Levels of Interaction Allowing Humans to *Command*, *Interrogate* and *Teach* a Communicating Object : Lessons Learned From Two Robotic Platforms

Peter Ford Dominey

CNRS, France

Alfredo Weitzenfeld

ITAM, Mexico

Objective

- Spoken Language interface for Human-Robot Interaction that allows
  - Commanding or directing behavior
  - Asking questions of the system
  - Teaching the system new behaviors
  - Recognising different users and their level of experience

Method: Platform 1
The Visual Scene Describer

- Off-The-Shelf
  - Color based segementation
  - Speech recognition, syntheses, dialog mgmt
- Development
  - Extraction of "Meaning"
    - Events from spatiotemporal schemas
  - Sentence-Meaning mapping
    - Grammatical Constructions

Vision, Event Extraction

What happended to the block?

Dialog Mgmt Speech synthesis & recogntion (CSLU RAD)

Human narrator

The block pushed the cylinder

Human manipulation of objects
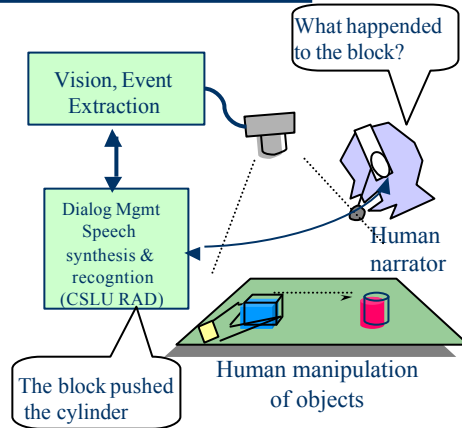
Event Description

- Object attention via motion
- Events via Contact Sequences
- Sentence Generation via Grammatical Constructions
  - meaning - sentence
- ' Pragmatics '
  - gave(moon, cylinder, block)
  - The moon gave the cylinder to the block.
  - The block was *gave* the cylinder by the moon. (given)
 - The cylinder was *gave* to the block by the moon. (given)

Learning Spatial Configurations

- Spatial attention directed to objects that have been moved
- Ensemble of primitive relations (horizontal and vertial) extracted
- Global form characterized
- User invited to name the demonstrated relation

Handling Unknown Relations

- System identifies user, and acts accordingly
- If the user asks the system to identifiy an unknown relation
- The system invites the user to name it for future reference

- Demonstration of interrogation and teaching
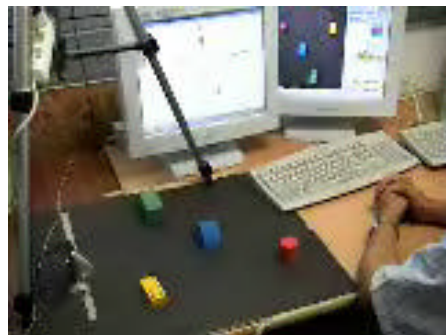
Method: Command, Interrogation and Teaching Platform 2 Aibo ERS7 Mobile Robot

- Off-The-Shelf
  - Vison & Motor Control
  - Speech recognition, syntheses, dialog mgmt
- Development
  - Extraction of "Meaning"
    - Events from spatiotemporal schemas
  - Sentence-Meaning mapping
    - Grammatical Constructions
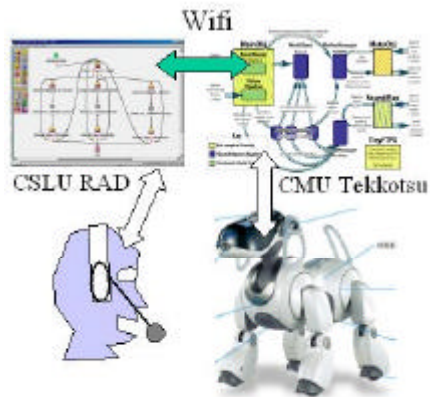
Learning Action Commands Introduction and New commands

- Explain system to new user
- Invite user to link a behavior with a command or button press
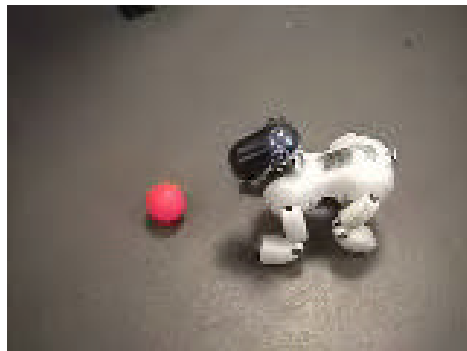
Spoken language for command and interrogation

- Access to behavior via spoken language
- Interrogation
- Goal directed telecommand

Lessons Learned

- Flexibility in Dialog
  - user should be able to say and do what they want when they want
- Less-constrained speech
  - richer inventory of constructions
  - learning new grammatical constructions on the fly
- Will require dynamic ontology
  - Project Rhone-Alpes Cluster « Presence »
 - Thesis project JD Boucher

| | Robot Platforms | |
	Platform 1. Event Vision and Description	Platform 2. AIBO Autonomous Robot
Capability		
1. Tell		Command different actions (shake, chase the ball, etc.)
2. Ask	Ask who did what in a given action	Ask what is the battery state ? Can you see the ball
3. Teach	This is a triangle This is a square, etc.*	Associate perceptual events with behaviors. Head-touch -> Bark.

* Teach grammatical constructions

(1) Extracting meaning from the environment: Event Perception

Give(2, 3, 1)

- Vision: Color based segmentation
- Extraction of perceptual primitives
  - *contact(agent, objet, duration)*
  - *Position,velocity*

Contact (2, 3)

- Describe event primitives
  - *touch, push, take* can be described as contacts, and durations.
  - duration: *Touch < Push < Take*
  - Agency =f(relative velocity)
- Compose Complex Events
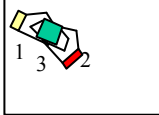 - Give(agent, object, recipient)
  - Take(agent, object, recipient)
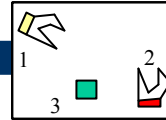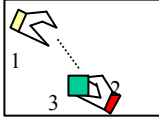
Contact (3, 1)
Contact (2, 1)

---

(3) Identifying-discriminating between different grammatical structures of input sentences

Inventoried via competing cues

form

ConstructionIndex
"__ was __ to __ by __"

"The ball was given to Jean by Marie"

Construction Inventory

Structured inventory of form to meaning mappings

Associative memory

ACTION	*gave*	*ball*	ball
AGENT	*Marie*	*gave*	given
OBJECT	*ball*	*Jean*	Jean
RECIPIENT	*Jean*	*Marie*	Marie

SceneEvent Array

SentToScene
Predicted
Referents Array

WordToReferent

OpenClassArray

meaning