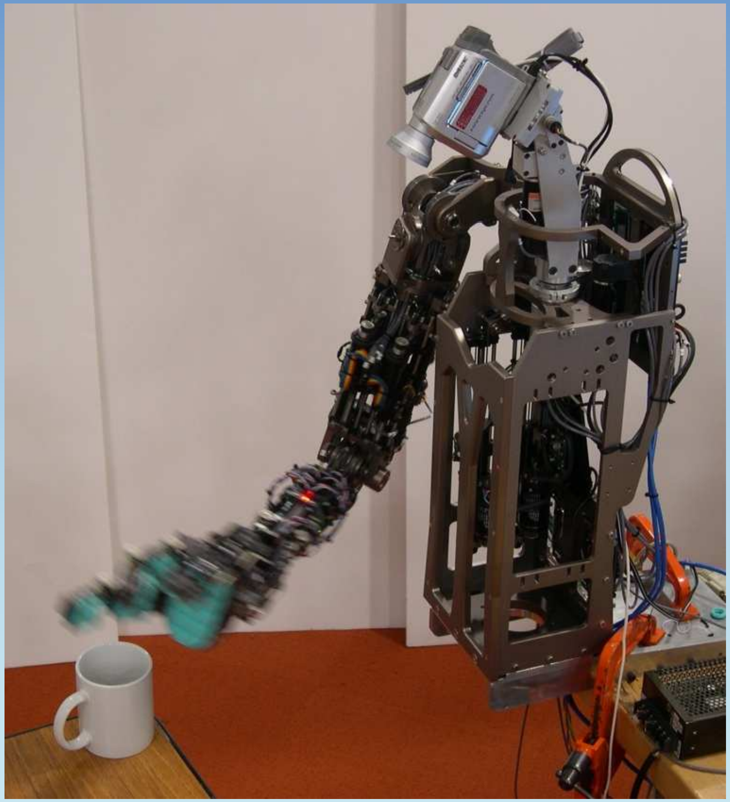
Haptic/tactile information is "directly" related to the task

Simpler than vision?

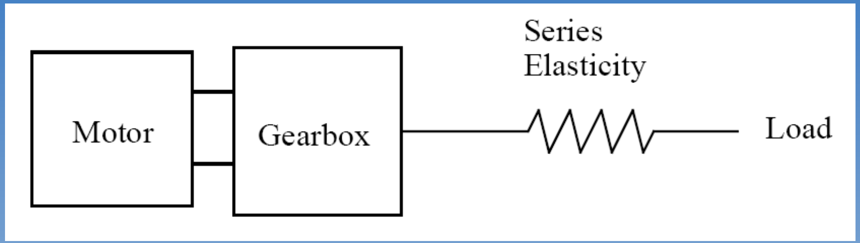
...but rarely investigated

A different robotic platform: Obrero

force sensing, tactile feedback, very limited vision



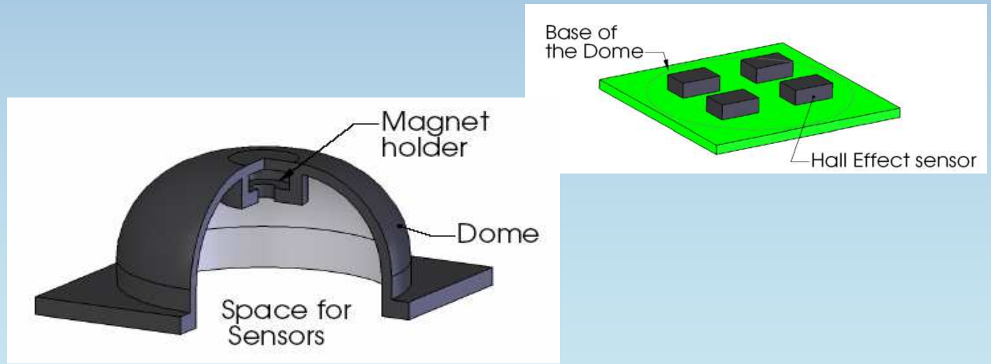
Actuation: series elastic actuators



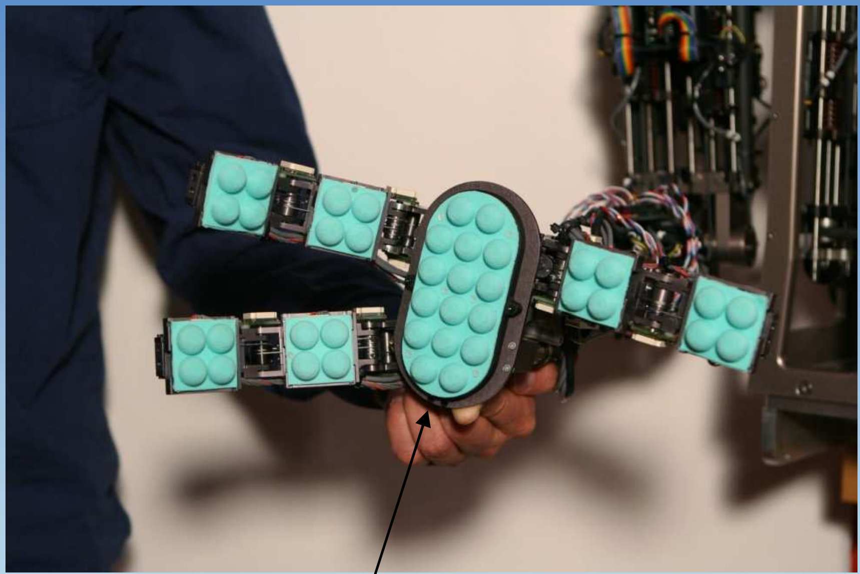
G. Pratt, M. Williamson (1995)

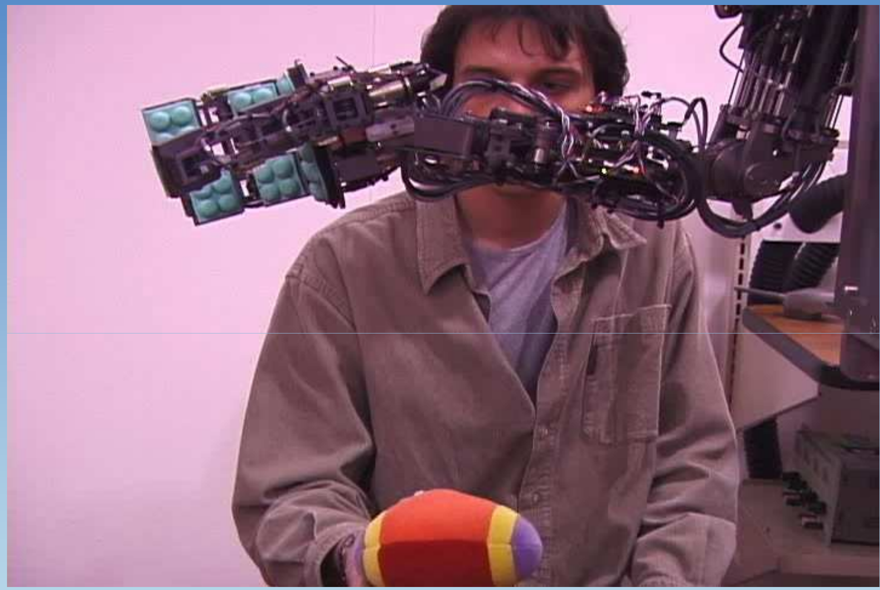
Tactile sensors

- Dome shaped, deformable
- Sensors favor compliance over resolution
- Friction, have you ever tried to grasp an object with a metallic hand?



Additional sensors: unreachable places





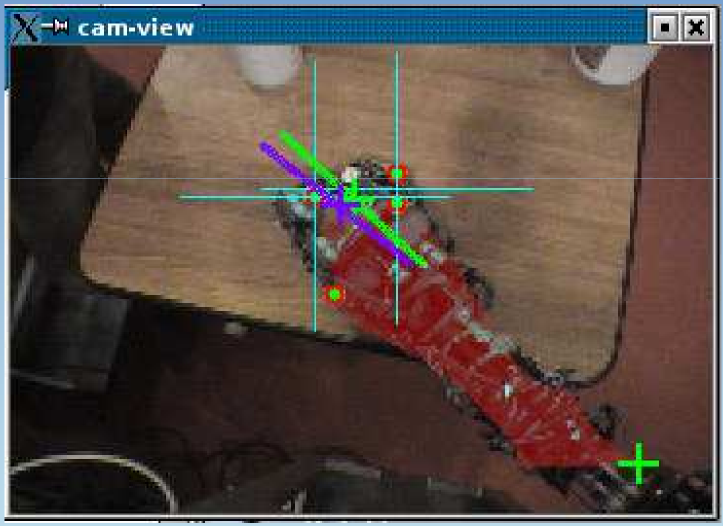
Additional sensors

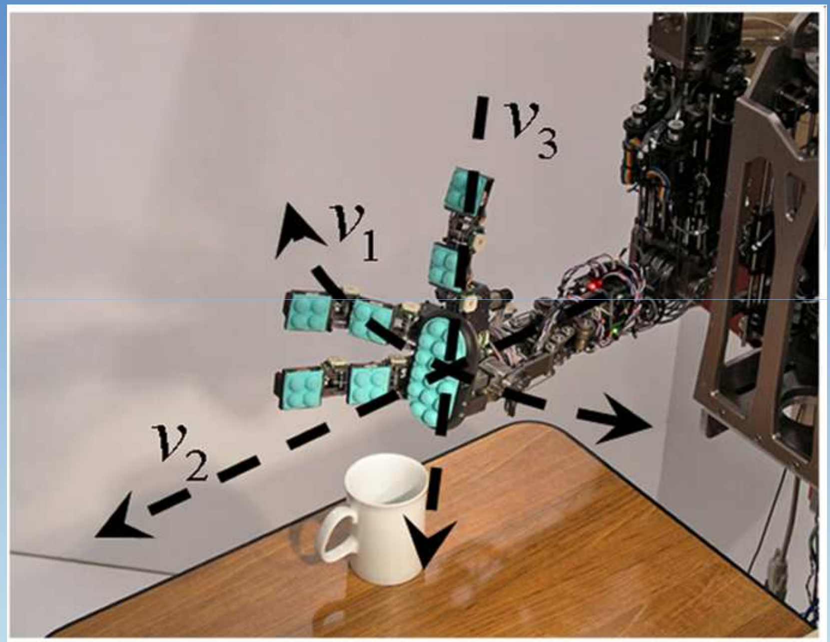
Visual attention and reaching

work with Charles Kemp

Grasping behavior

Exploration: adjust the position of the hand with respect to the object

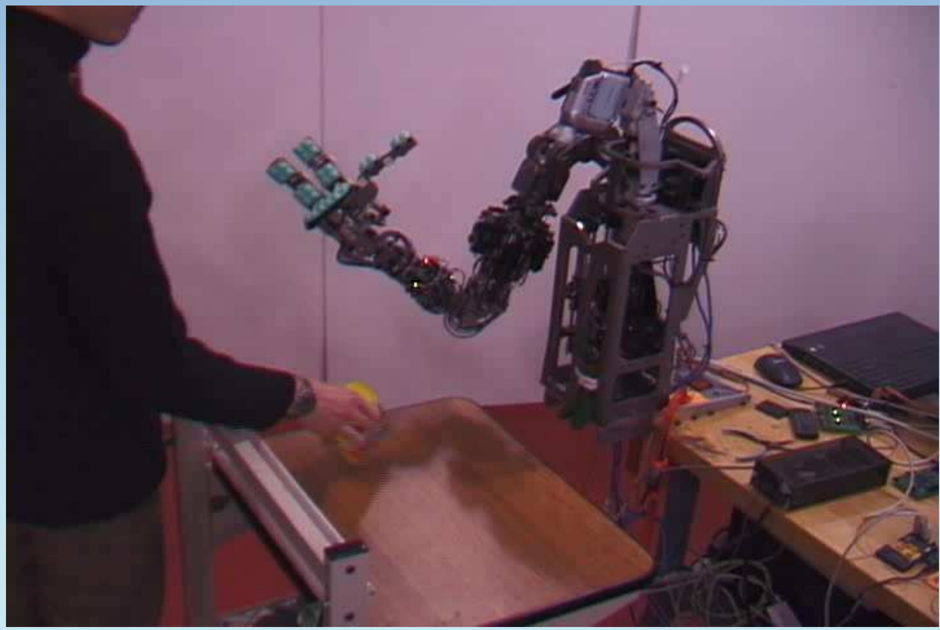


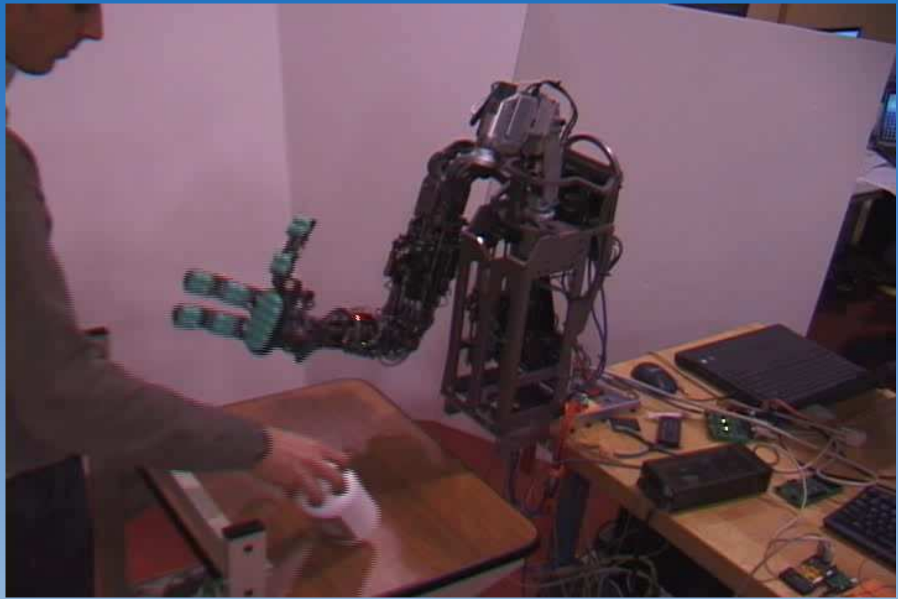


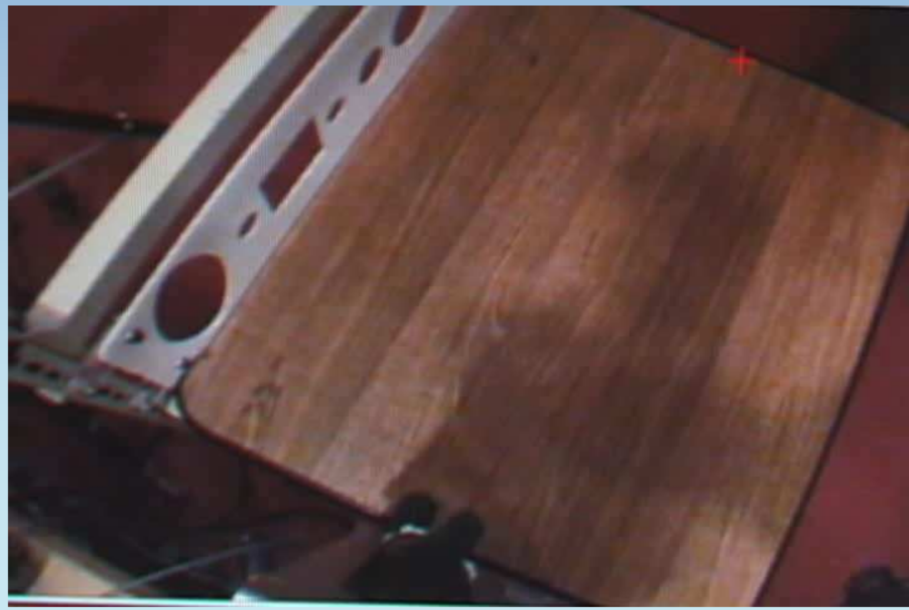
These explorative primitives are used by the following behaviors:
-*hovering behavior* moves the hand back and forth along v1
-*depth behavior* moves the hand toward the table, along v3
-*pushing behavior* moves the hand along v2 (towards the object)

These behaviors are modulated by tactile and force feedback:

- The *hovering behavior* stops and inverts the exploration when the object is detected
- The *depth behavior* is inhibited when contact is detected at the wrist
- The *pushing behavior* is activated when *both fingers* detect the object
- *Grasp behavior*: when contact is detected on the *palm,* close the fingers



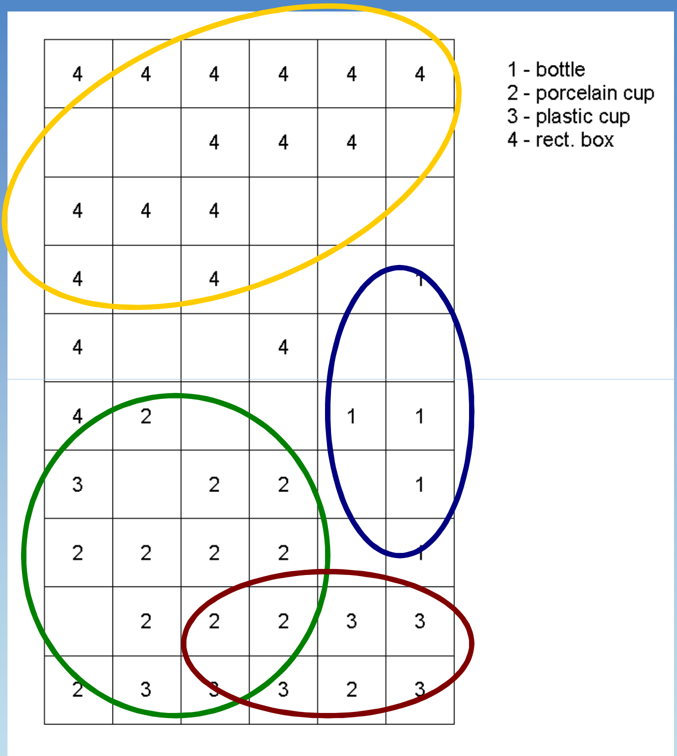




Some results...

Object	Weight	Trials	Failures
Plastic Bottle	265 [g]	22	0
Porcelain Cup	255 [g]	24	1
Box	240 [g]	34	2
Plastic Cup	220 [g]	24	4

Clustering of proprioceptive data



data is 13 encoders reading, for each grasp
Selft Organizing Map, left: most activated unit, right: U-matrix representation
In the SOM the same/similar objects activate neighboring units

L. Natale, E. Torres-Jara, Intl. Conference on Epigenetic Robotics, 2006.

To recap:

- Explore!
- Be soft, interact smoothly
- Sense and be reactive

- Controlled interaction produces rich sensory feedback that can be useful for learning

Tank you for your attention!