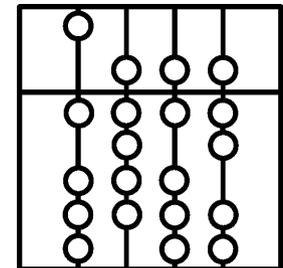


The iCat in the JAST Multimodal Dialogue System

Mary Ellen Foster
Technical University of Munich

First iCat Workshop
Eindhoven, 27 March 2005

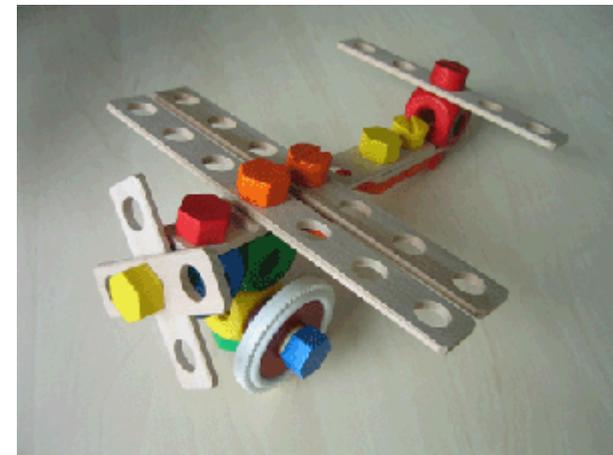


The JAST project

- ◆ “**J**oint **A**ction **S**cience and **T**echnology”
- ◆ <http://www.jast-net.gr/>
- ◆ Main objective: build jointly-acting autonomous systems that communicate and work intelligently on mutual tasks
- ◆ Research areas:
 - ◆ Cognitive and neural bases of joint action
 - ◆ Dialogue and joint action
 - ◆ Joint action in autonomous systems

The JAST dialogue system

- ◆ Task: robot and human jointly assemble *Baufix* construction toys
- ◆ Provides a testbed for implementing the results of the experimental joint-action studies



Target dialogue

- ◆ **User** Can you find a long slat?
- ◆ **JAST** What is a long slat?
- ◆ **U** A slat with five holes.
- ◆ **J** [*picks up a five-hole slat*]
- ◆ **U** You should insert the red bolt in the leftmost hole.
- ◆ **J** Which red bolt? There are three red bolts available.
- ◆ **U** Any red bolt.
- ◆ **J** [*picks up nearest red bolt*] Here is a red bolt. Can you insert it while I hold the long slat ?
- ◆ **U+J** [*action*]
- ◆ **U** We need a nut for this bolt.
- ◆ **J** Here it is. I'll hold the bolt and slat while you put the nut on them.

Current system



Roles of the iCat

- ◆ Feedback
 - ◆ Synthesised speech
 - ◆ Facial expressions
- ◆ Gaze control
 - ◆ User face tracking
 - ◆ Looking at objects on the table
- ◆ Blinking
 - ◆ “JustBlink” animation script (face only)
 - ◆ Send every 5 seconds, except while talking

Synthesised speech and facial expressions

- ◆ Voice: AT&T Natural Voices (SAPI 5)
- ◆ Expressions: built-in animation-module scripts, speech removed where necessary

CommandInput

```
load 3 Greet  
play 3 1  
set-var iCat.speech "Hallo, und willkommen bei Jast."
```

EventOutput

```
icat.speechevent -2  
[...]  
icat.speechevent -3
```

StatusOutput

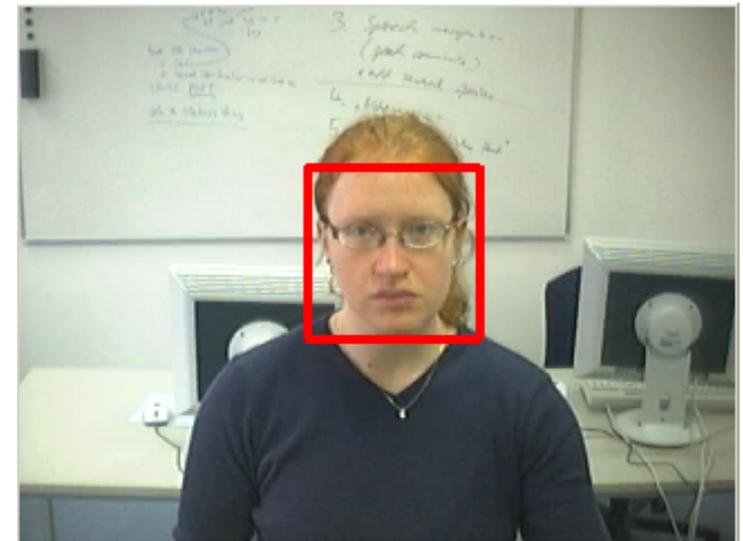
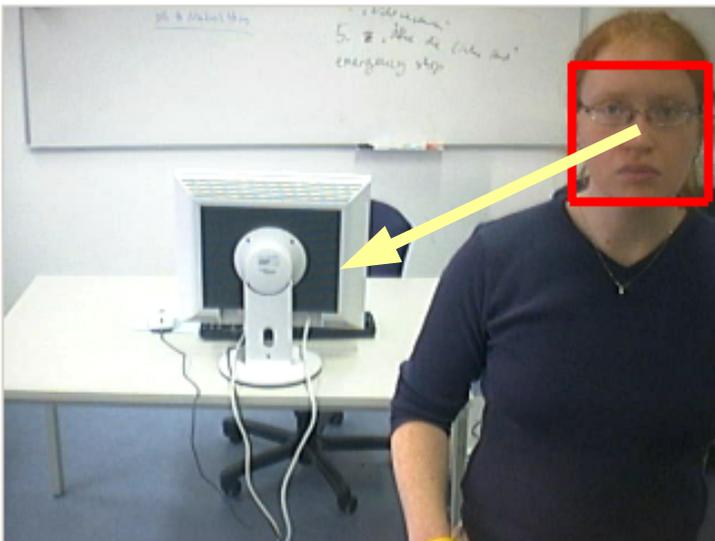
```
start 3 Greet  
stop 3 Greet
```



Either order

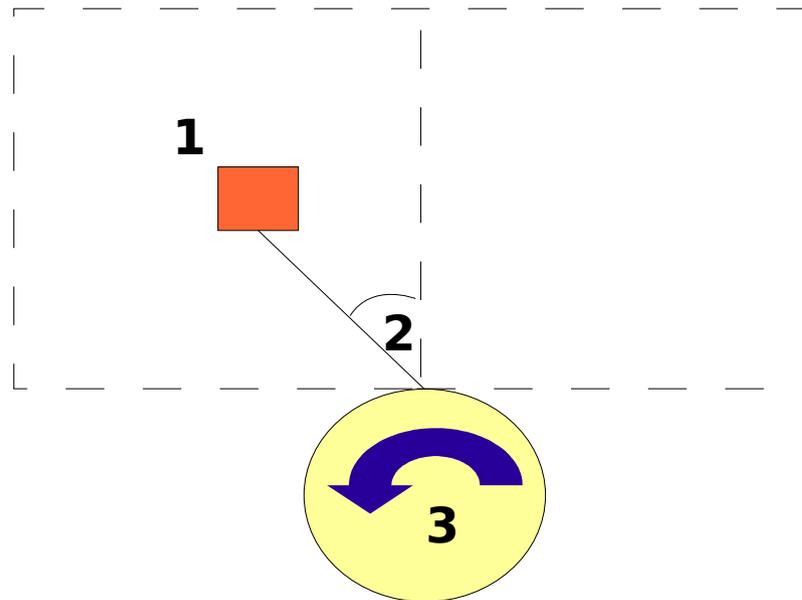
User face tracking

- ◆ OpenCV, using nose webcam
- ◆ Move head (*iCat.neck*, *iCat.body*) to put centre of user face at (160, 120)
 - ◆ $\text{newPos} = \text{curPos} - (\text{diff}/\text{SCALE})$
 - ◆ move cat if $|\text{newPos} - \text{curPos}| > \text{EPSILON}$



Looking at table objects

- ◆ Look at an object when it is used (picked up, put down, etc.)
 1. (x,y) from overhead camera
 2. Angle from centre
 3. Map to *iCat.Body* value ($45^\circ = 100$)



Implementation issues

- ◆ Integration with external event loop
 - ✓ Process OAA events within *vDoAction*
- ◆ Combination of speech and face motion
 - ✓ Wait for both to finish before continuing
- ◆ Coordination across output channels
 - ✓ Disable blinking and gaze during speech
- ◆ Interaction of PVM and Cygwin SSHD
 - ✓ Run SSH server as desired user
- ◆ Compiling with Eclipse+Ant

Next steps

- ◆ Coordination of facial motions with parts of the utterance
- ◆ More sophisticated gaze control
- ◆ Other forms of non-verbal feedback (e.g., nodding)
- ◆ Implement findings from dialogue experiments

Wish list

- ◆ Relative motion in animation scripts
- ◆ Animation-module events on SAPI
bookmarks in speech
- ◆ Controllable speed on neck and body
set-var commands
- ◆ Java API
- ◆ Support for Linux
- ◆ Lips that don't fall off :)