

HRI-2014 Workshop on Attention Models in
Robotics: Visual Systems for Better HRI

Attentional Top-down Regulations in a Situated Human-Robot Dialogue

Alberto Finzi

DIETI, Università degli Studi di Napoli "Federico II"

Riccardo Caccavale, Alberto Finzi, Lorenzo Lucignano, Silvia Rossi, Mariacarla Staffa



UNIVERSITA' DEGLI STUDI DI
NAPOLI FEDERICO II

Introduction

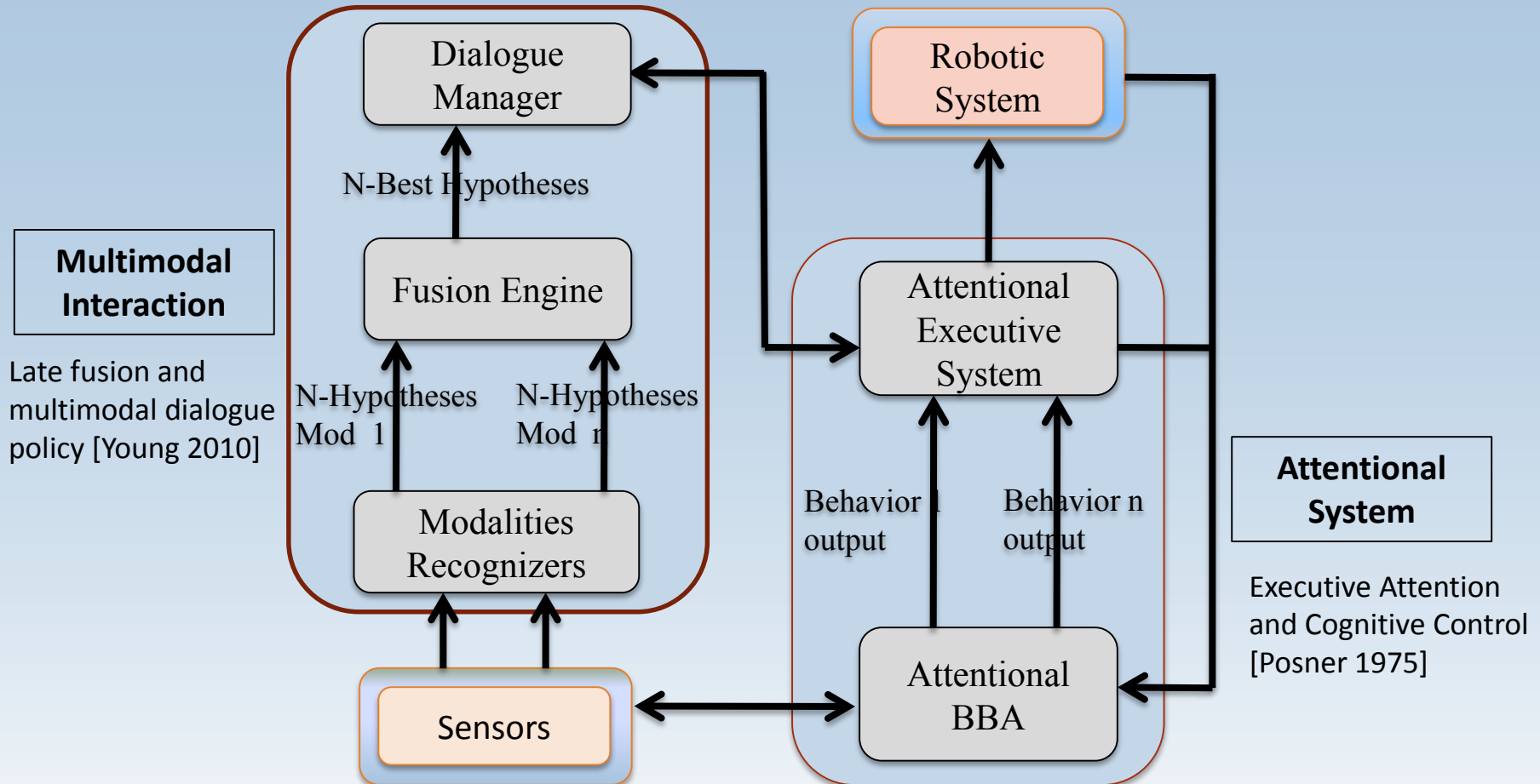
Multimodal interaction, Dialogue Manager, Attentional System

- **Integrated framework for multimodal HRI regulated by an attentional system:**
 - The interaction between humans and robots can be modeled as a multimodal dialogue flow, involving speech, gestures, gaze orientation, etc.
 - Attentional mechanisms can orient and focus the robotic perceptive, cognitive, executive processes during the interaction.
- **Attentional System:**
 - Executive attention and cognitive control [Posner 1975, Shallice 2000]
 - Bottom-up regulations (environment and internal stimuli)
 - Top-down regulations (structured tasks)
- **Attentional System and Dialogue Manager integration:**
 - The multimodal interaction policy is regulated and integrated by the Attentional System with contextual and task-related contents

Attentional Multimodal HRI

Attentional system and multimodal dialogue management

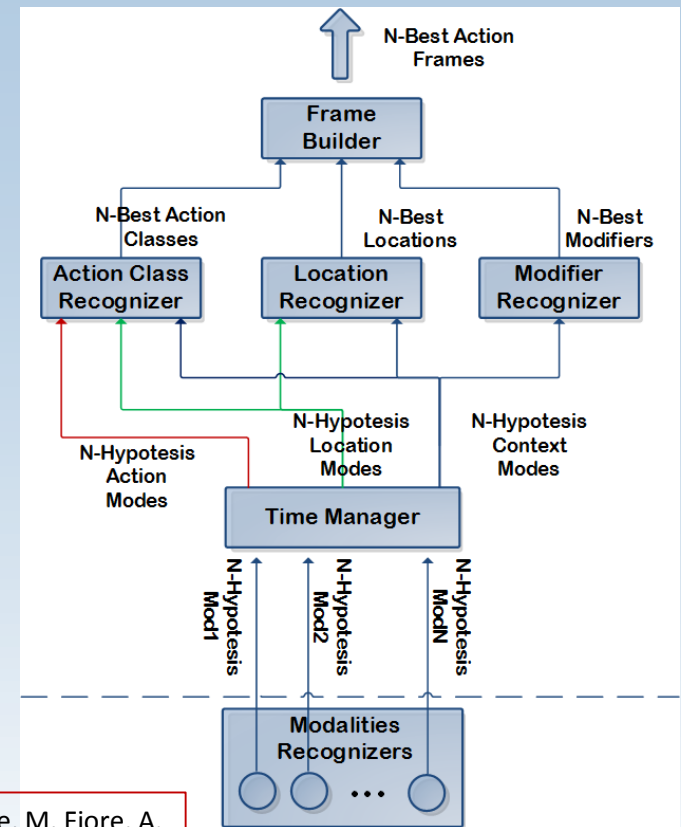
- Integrated Framework for Attentional Multimodal HRI:



Multimodal Interaction Module

Architecture

- **Multimodal interaction:**
 - Single-channel information can be ambiguous;
 - Ambiguities are resolved in cascade in the upper layers of the system;
 - Each layer provides the next layer with a list of possible interpretations;
 - Late fusion approach.
- **Classification of single modalities:**
 - Gesture, speech, etc.
- **Time Manager:**
 - Synchronization (temporal windows)
- **Action Classifier:**
 - User action recognition
 - Contextual weight
- **Location Classifier:**
 - Target of the action
- **Modifier Recognition:**
 - Execution modality
- **Frame Builder:**
 - N-best list of hypothesis



An Extensible Architecture for Robust Multimodal Human-Robot Communication, S. Rossi, E. Leone, M. Fiore, A. Finzi and F. Cutugno, in Proceedings of IROS 2013

Gesture Recognition

Classification of single modalities

■ Gestures:

■ Features:

- 3D coordinates of the shoulder, elbow, and hand joints.
- 3D angles between the shoulder and the elbow.
- 3D angles between the elbow and hand.
- Open, closed, pointing.
- Palm hand w.r.t. camera (boolean)

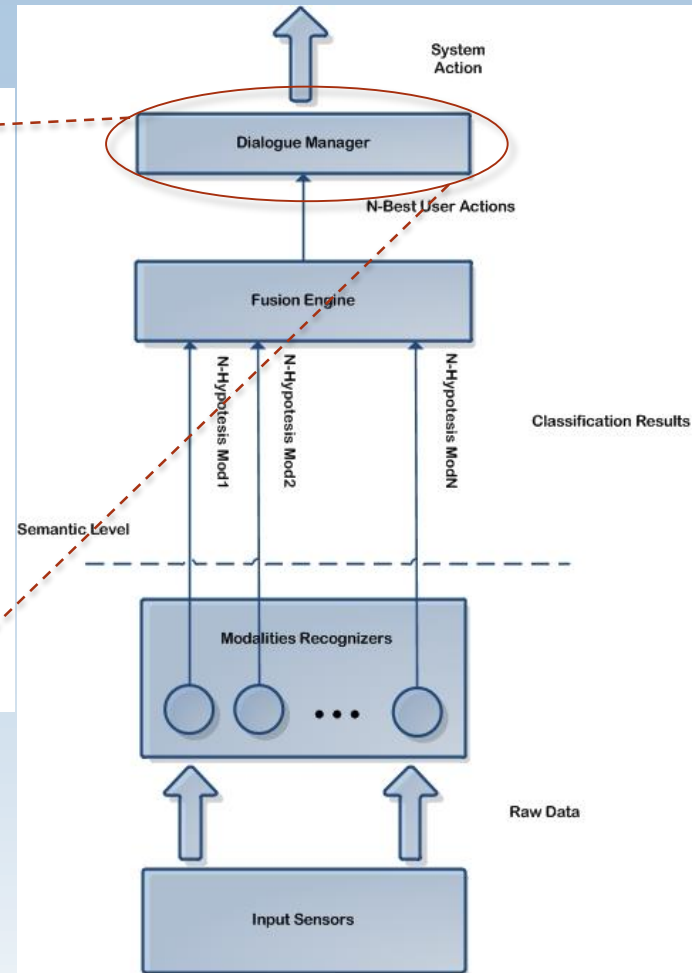
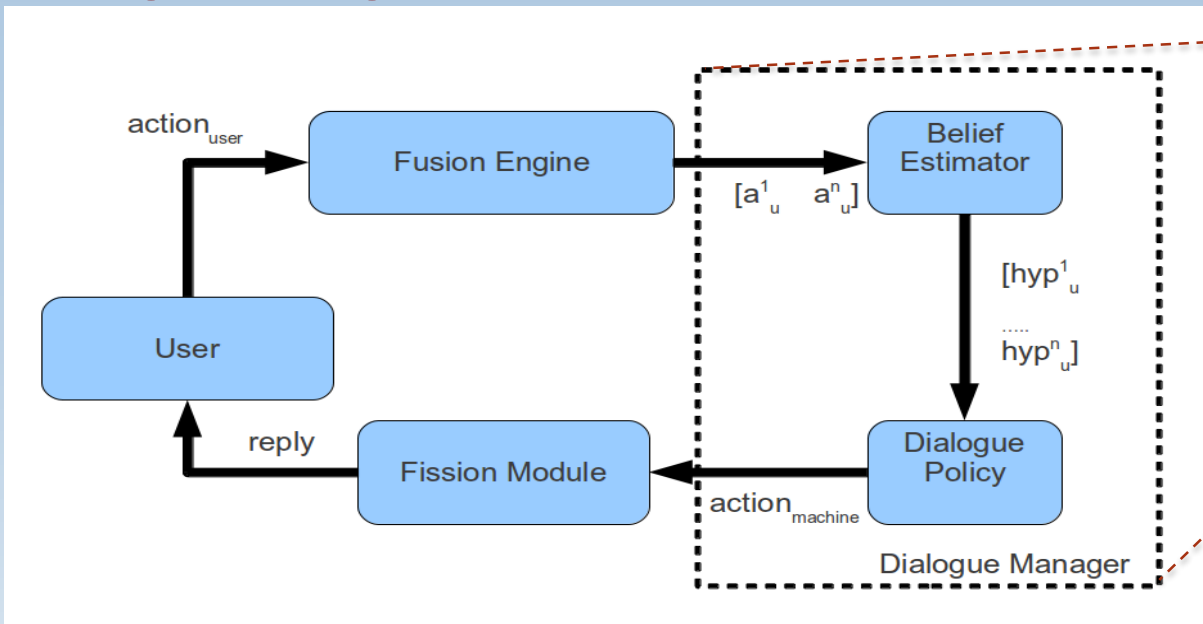
Point At	→	Go there, Specify object
Come Here	→	Come Here, Follow me
Hand's palm up	→	Give me, Show me
Idle	→	Take a decision
Walking	→	Follow me, Do nothing
Hand's palm Stop	→	Stop, Slow down, No
Grasp	→	Pick, Take
Circle in the air	→	Look for something
Release	→	Drop item



Dialogue Manager

Architecture

Dialogue Manager



- Dialogue state estimation according to the interaction history
- User intentions recognition from context and disambiguation of multiple hypotheses arising due to noisy or ambiguous situations.
- Dialogue coordination and action execution

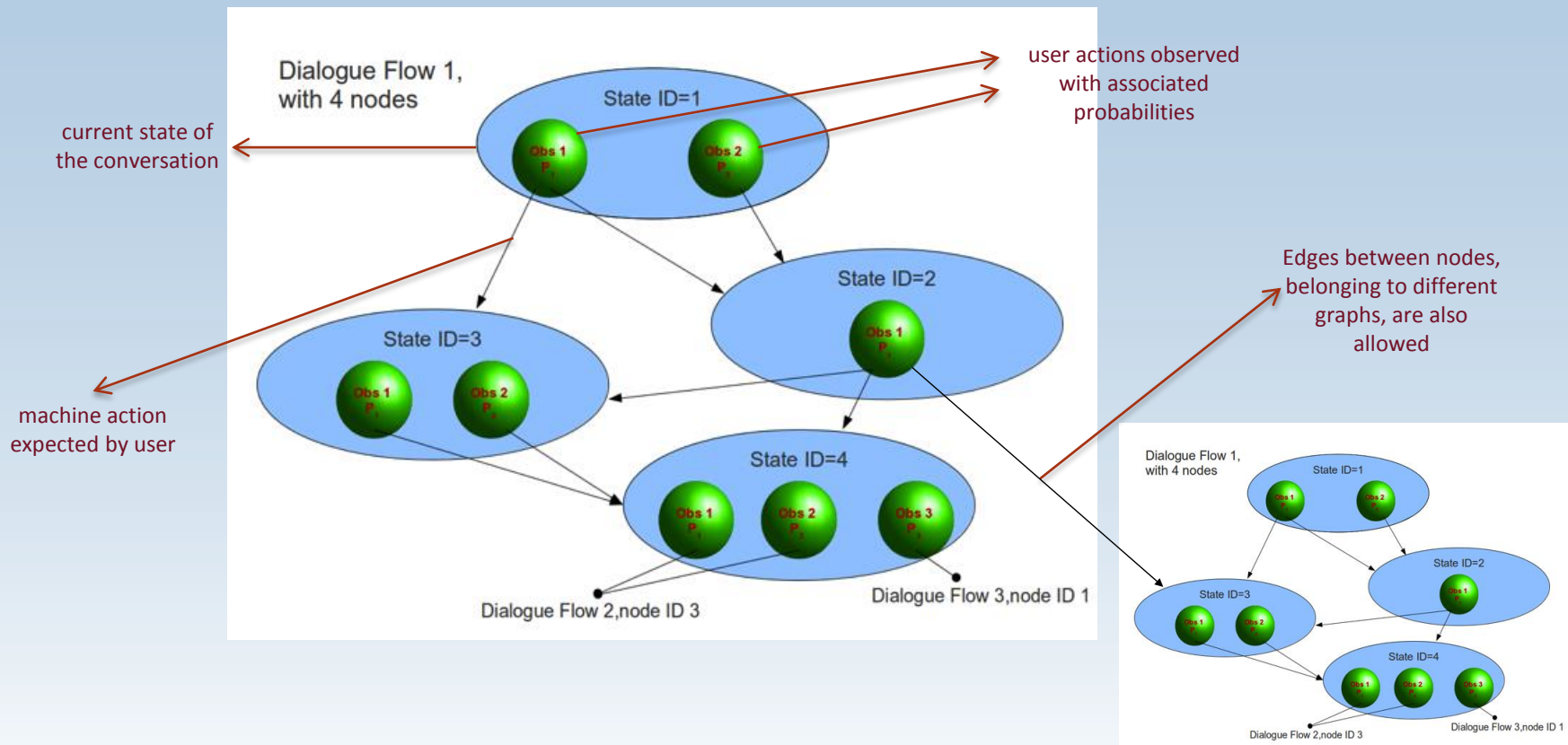
A Dialogue System for Multimodal Human-Robot Interaction , L. Lucignano, F. Cutugno, S. Rossi, A. Finzi, In Proceedings of 15th ACM International Conference on Multimodal Interaction - ICMI 2013

Dialogue Manager

Interaction Models

■ Interaction Models for Dialogue Management

- The system is provided with a set of interaction models named “dialogue flows”, which describe how the dialogue can develop



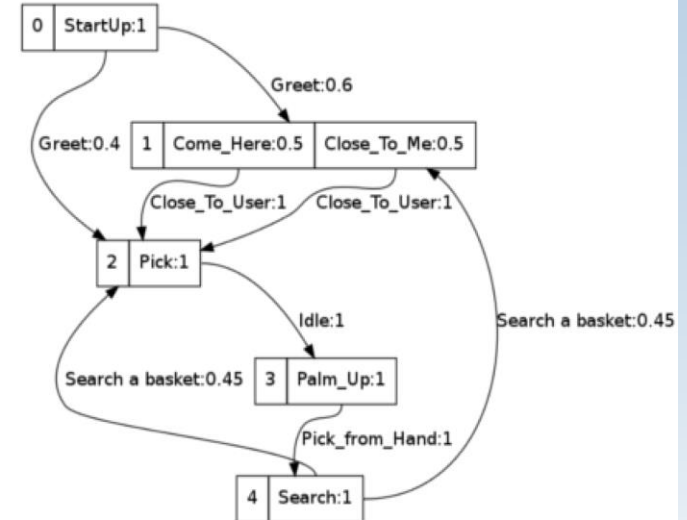
Dialogue Manager

Interaction Models

- Interaction Models for Dialogue Management
 - The system is provided with a set of interaction models, named “dialogue flows”, which describe how the dialogue can develop

XML description of a dialogue flow

```
<?xml version="1.0" encoding="UTF-8"?>
<Dialog label="Scenario1">
  <Node ID="0" label="Start" startingP="1">
    <Observe name="Startup" P="1">
      <MachineAction name="Greet">
        <Link_To ID="1">0.6</Link_To>
        <Link_To ID="2">0.4</Link_To>
      </MachineAction>
    </Observe>
  </Node>
  <Node ID="1" label="Robot is far, user wants it closer" >
    <Observe name="Come_Here" P="0.5">
      <MachineAction name="Close_To_User">
        <Link_To ID="2">1</Link_To>
      </MachineAction>
    </Observe>
    <Observe name="Close_To_Me" P="0.5">
      <MachineAction name="Close_To_User">
        <Link_To ID="2">1</Link_To>
      </MachineAction>
    </Observe>
  </Node>
  <Node ID="2" label="User ask to pick" >
    <Observe name="Pick" P="1">
      <MachineAction name="Idle">
        <Link_To ID="3">1</Link_To>
      </MachineAction>
    </Observe>
  </Node>
  <Node ID="3" label="User's palm is up with a ball " >
    <Observe name="Palm_Up" P="1">
      <MachineAction name="Pick from Hand">
        <Link_To ID="4">1</Link_To>
      </MachineAction>
    </Observe>
  </Node>
  <Node ID="4" label="Search a basket to place the ball" >
    <Observe name="Search" P="1">
      <MachineAction name="Search a basket">
        <Link_To ID="1">0.45</Link_To>
        <Link_To ID="2">0.45</Link_To>
      </MachineAction>
    </Observe>
  </Node>
</Dialog>
```



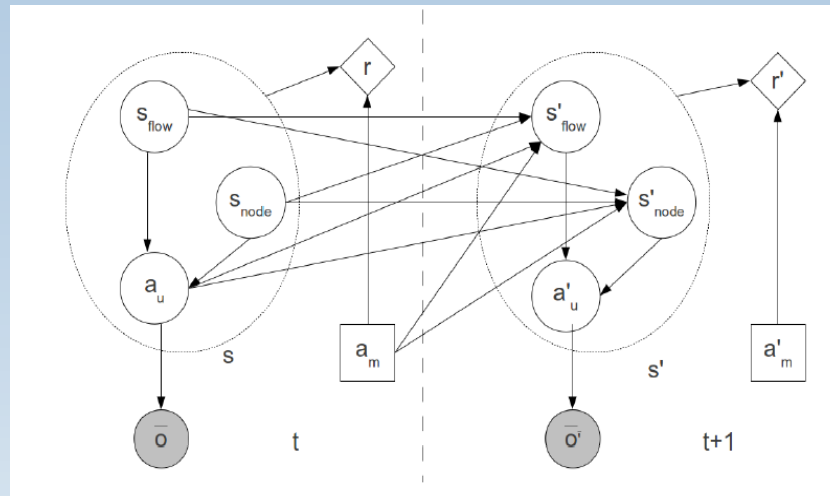
Dialogue Manager

- The Dialogue is represented by a Partially Observable Markov Decision Problem [Young10, Jurafsky00] extended to the multimodal case [Lucignano et al. 2013]

■ POMDP state is a tuple

$$\left(\underbrace{s_{flow}}_{\text{dialogue flow ID}}, \underbrace{s_{node}}_{\text{flow state ID}}, \underbrace{a_u}_{\text{last user's action}} \right)$$

■ POMDP Representation



■ POMDP solved using approximation methods:

- Point Based Value Iteration [Pineau et al. 2003], that approximates the value function only at a finite set of belief points
- Augmented MDP, that performs the optimization in a summary space rather than in the original space [Roy et al. 2000]

Attentional System

Attentional System and Cognitive Control for HRI

Attentional System:

It regulates both reactive and deliberative processes taking into account the interaction with the user (multimodal interaction, safety, naturalness and effectiveness)

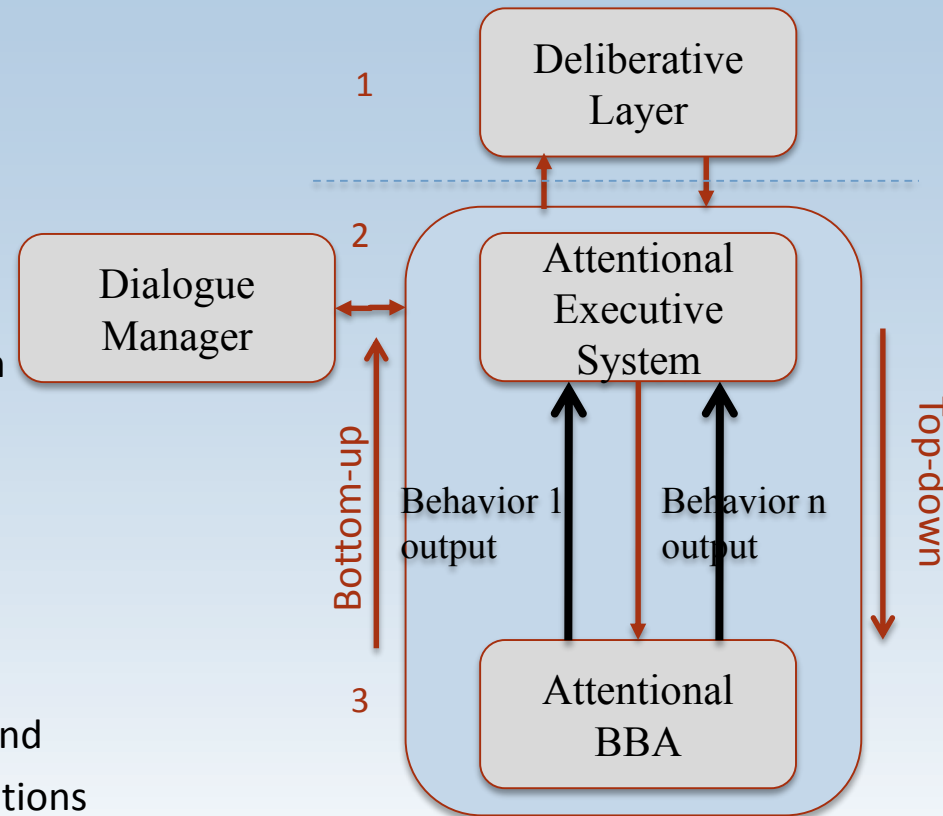
- We assume a layered architecture:

1. A **Deliberative layer**;

Attentional System:

2. An **Attentional executive layer** which orchestrates behaviours regulation, execution monitoring, and dialogue management;
3. An **Attentional behaviour-based layer** that provides adaptive and reactive control.

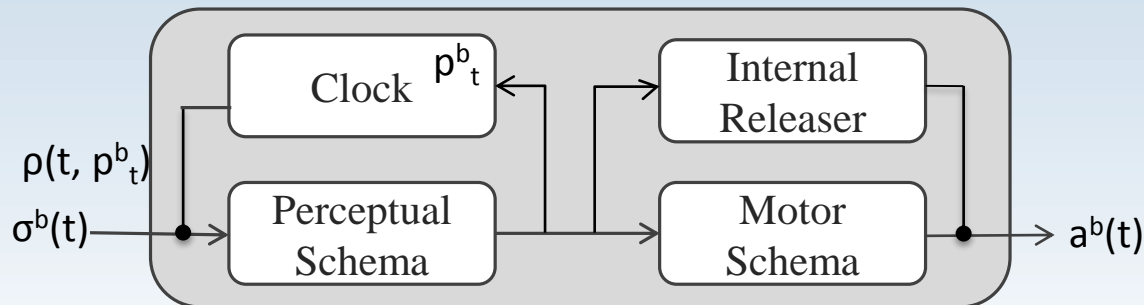
- The Attentional System integrates
 - **bottom-up** (event-based, stimulus-driven) and
 - **top-down** (task-based, goal-directed) regulations



Attentional System

Behavior-based attention system

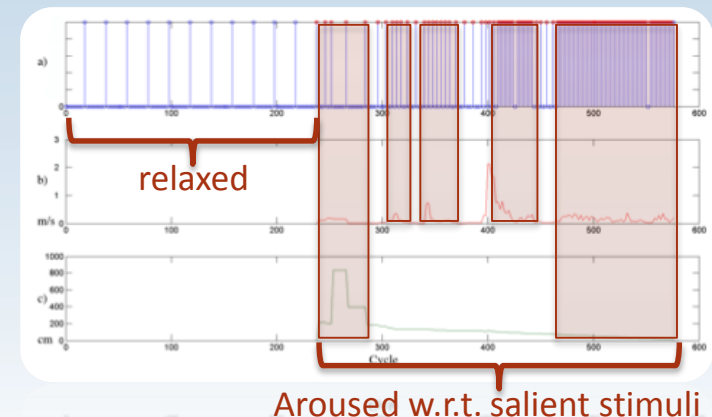
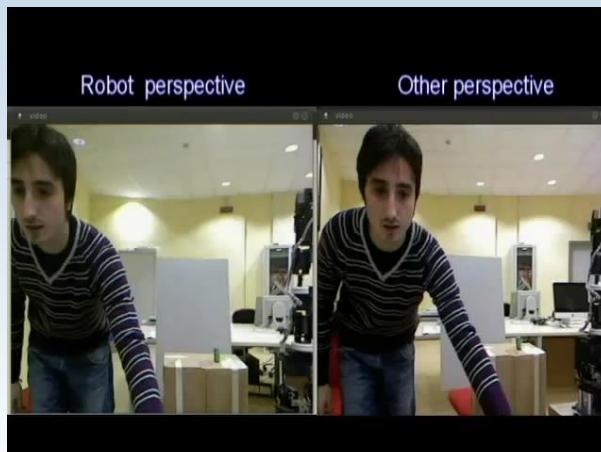
- **Frequency-based model of attention:**
 - The higher the attention the higher the resolution at which a process is monitored and controlled [Senders 1964, Posner et al. 1980].
- **Behavior-based architecture:**
 - Each behavior is endowed with an internal adaptive clock [Burattini, Rossi 2008] that represents an attentional mechanism [Burattini et al., 2010].
- **Internal Adaptive Clock:**
 - Attentional monitoring strategies increase/decrease the clock frequency of each behavior depending on salient internal/external stimuli (e.g. human disposition in the environment).



Attentional System

Behavior-based attention system

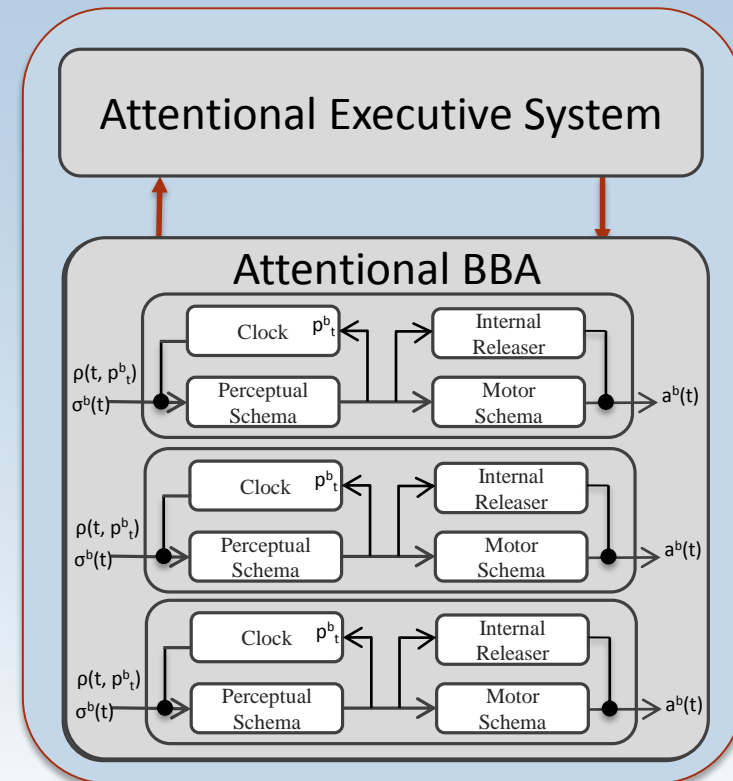
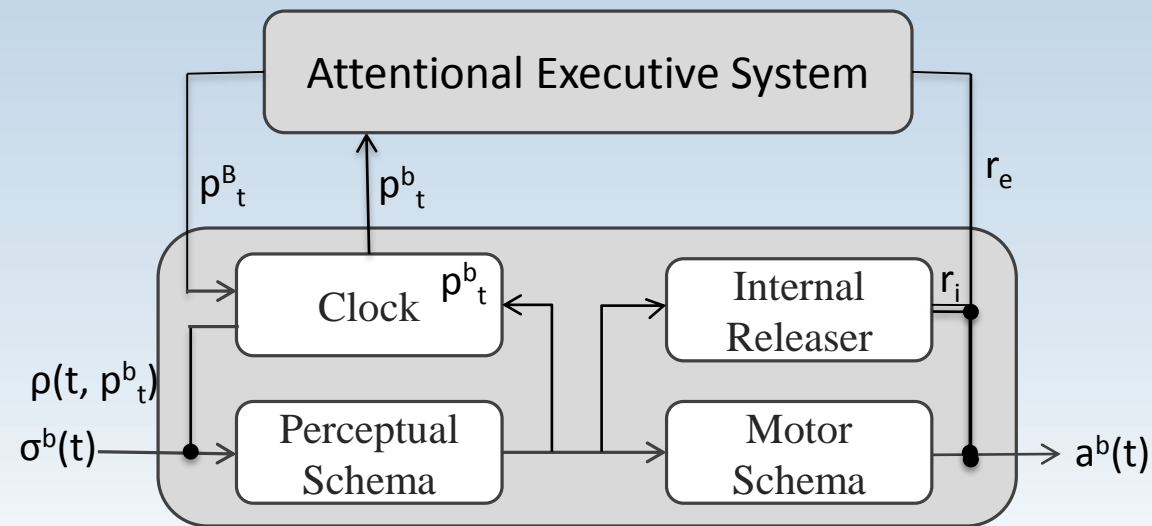
- **Frequency-based model of attention:**
 - The higher the attention the higher the resolution at which a process is monitored and controlled [Senders 1964, Posner et al. 1980].
- **Behavior-based architecture:**
 - Each behavior is endowed with an internal adaptive clock [Burattini, Rossi 2008] that represents an attentional mechanism [Burattini et al., 2010].
- **Internal Adaptive Clock:**
 - Attentional monitoring strategies increase/decrease the clock frequency of each behavior depending on salient internal/external stimuli (e.g. human disposition in the environment).



Attentional System

Attentional Executive System

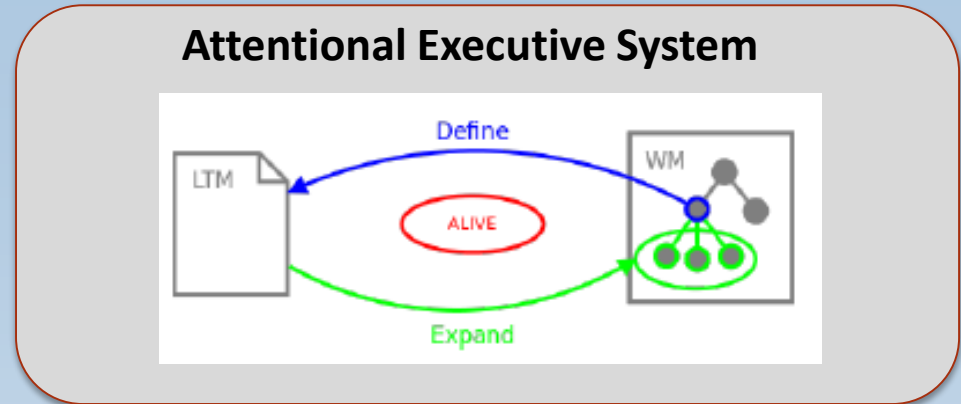
- Cognitive control and top-down regulations:
 - Execution monitoring, goal-directed behavior orchestration
 - Depending on the task/context/machine action (dialogue) it defines:
 - Behavior allocation;
 - Top-down attentional regulations.



Attentional System

Attentional Executive System: Cognitive Control Cycle

- **Long Term Memory (LTM):**
 - Repertory of hierarchical tasks
- **Working Memory (WM):**
 - Current executive state
 - Tasks in the attentional focus
- **Cognitive Control Cycle:**
 - Continuously updates the tasks hierarchy in the WM
 - Hierarchical tasks can activate and drive a hierarchy of attentional behaviors
 - Task-coherent behaviors are enhanced (high-frequency); inhibited otherwise



```

schema(alive,[
  [sonarStream,1,["TRUE"]],
  [engineStream,1,["TRUE"]],
  [blobStream,1,["TRUE"]],
  [inputStream,1,["TRUE"]],
  [requestStream,1,["TRUE"]]]).

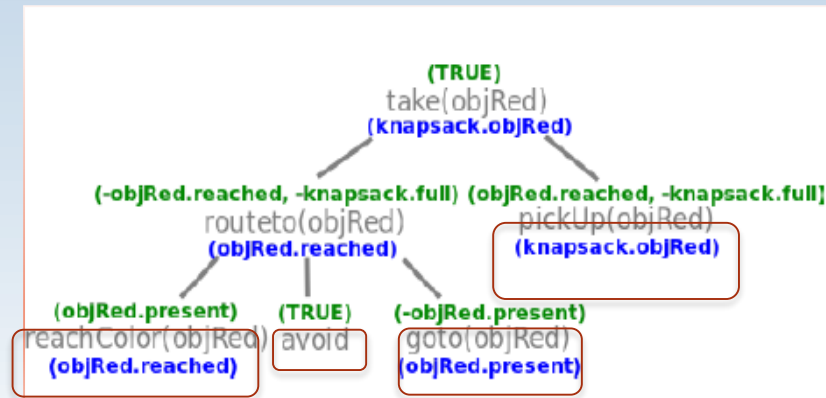
schema(goto(X,Y),[
  [avoid,1,["TRUE"]],
  [gotoxy(X,Y),1,["TRUE"]]]).

schema(followColor(Color),[
  [avoid,1,["TRUE"]],
  [reachColor(Color),1,["TRUE"]]]).

schema(searchColor(X),[
  [avoid,1,[-X.near]],
  [wander,1,[-X.present]],
  [reachColor(X),1,[X.present]]].
  
```

Long Term Memory

Top-down

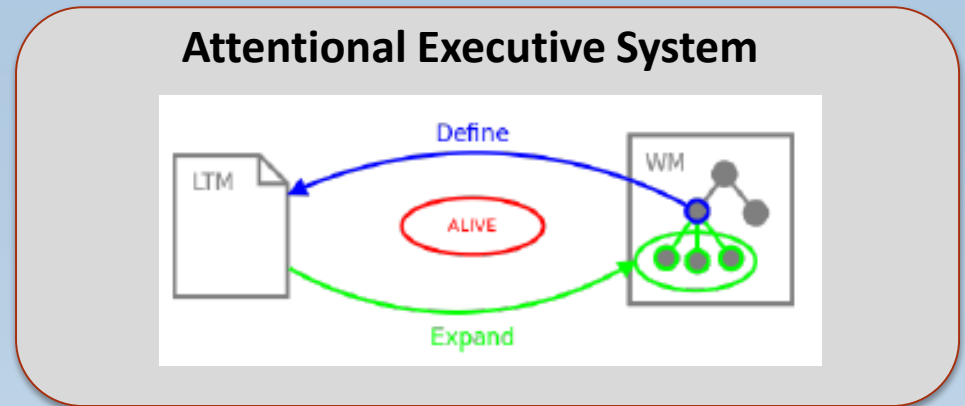


Hierarchical task in the WM

Attentional System

Attentional Executive System: Cognitive Control Cycle

- **Long Term Memory (LTM):**
 - Repertory of hierarchical tasks
- **Working Memory (WM):**
 - Current executive state
 - Tasks in the attentional focus
- **Cognitive Control Cycle:**
 - Continuously updates the tasks hierarchy in the WM
 - Hierarchical tasks can activate and drive a hierarchy of attentional behaviors
 - Task-coherent behaviors are enhanced (high-frequency); inhibited otherwise



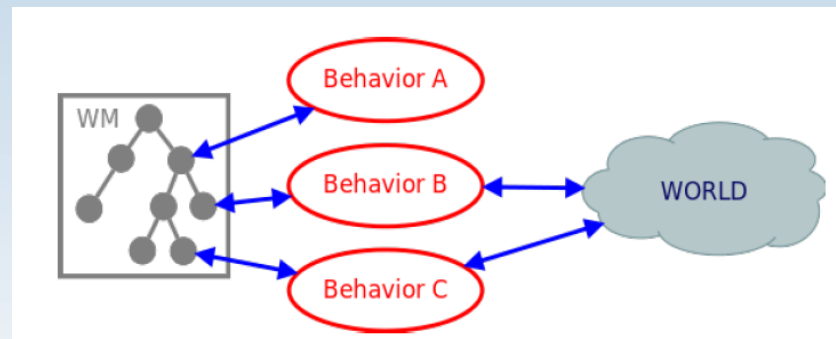
```
schema(alive,[
  [sonarStream,1,["TRUE"]],
  [engineStream,1,["TRUE"]],
  [blobStream,1,["TRUE"]],
  [inputStream,1,["TRUE"]],
  [requestStream,1,["TRUE"]]]).

schema(goto(X,Y),[
  [avoid,1,["TRUE"]],
  [gotoxy(X,Y),1,["TRUE"]]]).

schema(followColor(Color),[
  [avoid,1,["TRUE"]],
  [reachColor(Color),1,["TRUE"]]]).

schema(searchColor(X),[
  [avoid,1,[-X.near]],
  [wander,1,[-X.present]],
  [reachColor(X),1,[X.present]]].
```

Long Term Memory

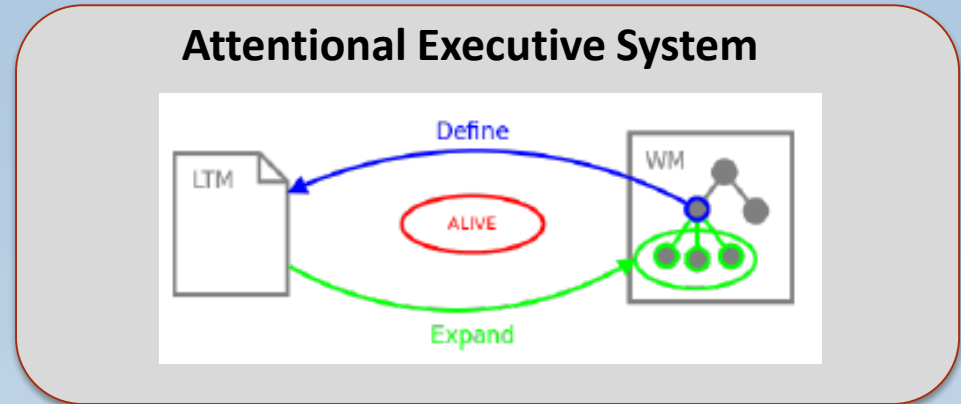


WM and attentional behaviors

Attentional System

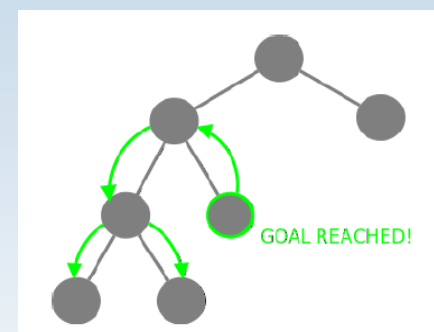
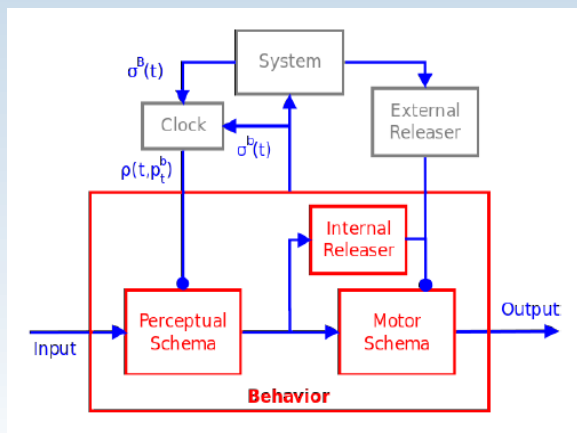
Attentional Executive System: Cognitive Control Cycle

- **Long Term Memory (LTM):**
 - Repertory of hierarchical tasks
- **Working Memory (WM):**
 - Current executive state
 - Tasks in the attentional focus
- **Cognitive Control Cycle:**
 - Continuously updates the tasks hierarchy in the WM
 - Hierarchical tasks can activate and drive a hierarchy of attentional behaviors
 - Task-coherent behaviors are enhanced (high-frequency); inhibited otherwise



$$p_t^b = f(\sigma^b(t), \sigma^B(t), p_{t-1}^b)$$

$$releaser^b = releaser_{int}^b \cdot releaser^s$$



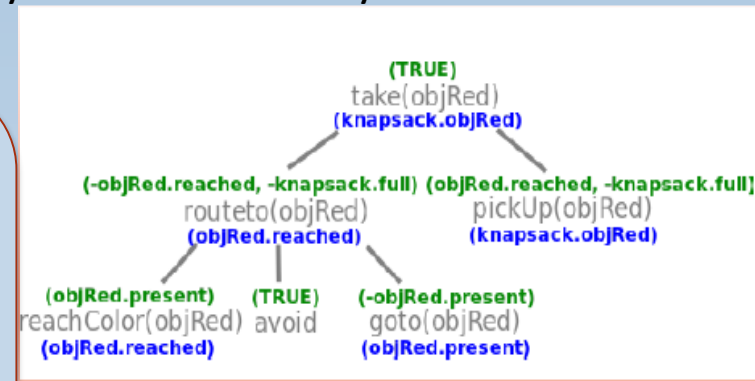
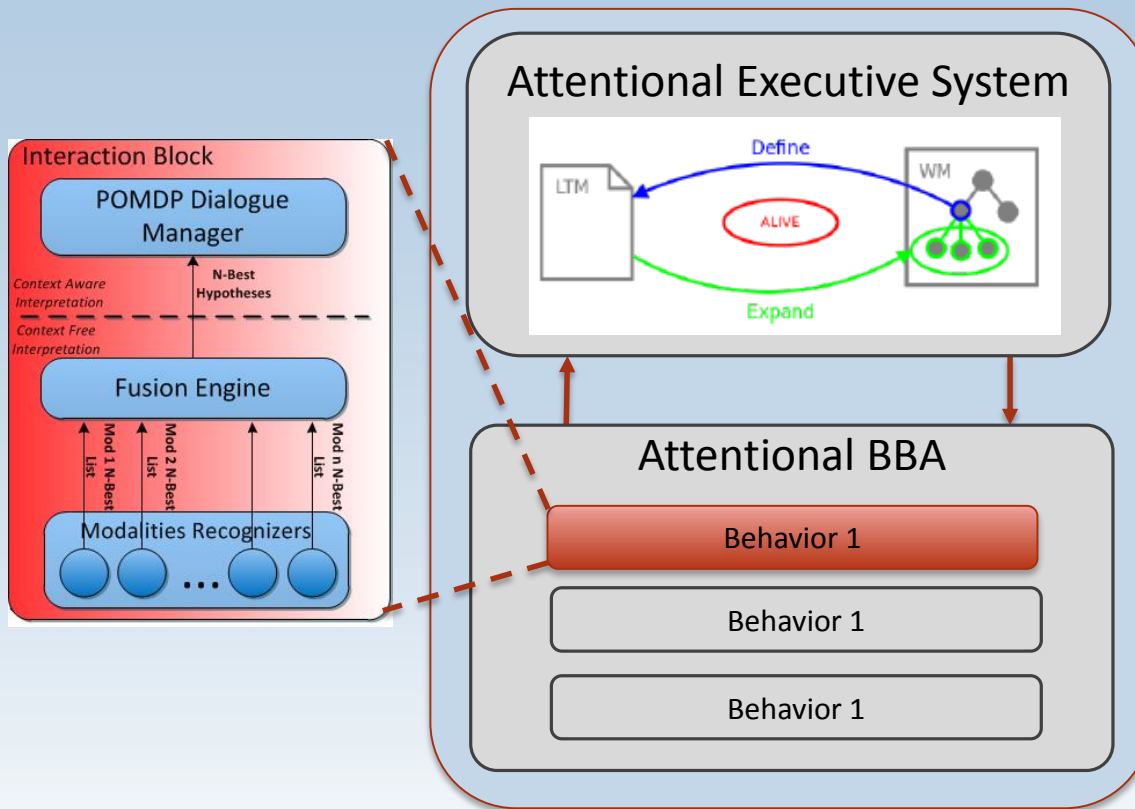
$$mag_{t+1}^{father} = mag_t^{father} + 1$$

Multimodal Attentional Interaction

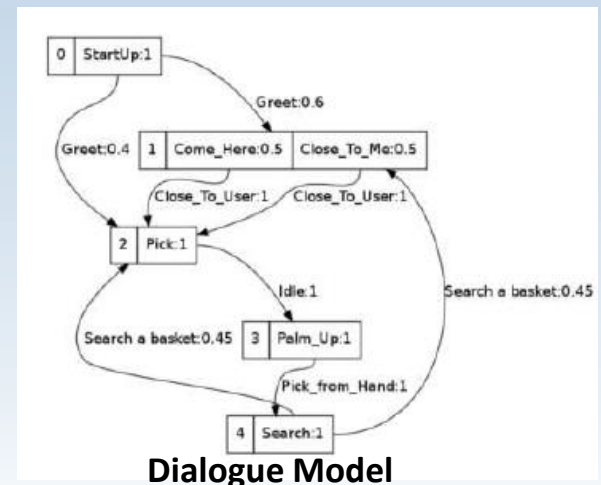
Attentional System, Dialogue Manager, and Multimodal Interaction

Attentional System:

- The dialogue manager is treated as a special interactive behavior
- The dialogue policy is integrated and regulated by the attentional system that provides contextual and task-related contents



Hierarchical task in the WM

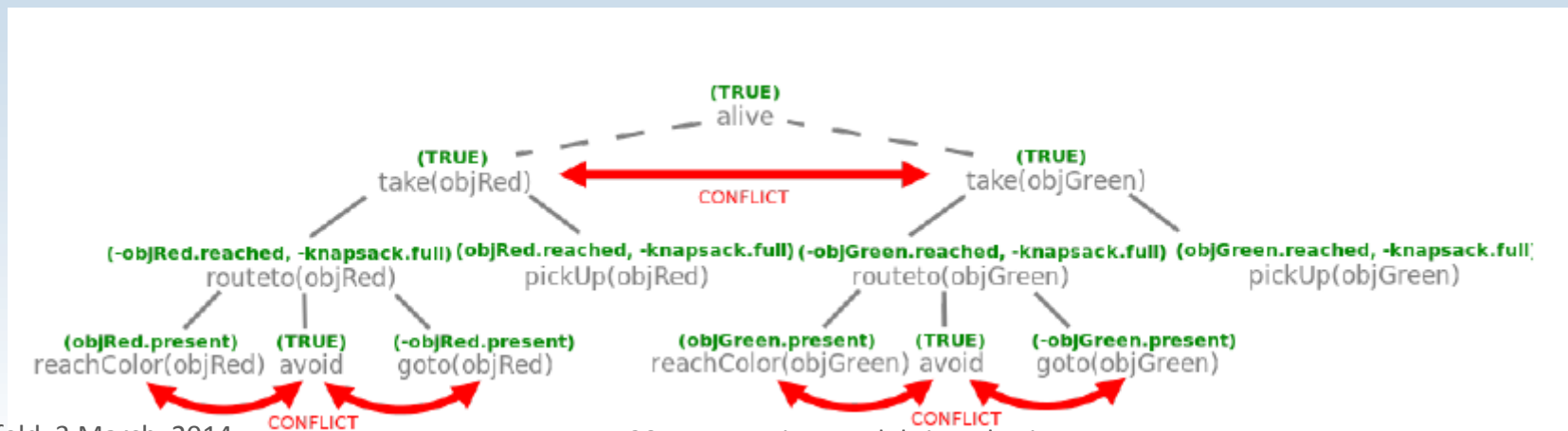
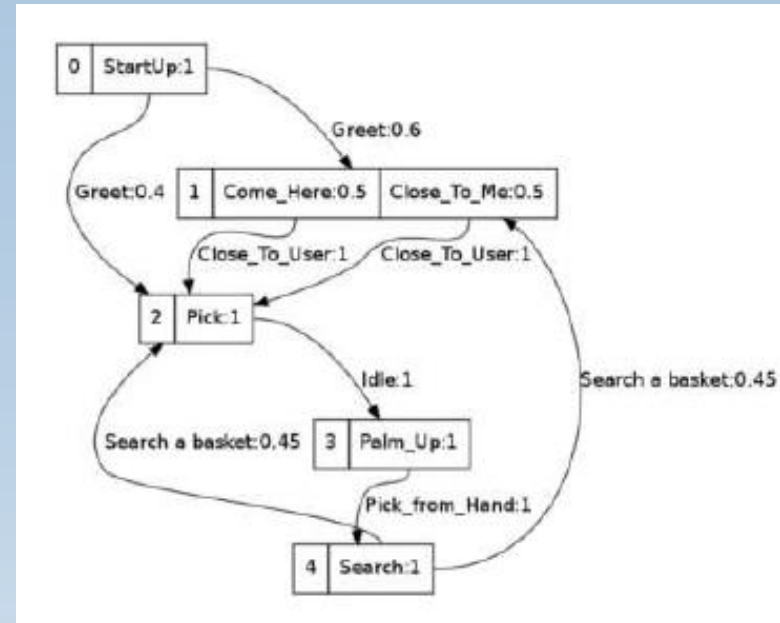


Dialogue Model

Attentional System and Dialogue Manager

Conflict resolution and policy integration

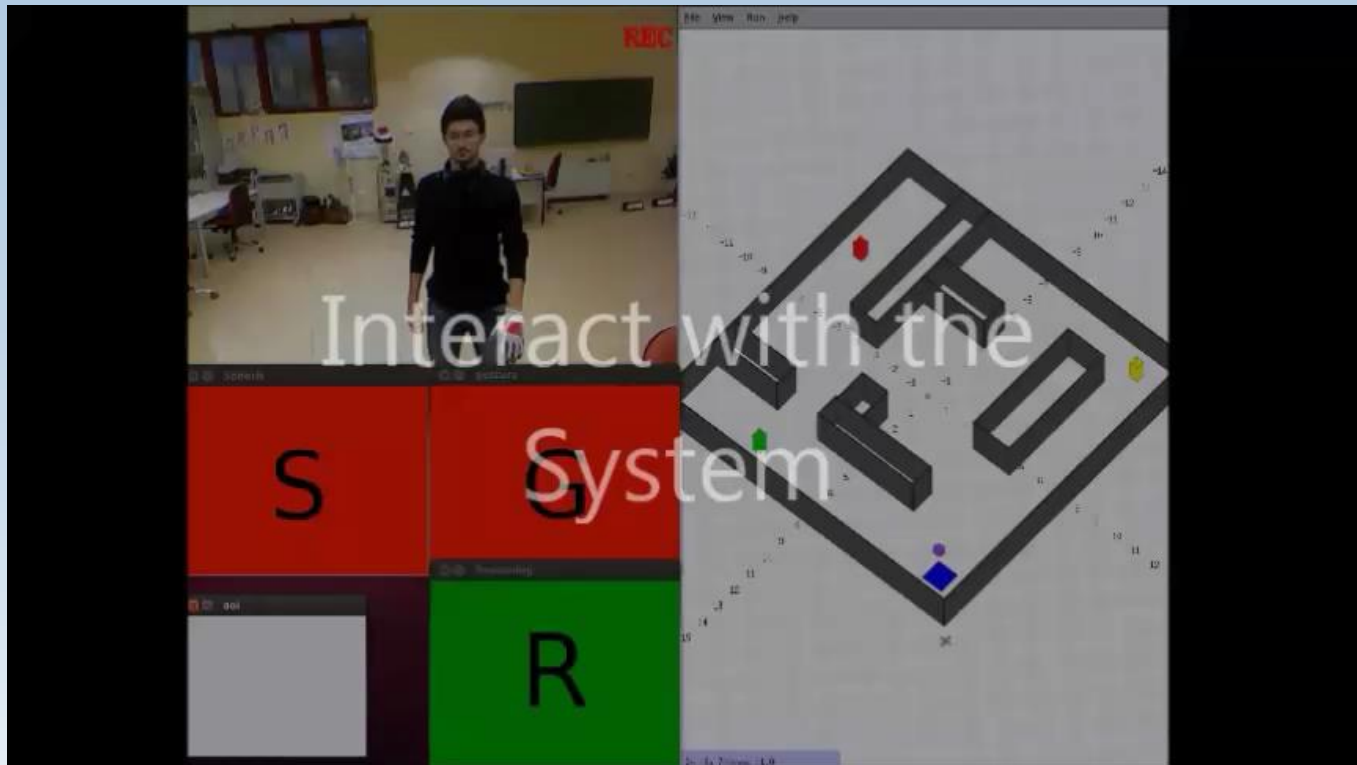
- **Dialogue policy integration:**
 - The dialogue policy regulates the interaction (communication, disambiguation, turn taking, etc.), but task and context/task-based data can be missed (e.g. gesture “take”, which object?)
- **Attentional system integration:**
 - Ambiguities and decisional impasses:
 - context/task coherence affects decisions about actions and parameters that instantiate the dialogue policy.



Multimodal Attentional Interaction

Top-down attentional mechanisms in collaborative activities: simulated scenario

- Multimodal interaction, Dialogue Management, Attentional Modulation:
 - Simulated pick-carry-place scenario



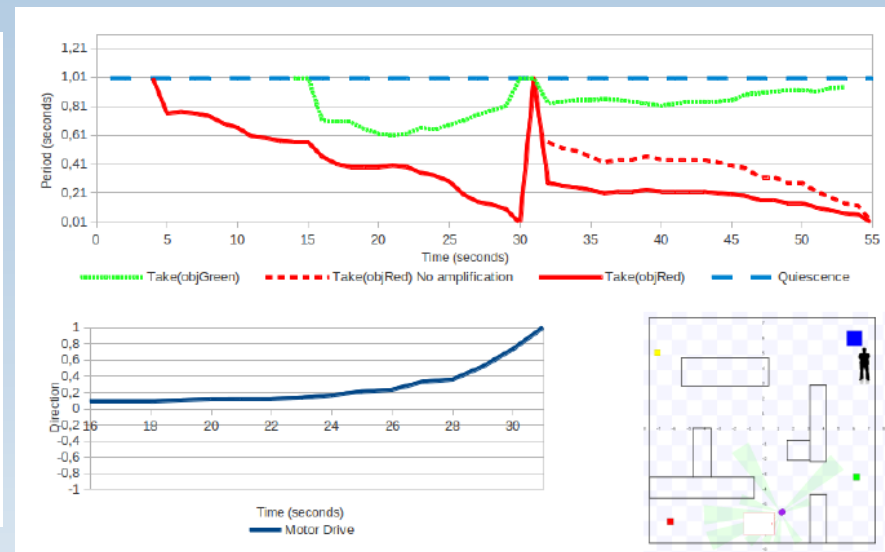
Multimodal Attentional Interaction

Top-down attentional mechanisms in collaborative activities: simulated scenario

- Multimodal interaction, Dialogue Management, Attentional Modulation:
 - Simulated pick-carry-place scenario
 - Preliminary tests (conflict resolution)

EXECUTION TIME			
Task Sequence	Time (min)	Task Sequence	Time (min)
TakeRed - TakeGreen		TakeRed - TakeGreen - TakeYellow	
Red Green Give	4.5	Red Green Give Yellow Give	9.19
Green Give Red Give	7.11	Green Give Red Give Yellow Give	8.19
Green Give Red Give	8.04	Red Green Give Yellow Give	7.21
Green Give Red Give	7.14	Yellow Give Green Give Red Give	9.08
Red Green Give	3.53	Yellow Green Give Red Give	7.28
Green Red Give	3.50	Red Green Give Yellow Give	6.41
Red Green Give	4.19	Red Green Give Yellow Give	7.02
Green Give Red Give	6.04	Red Green Give Yellow Give	7.05
Red Green Give	4.48	Yellow Give Green Give Red Give	9.43
Green Red Give	6.26	Red Green Give Yellow Give	8.48
AVG	STD	AVG	STD
5.48	1.64	7.93	1.07

Execution time of generic take() in different situations



Top-down and bottom-up influences

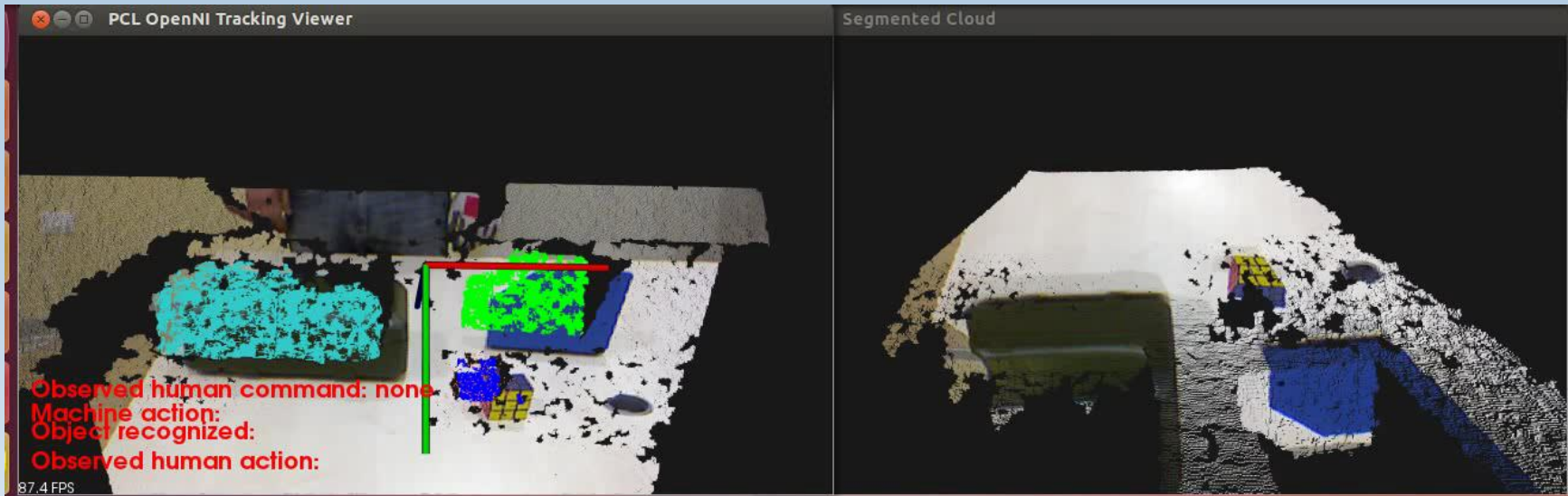
EXECUTION TIME (min)					
Take-Red		Take-Green		Take-Yellow	
avg	std	avg	std	avg	std
3.99	0.28	1.48	0.36	2.04	0.27

Execution time of a specific take()

Multimodal Attentional Interaction

Top-down attentional mechanisms in collaborative activities: real scenario

- Multimodal interaction, Dialogue Management, Attentional Modulation:
 - Table scenario, object recognition, and tracking



- **Multimodal:** Observed human action (command or action);
- **Dialogue Policy:** Machine action (command execution/initiative/explanation req.);
- **Executive System:** Attentional set, task-based behavior allocation, conflict resolution.

Multimodal Attentional Interaction

Cognitive executive control for a collaborative robot

- Multimodal interaction, Dialogue Management, Attentional Modulation:
 - Table scenario, object recognition and tracking

**Multimodal interaction
attentional regulation
object recognition and tracking**



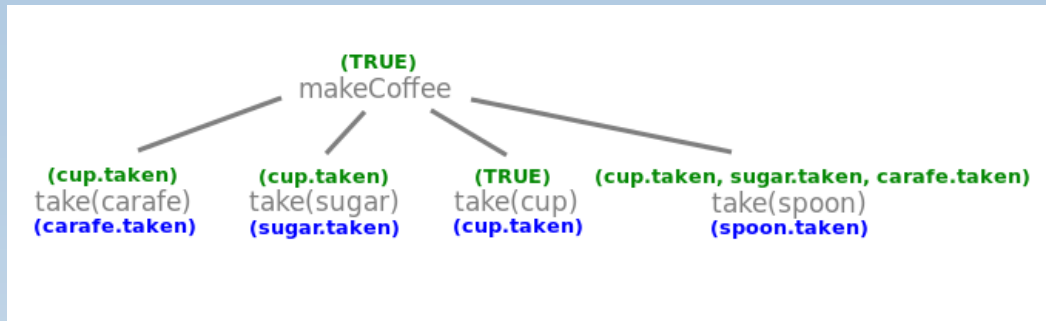
UNIVERSITA' DEGLI STUDI DI
NAPOLI FEDERICO II



Multimodal Attentional Interaction

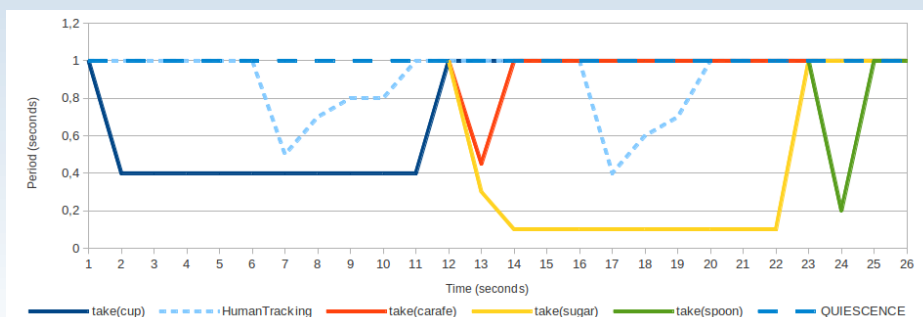
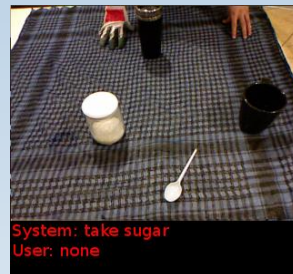
Cognitive executive control for a collaborative robot

- Multimodal interaction, Dialogue Management, Attentional Modulation:
 - Coffee scenario, object recognition and tracking



Coffee Scenario

4 objects on a table (cup, coffee carafe, a sugar bowl, spoon). The human is to prepare the coffee by collecting these objects in a suitable order: first the cup, then the sugar or the carafe, finally the spoon



The human can either take an object or receive it from the robot. Depending on the human-action target a gesture can be interpreted as a command or as an action.

Period modulation profile associated with this sequence of robot (solid line) and human actions (dotted line).

Conclusion

Summary and on-going work

■ Summary:

- Attention and dialogue management:
 - The dialogue policy provides an interaction template which is instantiated and continuously adjusted by the attentional system with respect to the environmental and the operative context.
- Attentional Executive System:
 - Structured tasks and reactive behaviors
 - Both bottom-up and top-down attentional modulations enable the system to execute structured tasks and solve decisional impasses

■ On-going work:

- Testing more complex interactive scenarios
- Integration of visual attentional mechanisms
- Integration of a deliberative layer