From natural to artificial cognitive systems

Giorgio Metta

LIRA-Lab
Dept. of Computer and Systems Sciences
University of Genoa, Italy

Dept. of Robotics, Brain and Cognitive Sciences
Italian Institute of Technology
Genoa
The Italian Institute of Technology
Robotics, Brain and Cognitive Sciences Department

Brain machine interface
- Design and fabrication of neural probes for in-vivo applications based on conductive and biocompatible nano-materials for electrodes

Human behavior
- Exploring the brain mechanisms at the basis of motor behavior and motor learning

Human behavior
- The development of new generation brain imaging systems: open magnet high resolution FMRI system targeted to brain research and natural behavior analysis (manipulation)

Brain machine interface
- The near-infrared spectroscopy: a new non-invasive technique to investigate brain functions in human subjects

Humanoid robotics
- Designing new generation, intelligent and interacting platforms

Humanoid robotics
- Toward next generation hybrid systems realized with soft materials and with cognitive abilities allowing operating in human environment
The tale of the Wright brothers (and three messages)

1. Reverse engineering:
   - Looking and copying bird flight → aircraft design
   - Reverse of reverse engineering → aerodynamics led to better understanding of bird flight (forward engineering)

2. Models:
   - Wright started from a previous model of Lilienthal (which was wrong) → but then they had (after 2 years) to produce their own models (and test them), they built a wind tunnel (very modern)!

3. Stability and control:
   - Separate models didn’t work well (either stability or control)
   - Discovered that the key to stability and control is by rolling → turn by rolling! Separate models don’t work, holistic approach is required and this was done by looking also at birds.
   - Understanding at the systems level

* From The Computational Neurobiology of Reaching and Pointing, Shadmehr and Wise, MIT press
...and the story goes

- Late 1903, first powered flight (35m, 10km/hour)
- 5 yrs later, 2 hours flight
- 8 yrs later, across North America
- 24 yrs later, New York to Paris
- 65 yrs later, three people to the moon
- Now, small seats and screaming infants
The three messages again!

1. Reverse and forward engineering
2. Mathematical modeling and empirical testing
3. Systems-level approach
Reverse engineering
Spinal behaviors

Walking behavior: cat rehabilitated to walk after complete spinal cord transection

Wiping reflex: an irritating stimulus elicits a wiping movement precisely directed at the stimulus location

Spinal sensorimotor coordination
(not simple reflexes after all)

- **Stimulus-response coordinate transformations**
  - **Evidence for:**
    - Combination of basic behaviors
    - Reach body parts that move respect to each other
    - Adapt and avoid obstacles
    - Use different sets of muscles
  - **Inverse dynamics (less compelling to me):**
    - Move $\rightarrow$ dynamics
    - Descending “kinematics” pathways (e.g. optic tectum, orienting behavior)
    - Walking, CPG’s, etc.
  - **Degrees of freedom problem:**
    - Evidence of synergies
    - Muscles activate together
    - Multi-joint muscles
Differences in the muscle activation

(a) Heel wipe
Heel wipe with obstacle

(b) Muscle activation

<table>
<thead>
<tr>
<th></th>
<th>No obstacle</th>
<th>Obstacle</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ST</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RA</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Differences in the muscle activation

TRENDS in Neurosciences
• Systematic stimulation of different regions of the spinal cord produced only a few types force fields (at least four).

Dissimilarity matrix of 49 force fields. Darker circles represent \( d(F_1,F_2) << 1 \)

The presence of only few units of motor output within the spinal cord is difficult to reconcile with the obvious ability of the nervous system to produce a wide range of movements.
Looking at muscles

H-reflex deviation from the mean (conductance of the spinal-muscle nerves)

Muscle anticipate the kinematics

Prone position, wrist ext/flexion (ECR, FCR muscles)

Supine position

As before, but different phase difference btw the kinematics and the muscular activation
Then compare with actual action

Which was 54° on average

Which was 112° on average

The movement doesn’t change

EMG signals from the FCR and ECR muscles
...but let me take a little leap forward

SHORT COMMUNICATION
Speech listening specifically modulates the excitability of tongue muscles: a TMS study

Luciana Fedigia, Lella Creighera, Giovanni Buscione and Giacomo Rizzolatti
1Dipartimento di Scienze Biomediche e Terapie Avanzate, Sezione di Fisiologia Umana, Università di Ferrara, via Fossato di Mortara 17-19, 44100 Ferrara, Italy
2Istituto di Fisiologia Umana, Università di Parma, via Volta 99, 43100 Parma, Italy

Keywords: mirror neurons, motor-evoked potentials, motor system, motor theory of speech perception
The experiment

- Listening: three categories of stimuli (words, pseudo-words, bi-tonal sounds).
- Two phonemes 'rr' requires strong tongue tip movement, 'ff' requires slight tongue tip movement.
- TMS of the under-threshold motor cortex.
- Recording of the MEP (motor-evoked potential) from the tongue muscles.
## Examples of word/pseudo-words

**Labiodental fricative consonant, ‘\(r^r\)’**

<table>
<thead>
<tr>
<th>Words</th>
<th>Pseudo-words</th>
</tr>
</thead>
<tbody>
<tr>
<td>birra (bier)</td>
<td>berro</td>
</tr>
<tr>
<td>carro (cart)</td>
<td>fira</td>
</tr>
<tr>
<td>cirro (cirrus)</td>
<td>forro</td>
</tr>
<tr>
<td>farro (spelt)</td>
<td>furra</td>
</tr>
<tr>
<td>ferro (iron)</td>
<td>marro</td>
</tr>
<tr>
<td>mirra (myrrh)</td>
<td>merro</td>
</tr>
<tr>
<td>morra (morra)</td>
<td>parro</td>
</tr>
<tr>
<td>porro (leek)</td>
<td>perro</td>
</tr>
<tr>
<td>serro (greenhouse)</td>
<td>vorro</td>
</tr>
<tr>
<td>terra (ground)</td>
<td>vurro</td>
</tr>
</tbody>
</table>

**Lingua-palatal fricative consonant, ‘\(l\)’**

<table>
<thead>
<tr>
<th>Words</th>
<th>Pseudo-words</th>
</tr>
</thead>
<tbody>
<tr>
<td>baffo (moustache)</td>
<td>biffo</td>
</tr>
<tr>
<td>beffa (hoax)</td>
<td>ciffo</td>
</tr>
<tr>
<td>buffo (funny)</td>
<td>leffa</td>
</tr>
<tr>
<td>ceffo (snoat)</td>
<td>meffa</td>
</tr>
<tr>
<td>coffa (crow’s nest)</td>
<td>paffo</td>
</tr>
<tr>
<td>goffo (clumsy)</td>
<td>peffo</td>
</tr>
<tr>
<td>muffa (mold)</td>
<td>poffa</td>
</tr>
<tr>
<td>puffo (smurf)</td>
<td>seffa</td>
</tr>
<tr>
<td>tuffo (dive)</td>
<td>viffa</td>
</tr>
<tr>
<td>zaffo (plug)</td>
<td>voffo</td>
</tr>
</tbody>
</table>
Results (in short)

- Modulation due to meaning
- Effect due to phoneme
“In all communication, sender and receiver must be bound by a common understanding about what counts; what counts for the sender must count for the receiver, else communication does not occur. Moreover the processes of production and perception must somehow be linked; their representation must, at some point, be the same.”

[Alvin Liberman, 1993]
More central

More peripheral

Even more peripheral (multiple body parts)
Grasping neurons
F5 canonical neurons

A: Observation + action

B: Observation only

1 s

20 sp/s
Mirror Neurons

The neuron is activated by “seeing” someone else’s hand performing a manipulative action and while the monkey is performing the same action.

Models

Total Drag = \( C_{dp} \times (\frac{1}{2} q V^2 \times SA) + \left( K \frac{w^2}{(\frac{1}{2} q V^2)} \times SA \right) \)

Where:
- \( C_{dp} \): Parasitic Drag Coefficient = \( \frac{D}{q \times SA} \)
- \( D \): Actual Drag
- \( q \): Dynamic Pressure
- \( SA \): Surface Area
- \( K \): Fluid Density
- \( V \): Velocity
- \( w \): Weight

\( K \) = constant of proportionality, varied with planform shape.

(not real useful when dealing with a golf ball :-))
Controlled Stiffness, viscosity

Feedback motor command

Realized trajectory

Desired trajectory +

Gain

Controlled object

Desired trajectory

Stiffness, viscosity

Feedback motor command

Realized trajectory

Inverse model

Feedforward motor command

Controlled object

Effect of delays

Make feedback control either poor or unstable altogether

- Engineering control systems
  - Delays: 500μs
  - Movement duration: seconds
  - Gain of the controller: can be made high

- Humans
  - Delays:
    - 20-50ms (spinal)
    - 150-250ms (vision)
  - Movement duration: 150-500ms
  - Gain of the controller: stiffness and muscle viscoelastic properties (comparatively low)

- BTW: maintaining fast control loops is not an easy feat
Building the internal models

• Rote-learning would be unpractical:
  - Too many possible actions (dof) for the available number of neurons (although they’re quite a lot!)

• Generalizing past experience:
  - Past experience is bound (unfortunately) to “represent” only a portion of the whole state space

• Developing and extending the control structure to new behaviors
  - Sequencing and combining primitive behaviors appropriately

• Predicting the future course of action
  - It might turn out to be useful!
From: Lackner JR, Dizio P. Gravitoinertial force background level affects adaptation to Coriolis force perturbations of reaching movements.
Internal models for interception

Fig. 1: Experimental setup: (right) a schema of the game played by subjects; (center) error measure; (left) a sketch of an experiment. 28 subjects participated in the experiment.

Stimuli (3 conditions)

With Alessandra Sciutti, Francesco Nori, Thierry Pozzo
Some results

Experiment with motor information

Mean error (distance between paddle and ball arrival point) in the three different conditions (control, test, upward direction and test, downward direction).
The wind tunnel
The iCub: quick summary

The iCub is the humanoid baby-robot being designed as part of the RobotCub project

- The iCub is a full humanoid robot sized as a three and half year-old child.
- The total height is 104cm.
- It has 53 degrees of freedom, including articulated hands to be used for manipulation and gesturing.
- The robot will be (once the software is done) able to crawl and sit and autonomously transition from crawling to sitting and vice-versa.
- The robot is GPL/FDL: software, hardware, drawings, documentation, etc.
Degrees of freedom

• 53: 9 in each hand
• Sensors: position, torque, temperature
• And also: cameras, microphones, gyroscopes, linear accelerometers
• For the future: tactile sensors, skin...
  - Low-resolution:
    • Distributed many sensing points
  - Fingertips:
    • Localized, high-resolution
A few examples

Custom design

Hand sensor sampling PIC-based card

Force/torque sensor fitted into the sensor

Fingertip sensorization

Design and documentation

Wired with 25micron coated wires
Current status

Design

Reality
Facial expressions

So, tell me about your problem
The soul of a new machine
More integration
... and robot Yoga
Skin, architecture by Marco Maggiali

With Giorgio Cannata (Univ. of Genoa)
Prototype covered by silicon rubber
Fingertip sensors

CAD/concept

Prototype

Some testing

Fingernail + microphone

By Alexander Schmitz, Marco Randazzo, Marco Maggiali and Lorenzo Natale
The hands
Joint-level torque sensing

Existing parts

FEM analysis of deformation

Changes (under implementation)

By Alberto Parmiggiani and collaborating with Nikos Tsagarakis
High-res tactile sensors

Passivated metal Exposed Au

Markers (METAL) for Piezo-polymer foil alignment

Contact (transistor gates) + PVDF

FETs

Testing the piezo-film deposition

Cross-talk btw. taxels

By Ravinder Dahiya, Maurizio Valle (Unige) & Leandro Lorenzelli (Trento)
Level 2 APIs: Prospective Action Behaviors

Coordinated operation: Ontogenic Development

Level 1 APIs: perception/action behaviors

Innate perception/action primitives
loose federation of behaviors

Level 0 APIs: data acquisition & motor control

Multiple YARP processes
Running on multiple processors

Gbit Ethernet

HUB pc104

DSP DSP .......... DSP DSP

Sensors & Actuators

Based on phylogenic configuration

Cognitive Architecture

Software Architecture

iCub Embedded Systems
more work in progress

- Ryo Saegusa: sensory prediction
- Andrew Dankers: models of vision
- Matteo Fumagalli: force control
- Boris Duran: dynamical systems control
- Lorenzo Jamone: grasping
- Serena Ivaldi: optimal control
- Massimiliano Izzo: internal models
Lots of people

- Lorenzo Natale, Francesco Nori: Software, testing, calibration
- Marco Maggiali, Marco Randazzo: firmware, DSP libraries, tactile sensing
- Francesco Becchi, Paolo Pino, Giulio Maggiolo, Gabriele Careddu: design and integration
- Gabriele Tabbita, Walter Fancellu: assembly
- Nikos Tsagarakis, William Hinojosa: legs and spine, force/torque sensors
- Bruno Bonino, Fabrizio Larosa, Claudio Lorini: electronics and wiring
- Luciano Pittera: wiring
- Mattia Salvi: CAD maintenance
- Alberto Zolezzi: managing quotes, orders and spare parts
- Giovanni Stellin: hand
- Ricardo Beira, Luis Vargas, Miguel Praca: design of the head and face
- Paul Fitzpatrick & Alessandro Scalzo: software middleware
- Alberto Parmiggiani: joint level sensing
- Alexander Schmitz: fingertips
- Nestor Nava: small Harmonic Drive integration
- Ravinder Dahiya: FET-PVDF tactile sensors
- Lorenzo Jamone: fingertips
- Daniel Roussy: construction
- Ludovic Righetti: simulation and initial torque specification
Other tunnels
Experimental setup...
The initial idea...
Objects come to existence because they are manipulated.

- Fixate target
- Track visual motion…
- (…including cast shadows)
- Detect moment of impact
- Separate arm, object motion
- Segment object

Which edge should be considered?

Maybe some cruel grad-student glued the cube to the table by Paul Fitzpatrick.

Color of cube and table are poorly separated.

Cube has misleading surface pattern.
Gesture “vocabulary”

- pull in
- side tap
- push away
- back slap
Exploring an affordance: rolling

A toy car: it rolls in the direction of its principal axis

A bottle: it rolls orthogonal to the direction of its principal axis

A toy cube: it doesn’t roll, it doesn’t have a principal axis

A ball: it rolls, it doesn’t have a principal axis
Forming object clusters
Into object affordances...

Bottle, “pointiness”=0.13

Rolls at right angles to principal axis

Car, “pointiness”=0.07

Rolls along principal axis

Cube, “pointiness”=0.03

Ball, “pointiness”=0.02

difference between angle of motion and principal axis of object [degrees]
The geometry of poking

- **Backslap**
  - Estimated probability vs. direction of movement [deg]

- **Pull in**
  - Estimated probability vs. direction of movement [deg]

- **Side tap**
  - Estimated probability vs. direction of movement [deg]

- **Push away**
  - Estimated probability vs. direction of movement [deg]
Behavior: poking according to affordance

Look for affordance

Search rotation

Identify and localize object

Previously-poked prototypes

Look for action to satisfy affordance
Behavior: poking according to affordance
Understanding a foreign manipulator

Object, goal connects robot and human action
Interpreting observations

“The robot could actually tell this was a side tap”

Initial position

Final position

A foreign manipulator (human) pokes an object
The direction of movement is compared with the object affordance
Interpreting observations

<table>
<thead>
<tr>
<th>Demonstration by human</th>
<th>Invoking the object’s natural rolling affordance</th>
<th>Going against the object’s natural rolling affordance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mimicry in similar situation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mimicry when object is rotated</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Mimicry
Data from human grasping

2 cameras → Frame grabbers → Images → To disk

Cyber-glove → RS232 → Tracker → RS232 → Other sensors → To disk

Tactile sensors
Bayesian classifier

\{G_i\}: set of gestures
F: observed features
\{O_k\}: set of objects

\( p(G_i|O_k) \): priors (affordances)
\( p(F|G_i,O_k) \): likelihood to observe \( F \)

\[
p\left( G_i \mid F, O_k \right) = p\left( F \mid G_i, O_k \right) p\left( G_i \mid O_k \right) / p(F \mid O_k)
\]

\( \hat{G}_{MAP} = \arg \max_{G_i} \left( G_i \mid F, O_k \right) \)
Two types of experiments

Learned by backpropagation ANN
Estimation

- \( p(G_i|O_k) \): affordances, by counting, estimated on the whole database
- \( p(F|G_i,O_k) \): EM algorithm on the parameters of a mixture of Gaussians (from Matlab implementation)
- VMM: Neural network, sigmoidal activation units, linear output, trained on the whole database
Role of motor information in action understanding

Object affordances (priors)

Visual space  Motor space

Classification (recognition)

Grasping actions


RobotCub.org
Some results…

<table>
<thead>
<tr>
<th></th>
<th>Exp. I (visual)</th>
<th>Exp. II (visual)</th>
<th>Exp. III (visual)</th>
<th>Exp. IV (motor)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Training</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># Sequences</td>
<td>16</td>
<td>24</td>
<td>64</td>
<td>24</td>
</tr>
<tr>
<td># of view points</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Classification  rate</td>
<td>100%</td>
<td>100%</td>
<td>97%</td>
<td>98%</td>
</tr>
<tr>
<td># Features</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td># Modes</td>
<td>5-7</td>
<td>5-7</td>
<td>5-7</td>
<td>1-2</td>
</tr>
<tr>
<td><strong>Test</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># Sequences</td>
<td>8</td>
<td>96</td>
<td>32</td>
<td>96</td>
</tr>
<tr>
<td># of view points</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Classification  rate</td>
<td>100%</td>
<td>30%</td>
<td>80%</td>
<td>97%</td>
</tr>
</tbody>
</table>
Speech again

With Michele Tavella & the CONTACT project team
System-level approach
Cognition: a process by which a system achieves behaviour that is

- robust
- adaptive
- anticipatory
- autonomous

Entails embodied perception, action, and interaction
Our approach

Guiding Philosophy

- Cognition cannot be hand-coded
- Is necessarily the product of a process of embodied development
- Initially dealing with immediate events
- Increasingly acquiring a predictive capability

Cognition and perception are functionally-dependent on the richness of the action interface
Our Approach (contd.)

Emergent embodied cognitive systems:

- Given a rich set of innate action and perception capabilities
- Develop over time an increasing range of cognitive abilities
- Recruiting ever more complex actions
- Achieving an increasing degree of prospection (and, hence, adaptability and robustness)
RobotCub Cognitive Architecture

Grounded in neuroscience and psychology

Rooted in action-dependent perception

Focused on adaptive & prospective capabilities

Designed to facilitate development

Cognitive architecture = (RobotCub) Phylogeny
In spite of the wiring problems…