Multimodal
Human-Computer Interaction

The design and implementation of a prototype

by Maurits André

TU Delft

CAIP

THE STATE UNIVERSITY OF NEW JERSEY
RUTGERS

Overview

- Introduction
  - Problem statement
- Technologies used
  - Speech
  - Hand gesture input
  - Gazetracking
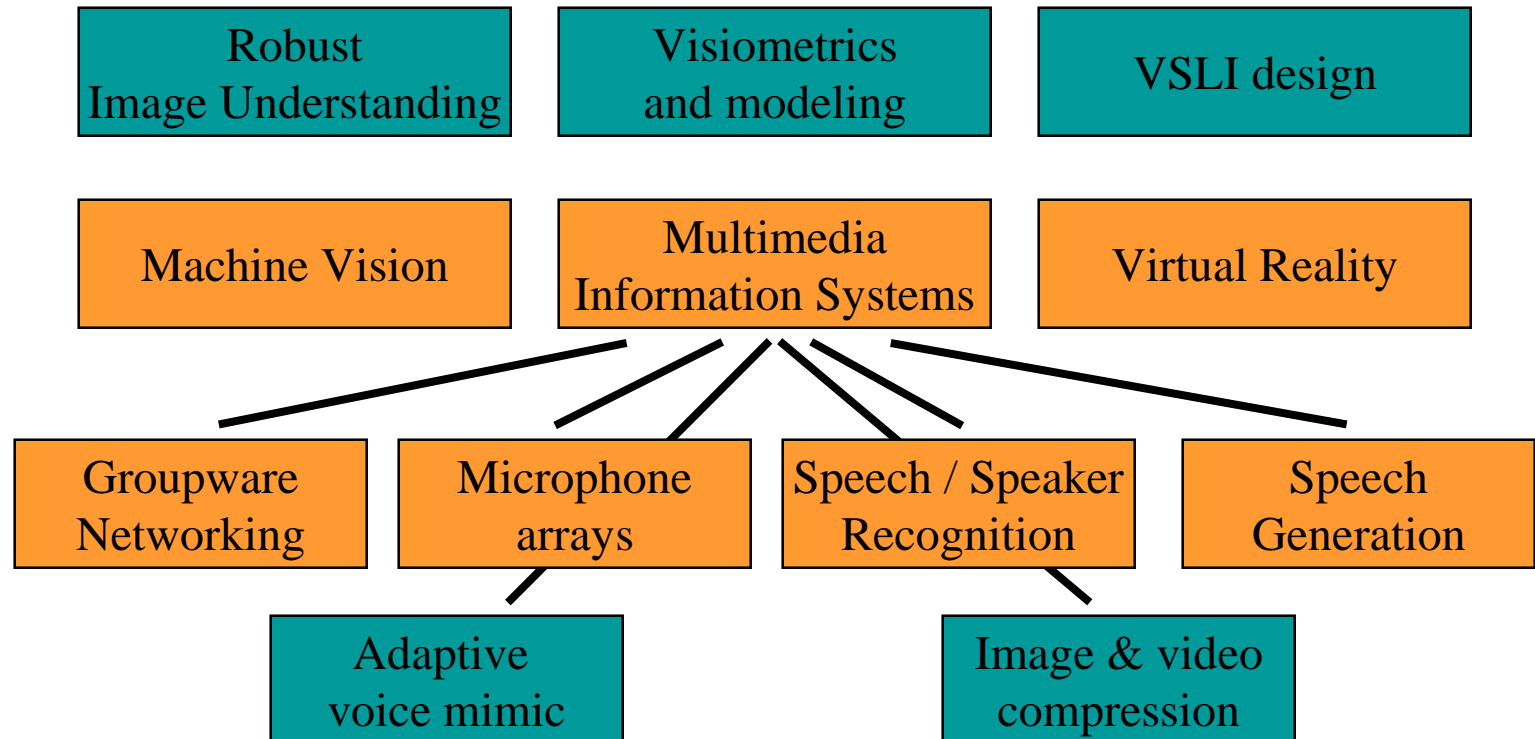- Design of the system
- Multimodal issues

Overview

- Testing
  - program tests
  - usability tests
  - human factors studies
- Conclusions and recommendations
- Future work
- Video

CAIP

Center for Computer Aids in Industrial Productivity
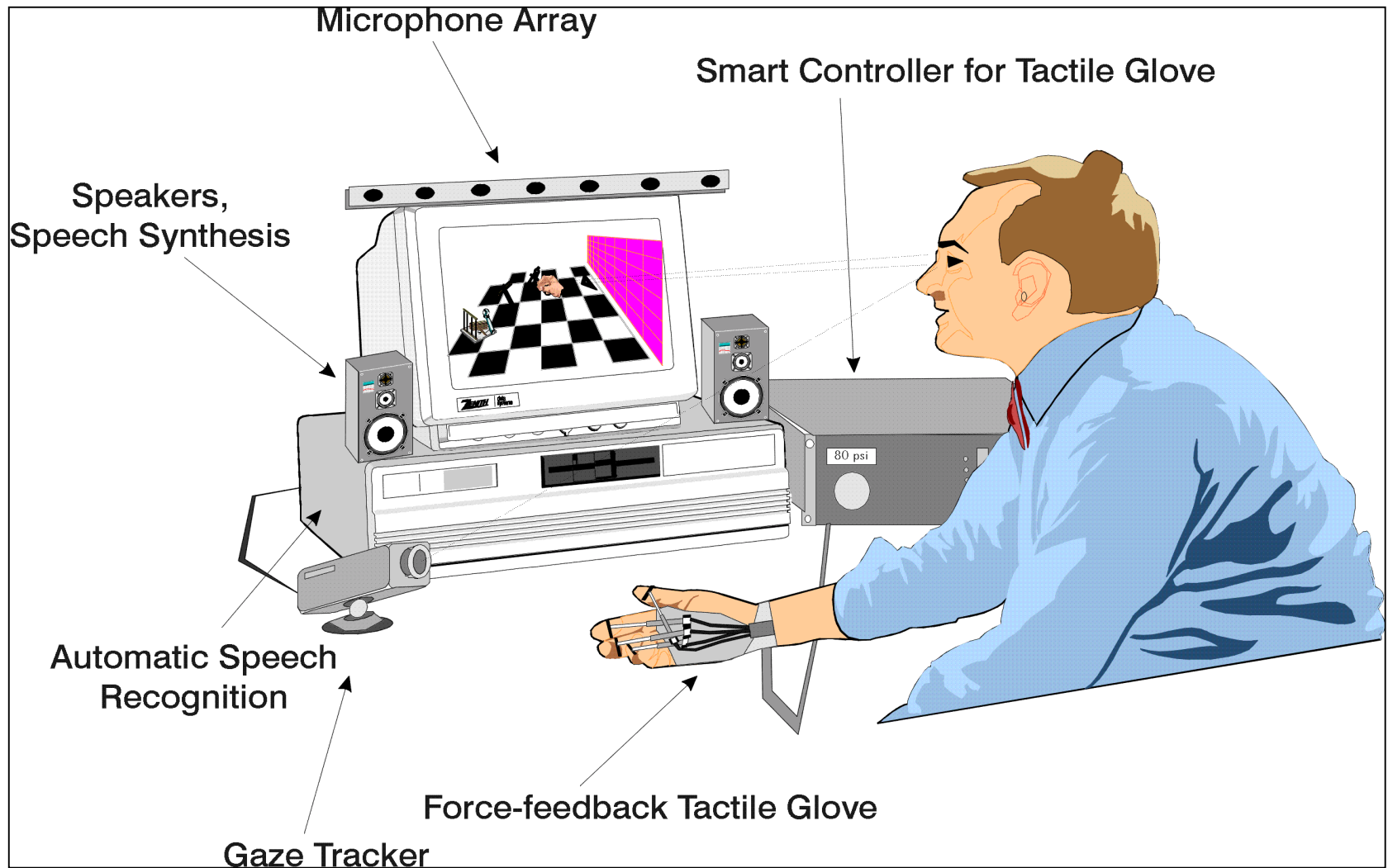
Prof. James L. Flanagan

Multimodal HCI

- Currently: mouse, keyboard input

- More natural communication technologies available:
  - sight
  - sound
  - touch

- Robust and intelligent combination of these technologies

Aim

Microphone Array

Smart Controller for Tactile Glove

Speakers,
Speech Synthesis

Automatic Speech
Recognition

Force-feedback Tactile Glove

Gaze Tracker

80 psi

Problem statement

- Study three technologies:
    - Speech recognition and synthesis
    - Hand gesture input
    - Gazetracking

- Design prototype (appropriate application)

- Implement prototype

- Test and debug

- Human performance studies

SR and TTS

- Microsoft Whisper system (C++)
- Speaker independent
- Continuous speech
- Restricted task-specific vocabulary (150)
- Finite state grammar
- Sound capture: microphone array

Hand gesture input

- Advantages:
  - Natural
  - Powerful
  - Direct
- Disadvantages:
  - Fatigue (RSI)
  - Learning
  - Non-intentional gestures
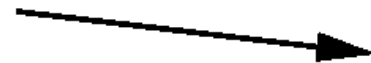 - Lack of comfort
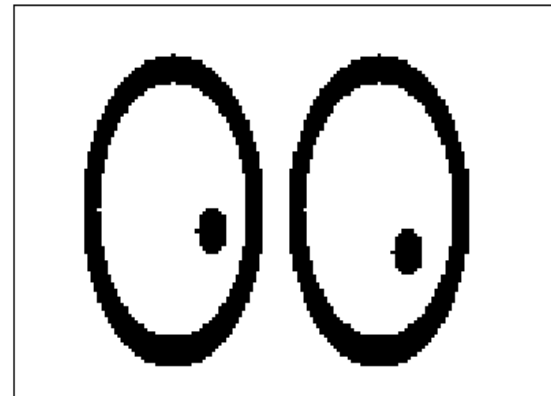
Force feedback tactile glove
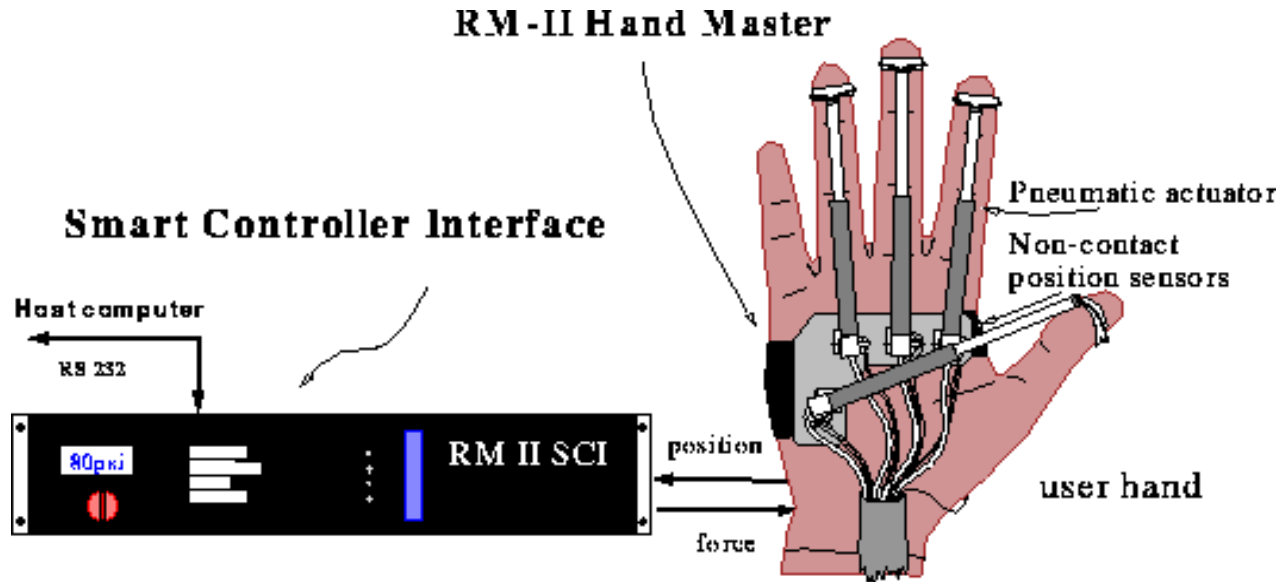
- Polhemus tracker for wrist position/orientation
- 5 gestures are recognized

Implemented gestures

- Grab   `"Move this"`
- Open hand   `"Put down"`
- Point at an object   `"Select"`

                         `"Identify"`
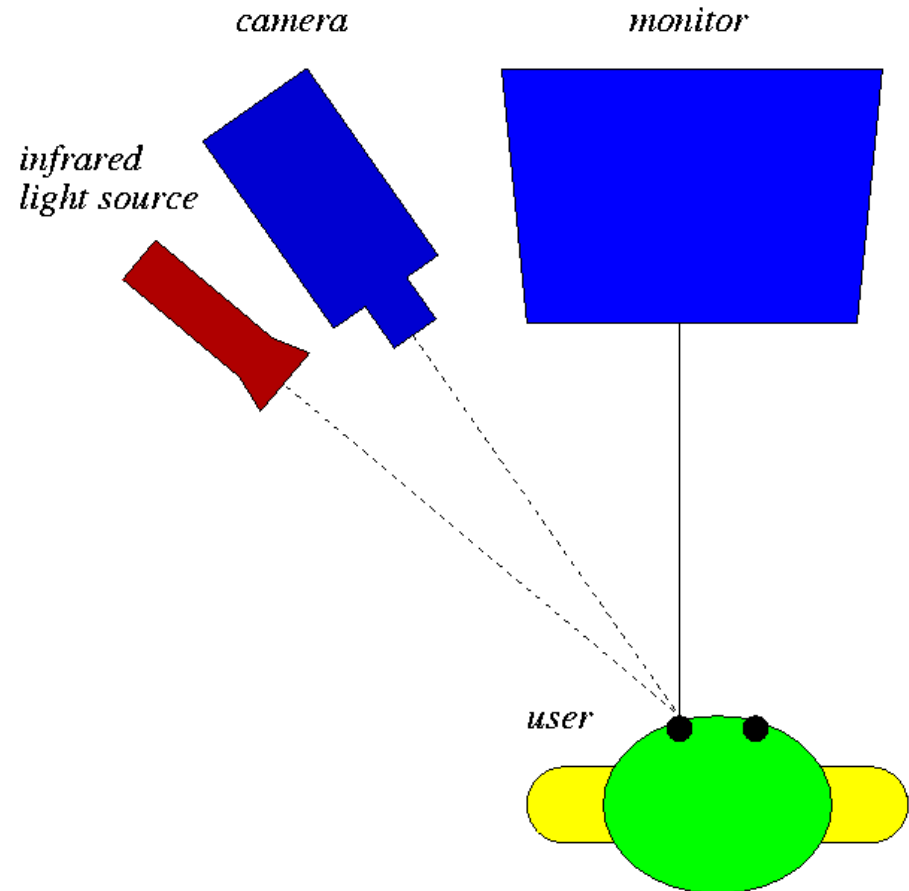
- Thumb up   `"Resize this"`

Eyes output

- Direction of gaze
- Blinks
- Closed eyes
- Part of emotion

Gazetracker

- ISCAN RK-726 gazetracker
- 60 Hz.
- Calibration

Application

- Requirements:
  - Multi-user, collaborative
  - Written in Java
  - Simple
- Choice:
  - Drawing program
  - Military mission planning system

Drawing program
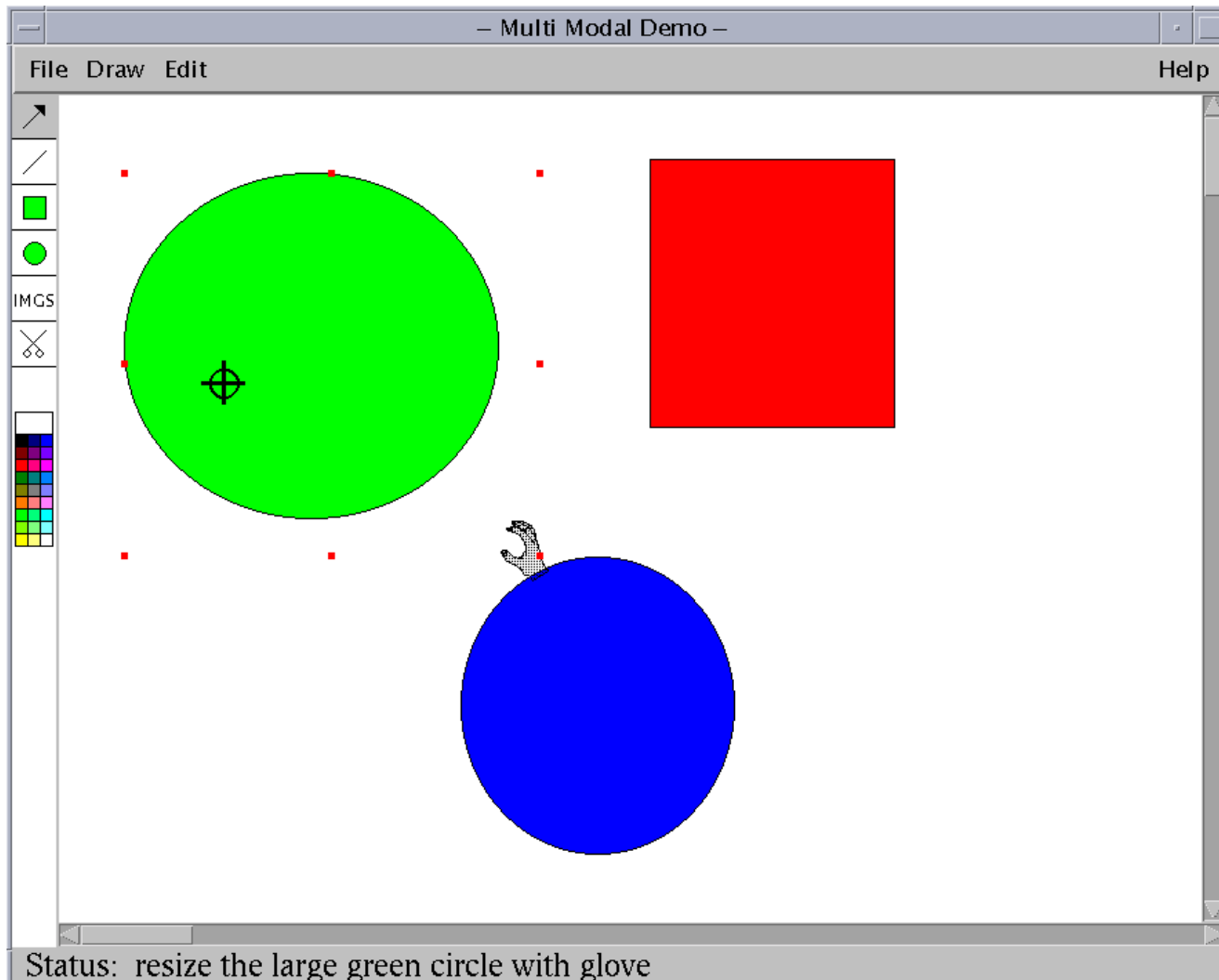
Military mission planning

Frames

- Slots
- Inheritance
- Generic properties
- Default values

Create Figure	
Type	circle/rect/line
Color	[white]/red/grn/...
Height	0..max_y
Width	0..max_x
Location	(x,y)

Command Frame

EXIT

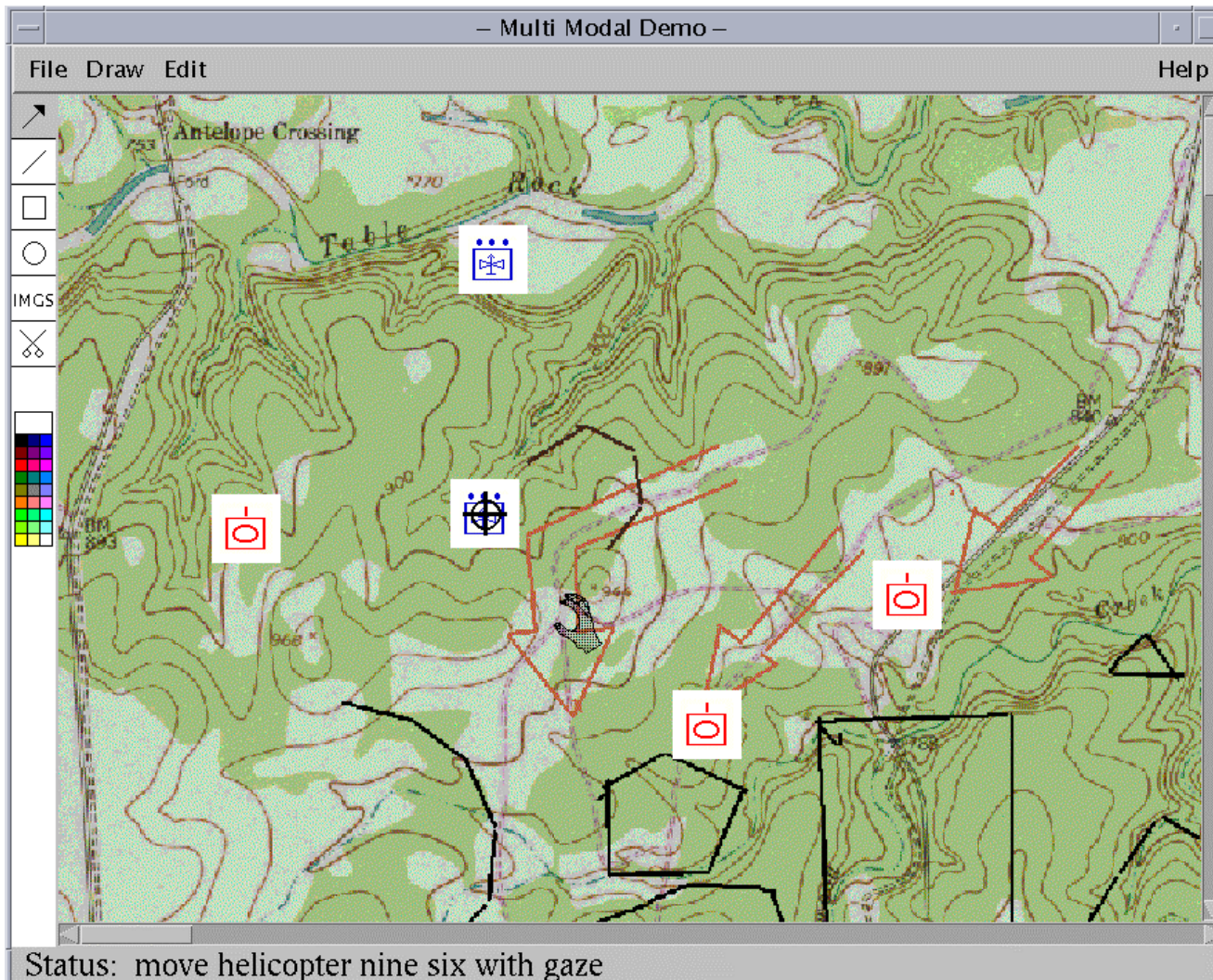| Confirm	YES

Cmd. with object

Object ID	0...max_int
Confirm	NO

Create Figure

Type	circle/rect/line
Color	[white]/red/grn/...
Height	0..max_y
Width	0..max_x
Location	(x,y)

Move

Destination	(x,y)

Design

Gazetracking

Speech Recognition

Gesture Recognition

Speech Synthesis

Fusion Agent

Collaborative Workspace
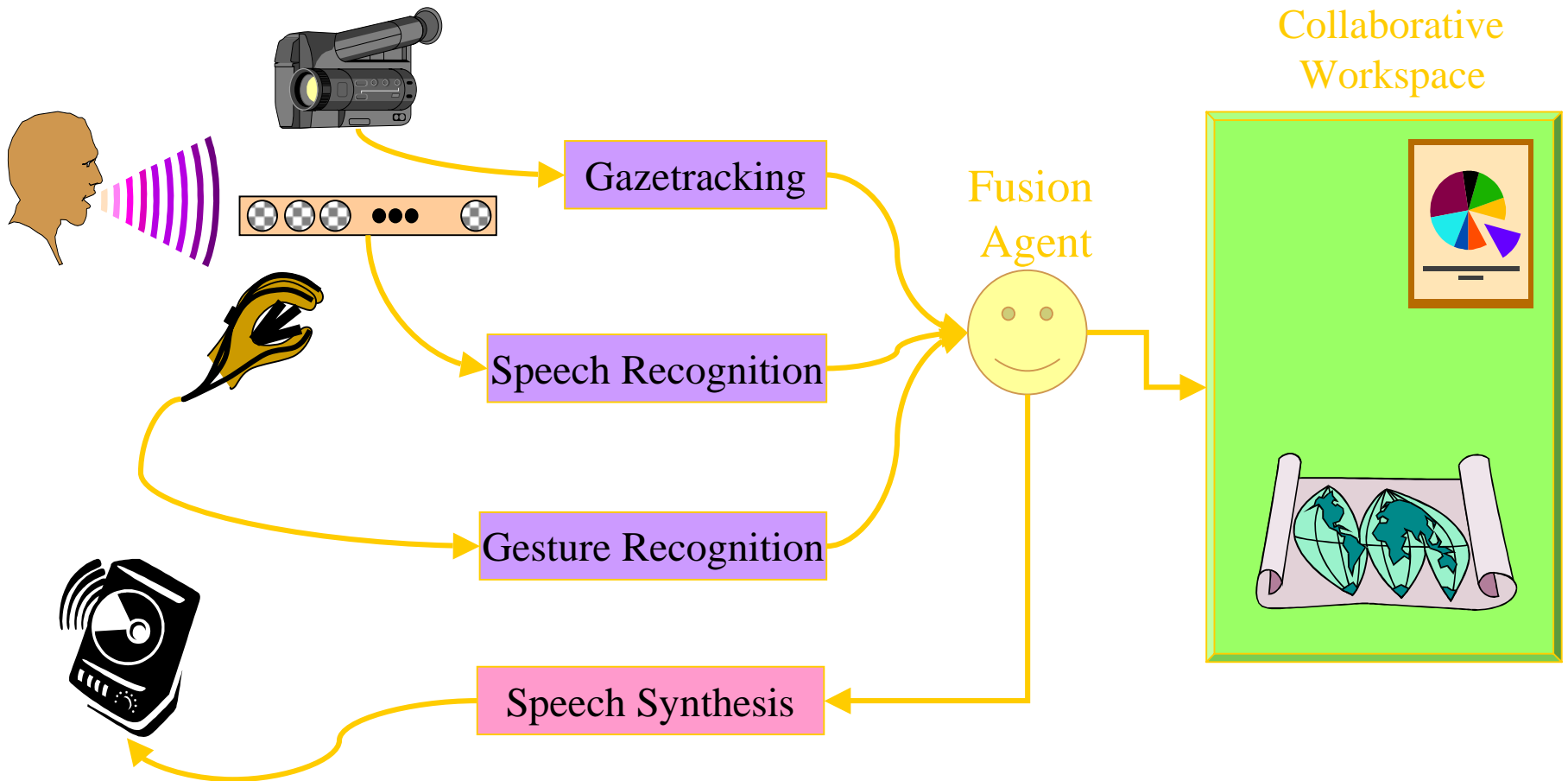
Fusion Agent

Example:"Move tank seven here."

(x1,y1)   (x2,y2)   (x3,y3)   (x4,y4)

Slot Buffer

Control

Parser

Frame: MOVE
object | Tank 7
where | (x4,y4)

Rule Based
Feedback
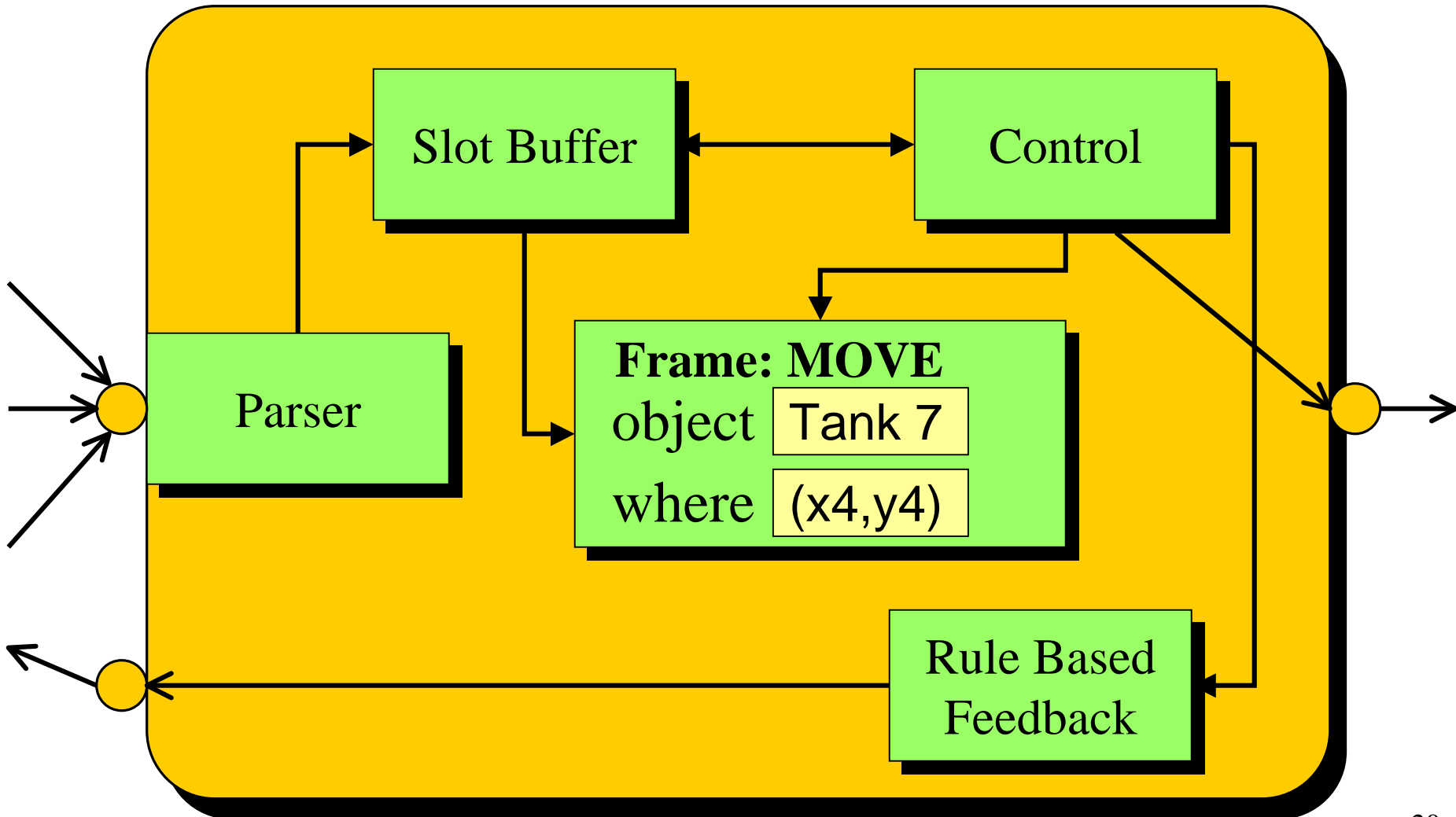
Classification of feedback

- ## Confirmation

  – Exit the system.   Are you sure?

- ## Information retrieval

  – What is this?   This is tank seven.

  – Where is tank nine. *Visual feedback.*

- ## Missing data

  – Create tank.   Need to specify an ID for a tank.

- ## Semantic error

  – Create tank seven.   Tank seven already exists.

  – Resize tank nine.   Cannot resize a tank.

Multimodal issues

- ## Referring to objects

  - describing in speech: *Move the big red circle*
  - using anaphora: *Move it*
  - by glove
  - gaze + pronoun: *Delete this*
  - glove + pronoun: *Delete this*

- ## Timestamps

  Create a   red rectangle from here to   here

  T1 T2 T3   T4   T5   T6 T7 T8

  xy1 xy2 xy3   xy4   xy5   xy6 xy7 xy8

Multimodal issues

- ## Ambiguity
    - saying x, looking at y   »»   x
    - saying x, pointing at y   »»   x
    - looking at x, pointing at y   »»   x
    - saying x, gesturing y   »»   xy or yx

- ## Redundancy
    - saying x, looking at x   »»   x
    - etc.

Program testing

- Implementation in Java

- Program testing and debugging

  - Module testing
  - Integration testing
  - Configuration testing
  - Time testing
  - Recovery testing

Testing

- Usability tests
  - Demonstration with military personnel
  - Human factors study:
    - Script for user
    - Questionnaire for user
    - Tables for observer
    - Log-file for observer

Lab

Conclusions:

Selecting

Modality	Accuracy	Speed	Learning
Speech	++	++	–
Gaze	+	++	++
Glove	+	+	+
Mouse	++	–	+

Conclusions / recommendations

- ## Speech
  - real-time
  - timestamps
  - grammar in help file
  - low error rate
  - misunderstanding

- ## Glove
  - real-time
  - low precision
  - 2D »» 3D
  - fatigue
  - non-intentional gesture
  - limited number of gestures

- ## Gaze
  - real-time
  - self-calibration
  - face tracker
  - head movements
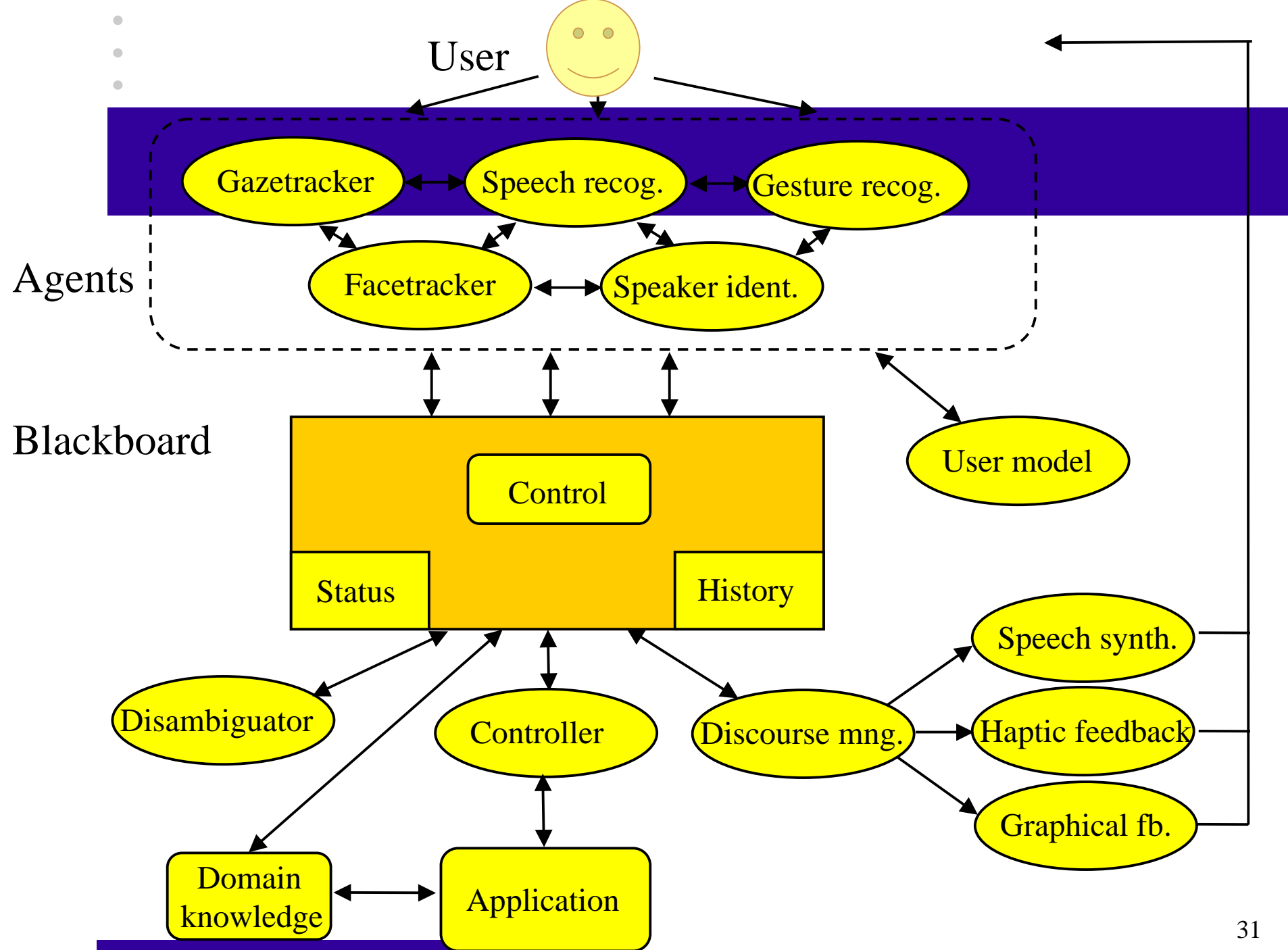 - jumpiness of eye movements
  - object of interest

General remarks

- Response time within 1 sec.

- Instruction, help files

- Application effective but limited

Future work

- Human performance studies
- Conversational interaction
- Context-based reasoning and information retrieval
- New design:

User

Agents

Blackboard

Gazetracker ⟷ Speech recog. ⟷ Gesture recog.

Facetracker ⟷ Speaker ident.

Control

Status   History

User model

Disambiguator   Controller   Discourse mng.   Speech synth.

Haptic feedback

Domain knowledge ⟷ Application   Graphical fb.

31

Problem statement

- Study three technologies:
  - Speech recognition and synthesis
  - Hand gesture input
  - Gazetracking
- Design prototype (appropriate application)
- Implement prototype
- Test and debug
- Human performance studies