



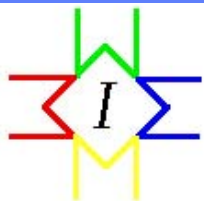
N
C
A
R
A
I

Multi-modal Interfacing for Human-Robot Interaction

Dennis Perzanowski, Alan Schultz,
William Adams, Magda Bugajska,
and Elaine Marsh

<dennisp | schultz | adams | magda | marsh> @aic.nrl.navy.mil

NAVAL RESEARCH LABORATORY
NAVY CENTER FOR APPLIED RESEARCH IN ARTIFICIAL INTELLIGENCE
Codes 5512 and 5515
Washington, DC 20375-5337



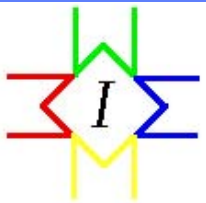
Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE 2001		2. REPORT TYPE		3. DATES COVERED -	
4. TITLE AND SUBTITLE Multi-modal Interfacing for Human-Robot Interaction				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Navy Center for Applied Research in Artificial Intelligence,Naval Research Laboratory,Code 5510,Washington,DC,20375				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES The original document contains color images.					
14. ABSTRACT see report					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES 21	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			



Objectives of Natural Language/Gesture Research

- Use a dialog-based approach to achieve:
 - ◆ Integrated multi-modal interface in a robotics domain
 - ◆ Dynamic autonomy
- Seamlessly integrate natural language and gestural communication
 - ◆ Ambiguous natural language utterances without gestures
 - ◆ deixis
 - “Go over there” <with/out gestures>
 - ◆ Contradictory inputs
 - ◆ “turn left” <while pointing right>
 - ◆ “Natural” and “synthetic” gestures coupled with speech and buttons
 - ◆ **Develop continuous dialog with/out interruptions**
 - ◆ **Facilitate dynamically changing levels of autonomy and interaction**

N
C
A
R
A
I





Working Hypotheses

■ Linguistic Hypothesis:

- ◆ Gesture disambiguates and contributes information in human-human dialog

■ Gestural Hypothesis:

- ◆ Gesturing is natural in human-human dialog
 - Hand/arm movement vs. Electronic devices, e.g. mouse, light-pen, or touch-screen

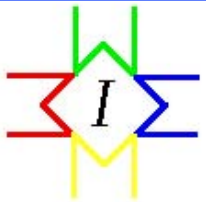
■ Assumption of natural language/robotics research

- ◆ “...With just a very few human-like cues from a humanoid robot, people naturally fall into the pattern of interacting with it as if it were a human.”--
 - ◆ Quote taken from The COG Shop website: <http://www.ai.mit.edu/projects/cog/>

And as we all know,

- humans can be pretty independent
- humans desire human-like cooperation in the systems they design

N
C
A
R
A
I

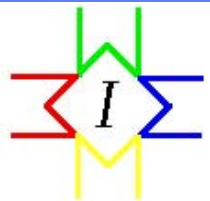
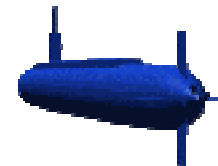




Dynamic Autonomy

■ Re-deployment during mission interspersed with periods of autonomy

- Micro air vehicles launched
- autonomous underwater vehicles
- planetary rover

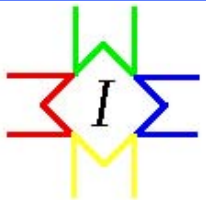




Mixed-Initiative Systems and Dynamic Autonomy

N
C
A
R
A
I

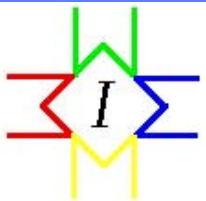
- **Command and control situations**
 - ◆ Characterized by tight human/robot interactions.
 - ◆ Instantiating a goal is a function of either agent in these situations.
 - ◆ These are by definition “Mixed-initiative” systems.
- **Levels of independence, intelligence and control are necessary in “mixed-initiative” systems**
 - ◆ Dynamic autonomy is necessary to achieve these varying levels.





Hardware and Software

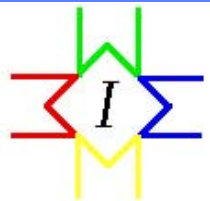
- ◆ COTS speech recognizer
 - ◆ IBM ViaVoice Pro Millennium Edition
- ◆ NAUTILUS natural language processor
 - ◆ In-house natural language understanding system
 - Programming in Allegro Common Lisp and C++ on PCs
 - Programming in Gnu Common Lisp and C++ on Suns
 - ◆ Messages are passed via “foreign functions” between modules in a kind of “blackboard” architecture
- ◆ COTS mobile robots
 - ◆ Nomadic Technologies 200 and XR-4000 mobile robots
 - ◆ RWI ATRV-Jr.
- ◆ Personal Digital Assistant
 - ◆ Palm family, e.g. 3 COM Palm V Organizer





Linguistic Constructs

- We are using two linguistic variables, “context predicates” which contain location information and goals, to track both interrupted and non-interrupted goal completion in a command, control and interaction environment (“C₂I”).
- Context predicates and goal information are being used to enable greater independence and cooperation between agents in a C₂I environment.





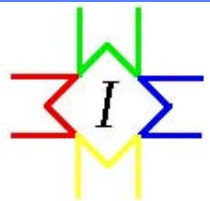
Predicates and Goals

■ CONTEXT PREDICATES

- ◆ a stack of goals and their status (attained vs. unattained)

■ GOALS

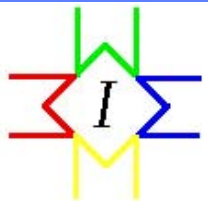
- ◆ Event goals
 - ◆ “turn left/right”--arriving at the final state of having turned in a particular direction
- ◆ Locative goals
 - ◆ “there” “the waypoint” “table”





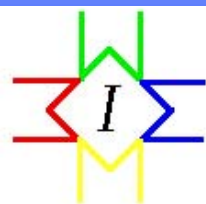
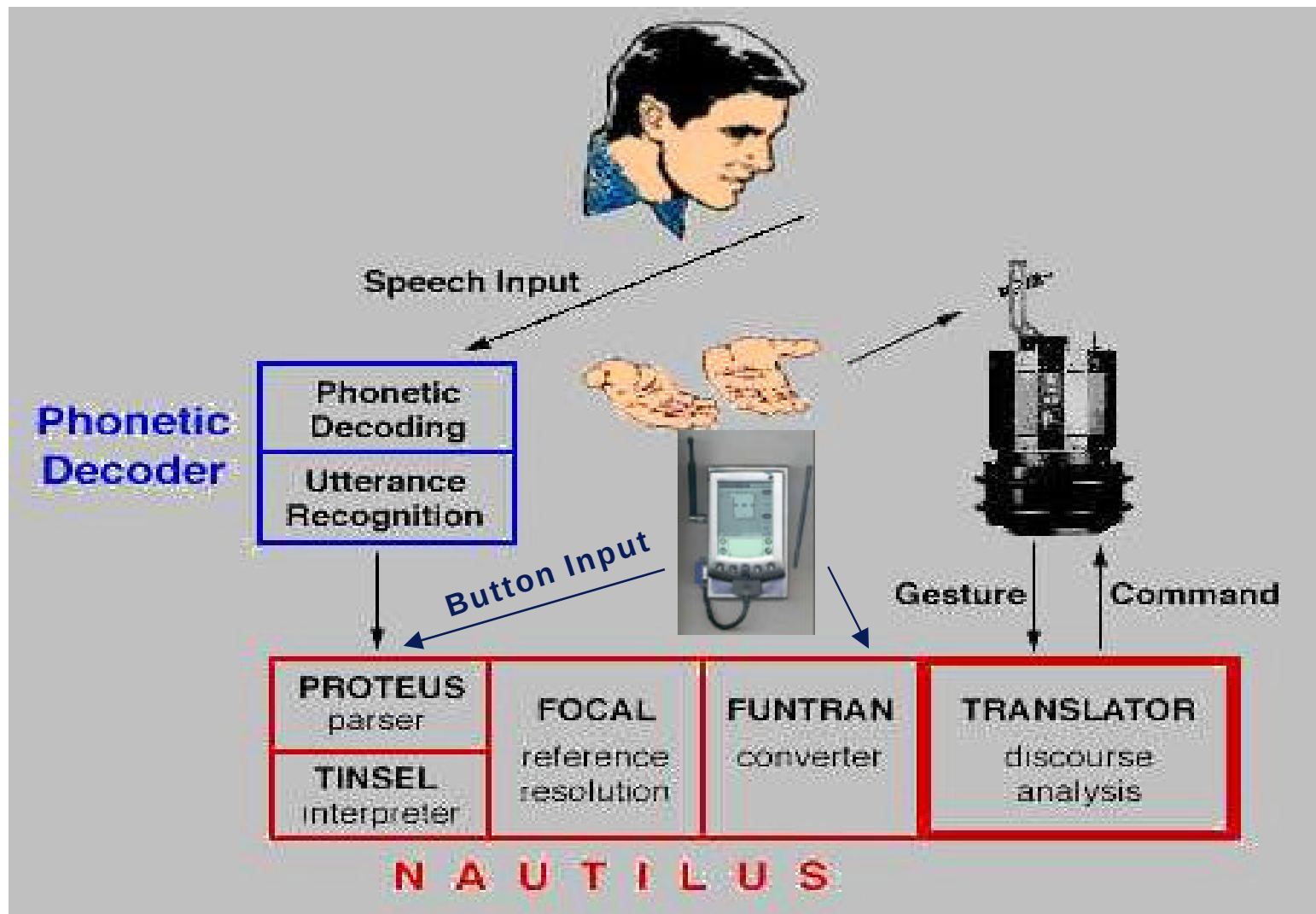
Object and Gesture Recognition

- Structured-light range finder (camera + laser)
 - ◆ output: 2D range data
- 16 ultra-sonic sonars
 - ◆ output: range data out to 25'
- 16 active infrared sensors
 - ◆ output: delta of ambient light and current light
(detects if object present)
- bumpers for collision avoidance





Interaction with NL Component





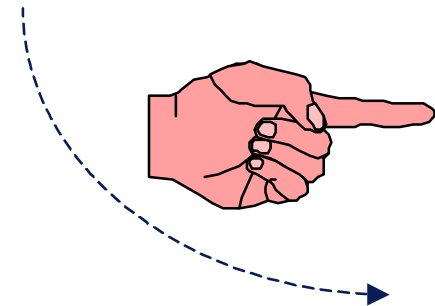
Processing Speech and Gesture



GO TO THE WAYPOINT OVER THERE



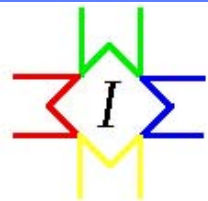
```
(IMPER (:VERB GESTURE-GO
(:AGENT (:SYSTEM YOU))
(:TO-LOC (WAYPOINT-GESTURE))
(:GOAL (THERE))))
```



(2 42 123456.123456 0)



(sending message: "17 42 0")



N
C
A
R
A
I



Processing PDA Commands & Gestures

N
C
A
R
A
I



Go To

Palm-click:

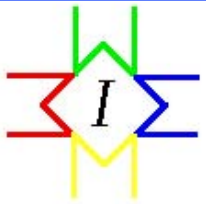
X = 30

Y = 120

(0x83)

(30 120)

(sending message: "17 30 120")





Integrated Inputs

N
C
A
R
A
I



GO OVER THERE



Palm-click:

X = 30

Y = 120

(IMPER (:VERB GO
(:AGENT (:SYSTEM YOU))
(:TO-LOC (HERE)))

(30 120)

(sending message: "17 30 120")

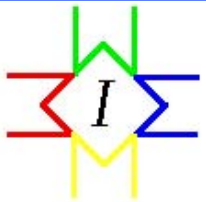
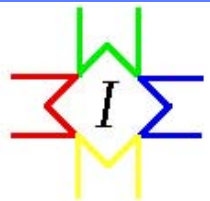
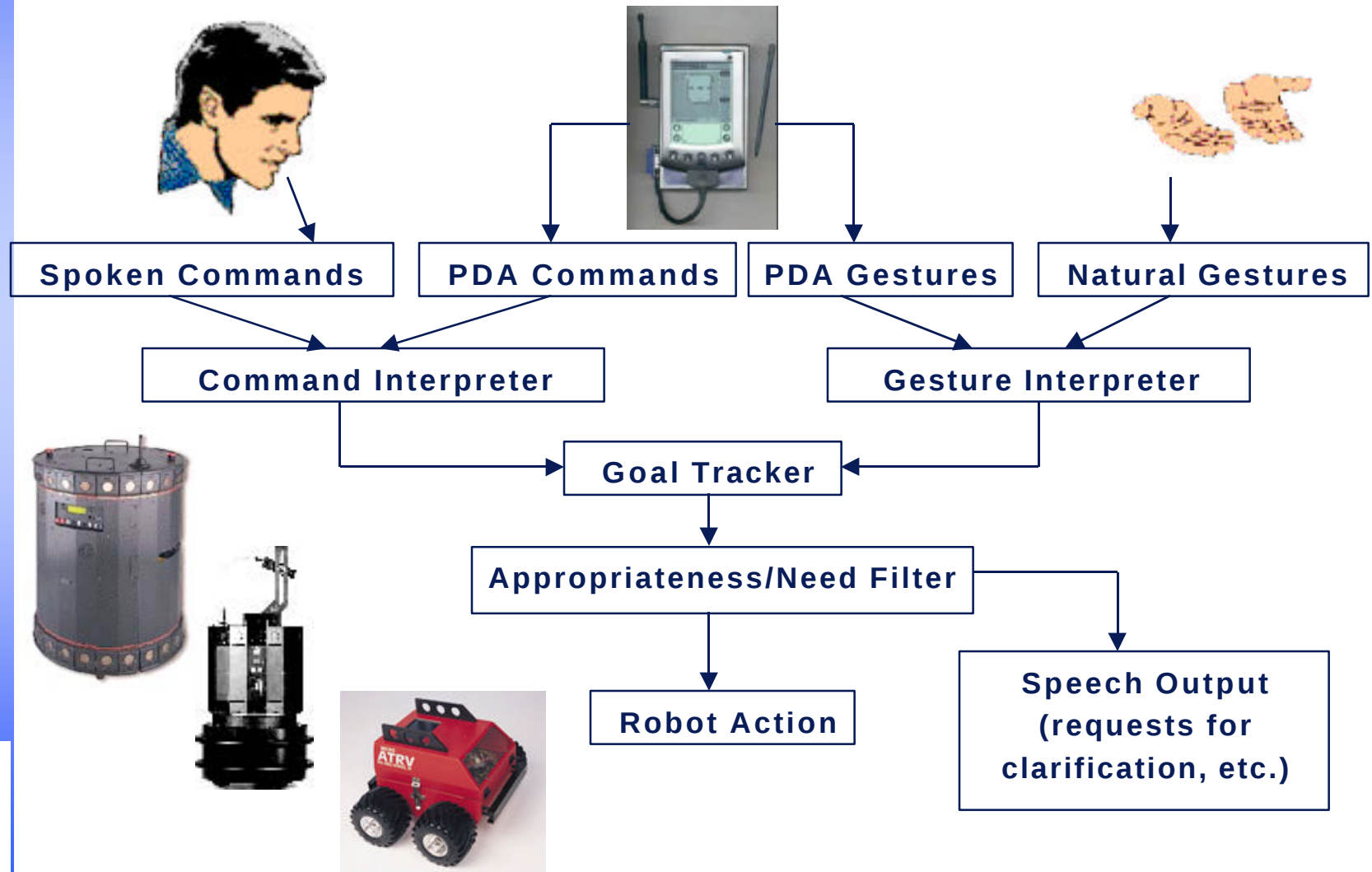




Diagram of Multi-modal Interface





TYPES of DATA CURRENTLY HANDLED

N
C
A
R
A
I

■ Simple commands

- Roadrunner, turn left.
- Coyote, go to waypoint 1.

◆ Line Segment commands (with/out natural/PDA gesture)

- Roadrunner, move back ten inches.
- Coyote, move up this far.

◆ Vectoring commands (with/out natural/PDA gesture)

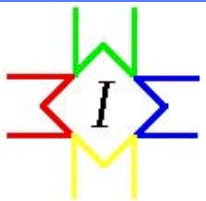
- Roadrunner, turn left 30 degrees.
- Coyote, turn right this far.

■ Complex commands (with/out natural/PDA gesture)

- Roadrunner/Coyote, go to the waypoint over there.

■ Interrupted sequences

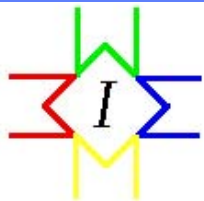
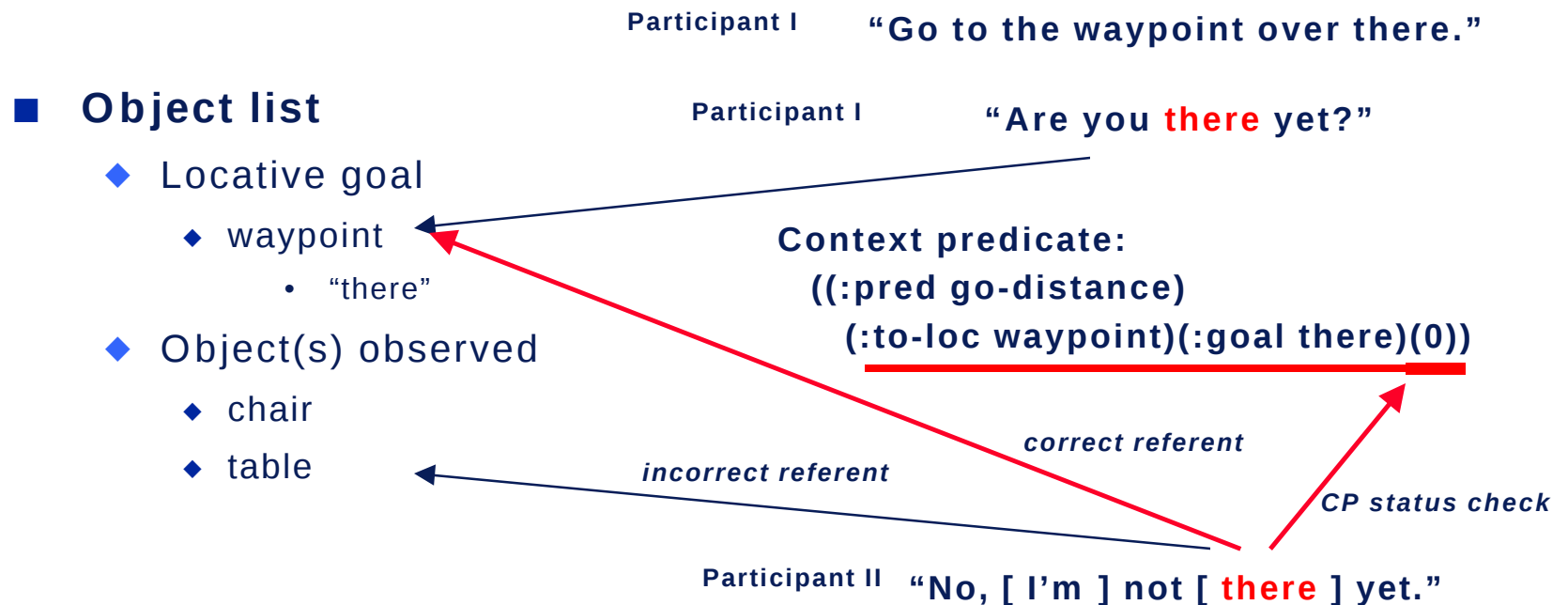
- | | |
|--------------------------|--|
| • Coyote, go over there. | • Roadrunner, go to waypoint 2. |
| • Where? | • Roadrunner, stop. |
| • Coyote, over there. | • Roadrunner, continue (to
waypoint 2/3). |





Disambiguating Locative Data

Systems having **robust vision capabilities** and **complex goal-directed activity** can have locative reference problems. As in the following dialog where the system sees something prior to the completion of a goal. Using a status check of the “context predicate” disambiguates the referent.

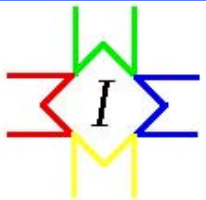




Data we still want to handle

- A typical dialog with two or more autonomous robots:
 - ◆ 1. Roadrunner, go to waypoint 1.
 - ◆ 2. Coyote, go to the door over there.
 - ◆ 3. Roadrunner, stop.
 - ◆ 4. Coyote, stop but now continue doing what Roadrunner was doing.
 - ◆ 5. Roadrunner, I want you to go to the door instead.
- Above interchange requires additional dialog capabilities
 - ◆ fill in “elliptical information”
 - ◆ “doing what Roadrunner was doing” (sentence 4 above).
 - ◆ disambiguate referents across the dialog,
 - ◆ “the door” in sentence 5.
- Self-appointed team membership
 - ◆ Given a particular goal or goals and several robots tasked to complete the goal(s), robots assign themselves to various teams to complete the task.

N
C
A
R
A
I





Additional Requirements: Interleaved Planning

N
C
A
R
A
I

■ Predicate List

- ◆ go to object <locative>
- ◆ stop
- ◆ go to object <table>
- ◆ pick up object <book>
- ◆ go to object <locative>

GOAL STACK (attained?)

YES
YES
YES

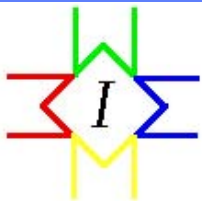
NO

NO

**situation, plan of action, goals unknown to speaker occur*

- ◆ **move obstacle <X>*
 - ◆ *<determine if moveable>*
 - *<if not, report>*
 - *<if moveable, determine how>*
 - ◆ *<if moveable by self, move obstacle <x>>*
 - ◆ *<if not, acquire assistance>*
 - *<move obstacle <x>*
 - *<re-deploy assistant>*
- ◆ go to object <locative>

YES
YES

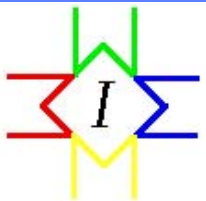




Conclusions

1. By using “context predicates” we track actions occurring during a dialog to determine which goals (event and locative) have been achieved or attained and which have not.
2. By tracking “context predicates” we can determine what actions need to be acted upon next; i.e. predicates in the stack that have not been completed.
3. “Locative” expressions, e.g. “there,” give us a kind of handle in command and control applications to attempt error correction when locative goals are being discussed.
4. By interleaving complex dialog with natural and mechanical gestures, we hope to achieve dynamic autonomy and an integrated multi-modal interface.

N
C
A
R
A
I

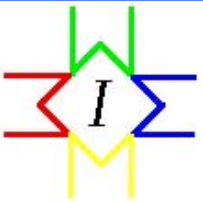




Future Plans

- Extend gesture recognition via better vision capabilities, etc.
- Integrate symbolic gestures with natural gestures
 - ASL, canine obedience, etc.
 - for use in noisy or secure environments
- Integrate 3D audio with multimodal interaction
 - orientation of speaker and “hearer” and directionality issues between participants
- Integrate speaker recognition via visual input
- Develop dialog-based planning for teams of dynamically autonomous robots

N
C
A
R
R
I





Video Clip

N
C
A
R
A
I

QuickTime™ and a
Cinepak decompressor
are needed to see this picture.