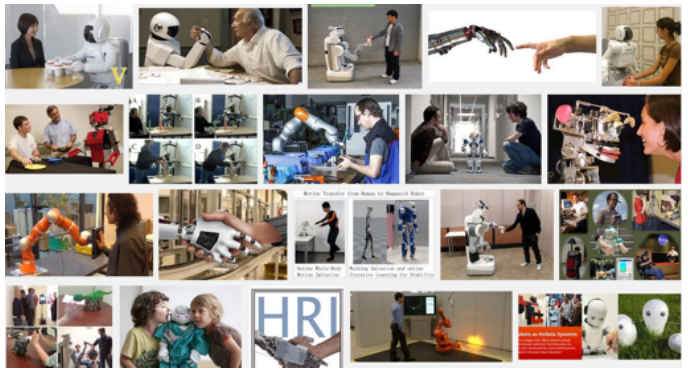
Spoken Human Robot Interaction

Daniele Nardi,
Andrea Vanzo
Dipartimento di Ingegneria
Informatica, Automatica e Gestionale
nardi@dis.uniroma1.it
http://rococo.dis.uniroma1.it

SAPIENZA
UNIVERSITÀ DI ROMA

Robots and Humans



Robots and Humans

Service robots

Plan of this introduction

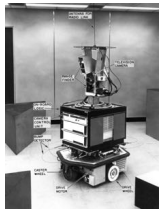
- Human Robot Interaction
  - Social Robots
  - Some examples
- Talking with robots
  - Command Interpretation
  - Dialog

What is HRI?

"*Human-robot interaction is the field of study dedicated to understanding, designing, and evaluating robotic systems for use by or with humans*"

Dynamic Interaction (Goodrich, Schultz 2007, HRI)

Some temporal references



- 1992 – First IEEE Int. Symp. on Robot & Human Interactive Communication (RoMan)
- 1997 AAAI Hors d'oeuvres Anyone?
- 1999 IEEE RAS TC on Human-Robot Interaction & Coordination
- 2000 IEEE Humanoids
- 2006 ACM Int. Conf. Human-Robot Interaction
- 2006 RoboCup@Home
- 2009 Int. Conf. Social Robotics

HRI vs HCI

- Robots have (physical) bodies
- Robots act in the real world
- Robots are perceived as living entities
- Human-robot interaction is asymmetric (robots have not the same cognitive skills of humans)
- HRI is bidirectional (robots are not passive entities like computers!)

HRI is not a special case of HCI

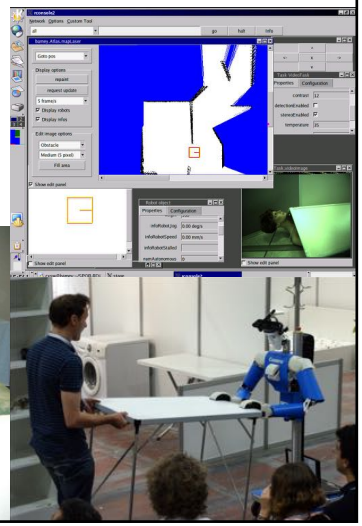
HRI: multidisciplinarity

HRI brings together a **variety of fields**, including:

- **engineering** (electrical, mechanical, industrial, and design),
- computer science (**human-computer interaction**, **artificial intelligence**, **robotics**, natural language understanding, and computer vision),
- **social sciences** (psychology, cognitive science, communications, anthropology, and human factors),
- **humanities** (ethics and philosophy).

HRI: settings

- Remote (Telepresence)
- Proximate (Co-location)
- Physical



Possible inputs (for the robot)

- Hand-held devices
- Speech
- Sound
- Touch

- Temperature
- Olfaction

People

- Locomotion
- Gestures
- Race? Gender?

Head

- Gaze
- Facial Expressions

Possible inputs (vision)

– Face detection / tracking / recognition (including expressions)

– Person detection / tracking / recognition

Possible outputs (of the robot)

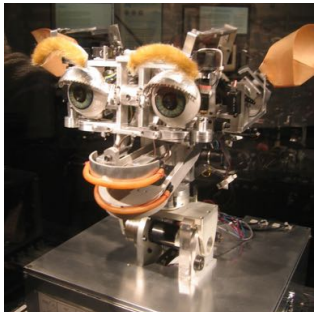
Body
- Position
- Speed

Head
- Turning
- Eye motion
- Facial expressions

Arms
- Grab/Hand objects
- Gestures
- Shake hands

- Speech
- Sound
- Lighting
- Smell

Possible outputs (of the robot)



HRI: application domains

ANY ROBOT !!!

Service Robotics and robots in our homes is one of the most compelling cases.

Social Robots

(e.g. Robots as dietary assistants)

"Intelligent" Robots

- at least: `not(stupid) robots`
- Gap between user expectations and robot functionality.
- Why?
  - limited capabilities of perception systems
  - difficulty of communicating with humans
  - ability to acquire, maintain and use knowledge

Our Approach: Symbiotic Autonomy

The concept

Enable a robot to get help from humans in the same fashion a person might be helped by another individual.

- Symbiotic autonomy[1] and symbiotic robotics[2]
- Exploit HRI to overcome the limitations of the robot

[1] *An effective personal mobile robot agent through symbiotic human-robot interaction*. Rosenthal, Biswas and Veloso.

[2] *Symbiotic robotic systems: Humans, robots, and smart environments*. Coradeschi and Saffiotti.

Are humans willing to help?

In the context of **Symbiotic Autonomy**

evaluate the **Collaboration Attitude** of humans

varying different factors:

- **Activity context**
- **Proxemics**
- **Gender**
- **Height**

Our approach: Small is beautiful

The concept

Acquire very detailed knowledge about the operational requirements through a continuous interaction

1. Environment (Semantic Map)
2. Tasks
3. User

Semantic Mapping

Herzberg & Nuchter, 2008

A semantic map for a mobile robot is a map that contains, in addition to spatial information about the environment, assignments of mapped features to entities of known classes.

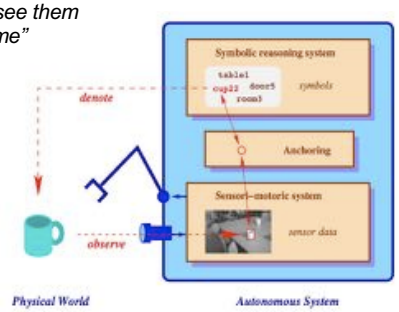
Semantic maps allow the robot to…

- …perform reasoning over environments, objects and properties
- …communicate with humans, understanding complex commands
- …perform complex tasks

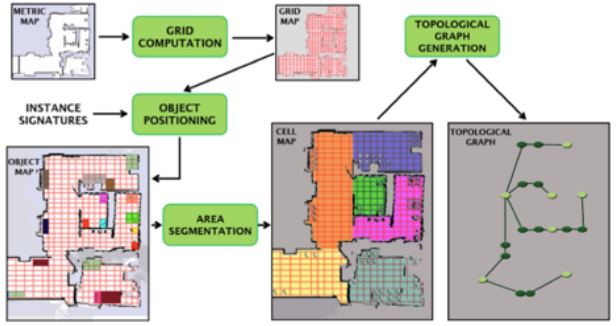
Symbol Grounding

"For things to exist there are two essential conditions, that a man should see them and be able to give them a name"
(*Saramago, 1995*)

- Symbol grounding (*Harnad*, 1990)
- Anchoring (*Saffiotti*, 1994)

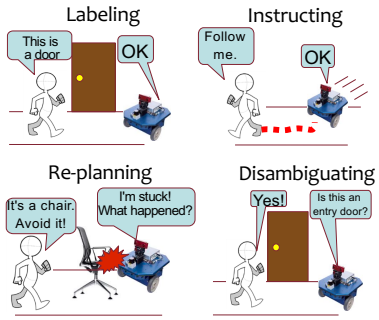


Cell Map and Topological Graph Generation



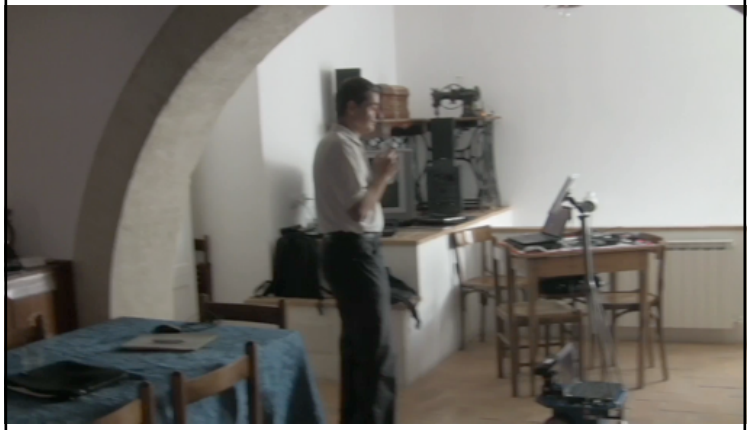
Human Augmented Methods

- Human-in-the-loop
- Human-Robot uni-modal or multi-modal interaction
- Integration with perception
- Clarification dialogues



Diosi *et al.*, Zender *et al.*, Kruijff *et al.*, Nieto-Granda *et al.*, Pronobis and Jensfelt, Randelli *et al.*

Basic Knowledge Acquisition (home)



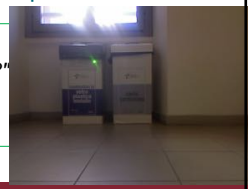
Example dialogs

Knowledge from the map for disambiguation

User: *"Go to the socket.".*
Robot: *"There are many sockets. Which one do you mean?"*
User: *"The one close to the emergency door."*
Robot: *"OK. I am going to the socket close to the emergency door."*

Knowledge from the map combined with perception

User: *"Throw the paper in the garbage bin.".*
Robot: *"There are many garbage bins. Which one?"*
User: *"The blue one."*
Robot: *"OK. I am going to the blue garbage bin."*



RoboCup@Home

- Development of Domestic Service Robots
  - Complex Integrated Systems
  - Large variety of tasks
  - Human-Robot Interaction

HRI and semantic map

Take the screwdriver on platform 1 …

Human Robot Interaction
within
RoboCup@Work scenarios

Learning Tasks from the User

- Robots face difficulties not envisioned by their developers at programming time
- Tasks specialized and adapted to the needs of specific users and environments

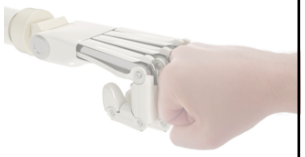
1. Learning parametric task descriptions that are defined as a combination of primitive actions
2. Learning primitive actions (e.g. handover)

Teaching parametric tasks

Bring @object to @location

Spoken Human-Robot Interaction

- Design robotic systems that exhibit a natural and effective interaction with users
  - spoken language
  - guiding touch
  - gestures
  - gaze
  - visual demonstration



- Natural Language is a ***natural*** way of communicating

Not really convinced…

- Why don't we use a dedicated UI?
  - Touch
  - Artificial Language
  - Gestures
- Because an artificial UI requires training and NL is the most efficient and natural way of communicating
  - Elder
  - Kids
  - Lazy people

Challenges in Spoken Human Robot Interaction

- The input signal (sound) is highly noisy (unless very constrained)
- Natural Language is inherently ambiguous
  - Natural Language (English or Italian are highly ambiguous)
    - Syntactic: Jordan could write more profound essays – "more" what?
    - Semantic: *Prostitutes appeal to the Pope* (real life headline) - appeal may mean both "to be liked by someone" and "to seriously request for help"
- Natural Language Interpretation is highly dependent on context

Interpreting commands – increasing complexity

could you please find the remote controller of the television and bring it to the kitchen

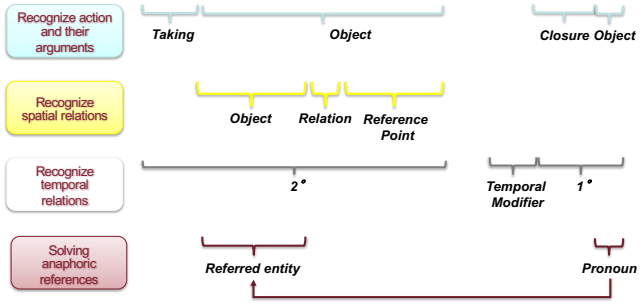
```
LOCATING(phenomenon:"the remote controller of the television")#
BRINGING(theme:"it",goal:"to the kitchen")
```

go to the bathroom, take the soap, and bring it to the side-table

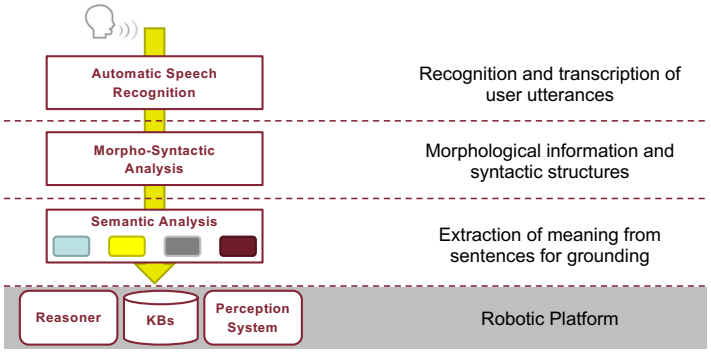
```
MOTION(goal:"to the bathroom")#TAKING(theme:"the
soap")#BRINGING(theme:"it",goal:"to the side-table")
```


Interpreting commands

"take the bottle on the table but first open it"



SLU for interpreting commands: typical pipeline



SLU for interpreting commands: approaches

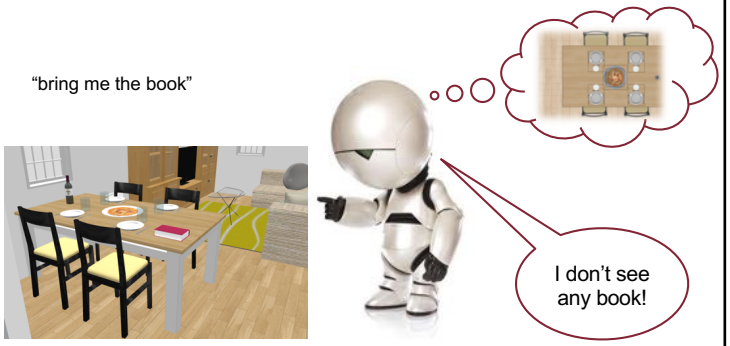
Grammar-based	Data-driven
Hand-crafted grammar development	Learning from data
Parsing	Statistical models rather than grammars
Limited coverage (structure+lexicon)	Generalization (structure+lexicon)
High performance on specific domains	Over/Under-fitting risk

Grammar-based vs. data-driven

- **S4R[1]** (Speaky 4 Robots)
  - Grammar-based
  - Language is domain-specific
  - Interpretation is application-dependent
- **LU4R[2]** (adaptive spoken Language Understanding 4 Robots)
  - Data-driven
  - Language is domain-driven
  - Interpretation is context-sensitive

Connection with perception

"bring me the book"

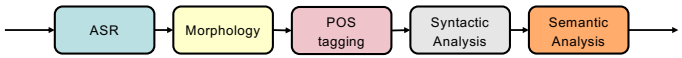


Command Interpretation vs. Spoken Dialog Systems

- **Command interpretation**
  - Communication is an atomic processing of sentences
  - Each sentence is independent
  - Linguistic/Physical context irrelevant
  - The SLU process is enough to carry out the task
- **Dialogic interactions**
  - Communication is a sequence of turns (sentences)
  - Each sentence depends (at least) on the previous one
  - The dialogue state influences the flow
  - Implies: dialogue manager, SLU, dialogue state tracking, natural language generation, …

Plan of the lectures

- Command Interpretation
  - ASR (1)
  - Morpho-syntactic analysis (1,2)
    - Morphology
    - POS tagging
    - Syntactic Parsing
  - Semantic Analysis (3)



- Dialogue Management (4)
  - Rule-based
  - Statistical DM

NLP has many other application contexts

- Google Assistant
- Siri
- Smart cars

- Machine translation

More in the AI course and then on Natural Language Processing

Summary of introduction

- ***HRI is a critical feature for (service) robots***
- ***Speech is a powerful HR interaction channel***
- ***Spoken HRI is a an ideal framework to address the integration of symbolic and numerical reasoning***