Human – Robot Communication

Paul Fitzpatrick

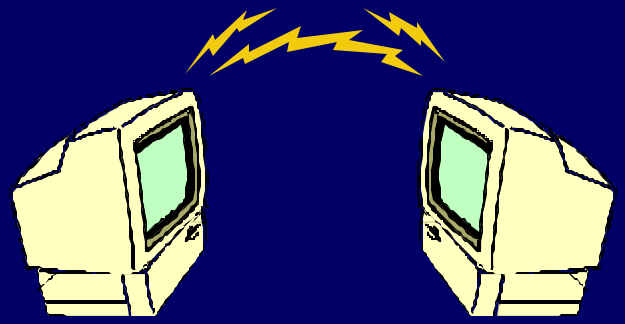
Human – Robot Communication

- Motivation for communication
- Human-readable actions
- Reading human actions
- Conclusions

Motivation

- What is communication for?
  - Transferring information
  - Coordinating behavior
- What is it built from?
  - Commonality
  - Perception of action
  - Protocols

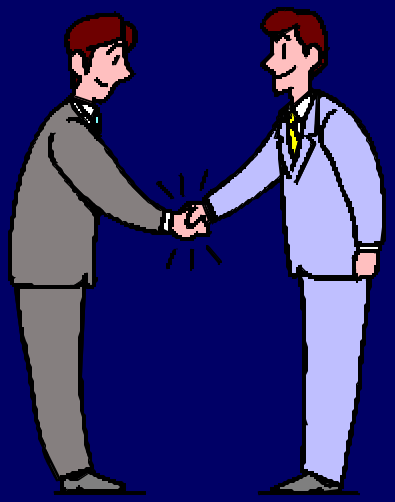
Communication protocols



Computer – computer protocols

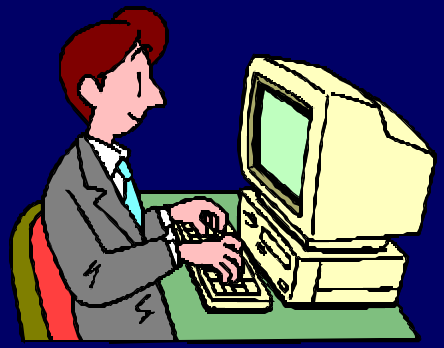
TCP/IP, HTTP, FTP, SMTP, …

Communication protocols



Human – human protocols

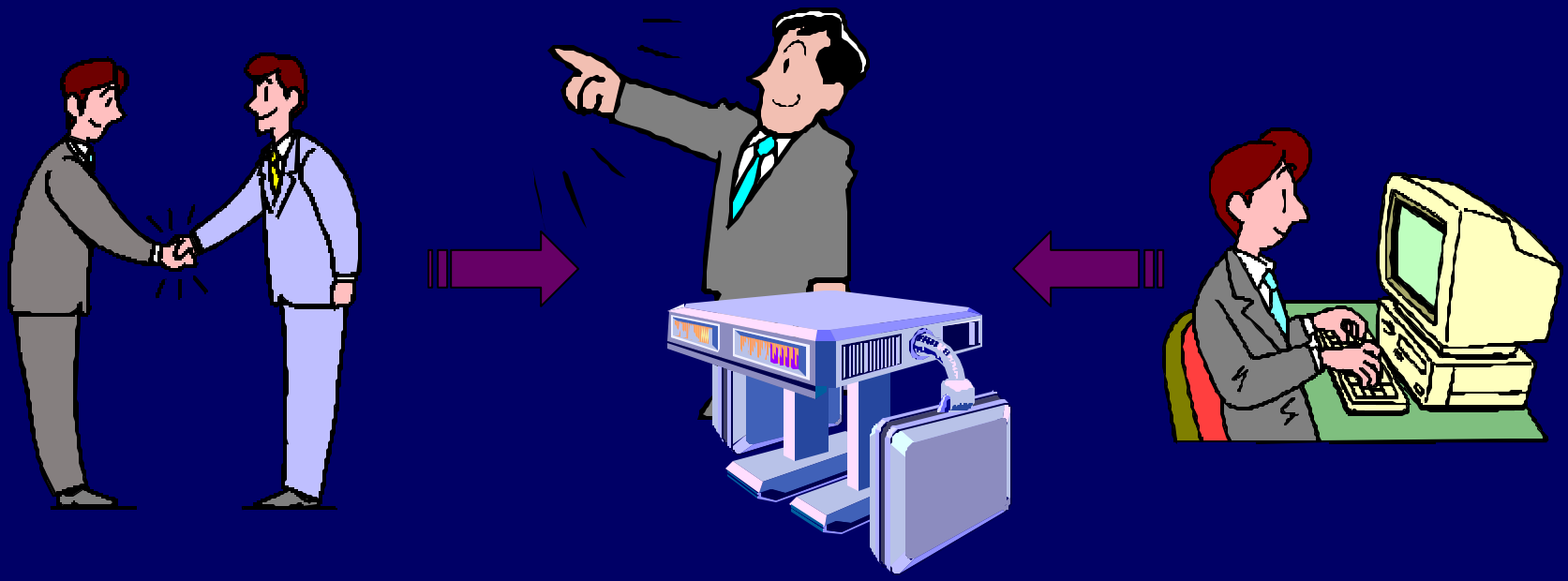
Initiating conversation, turn-taking, interrupting, directing attention, …



Human – computer protocols

Shell interaction, drag-and-drop, dialog boxes, …

Communication protocols



Human – human protocols

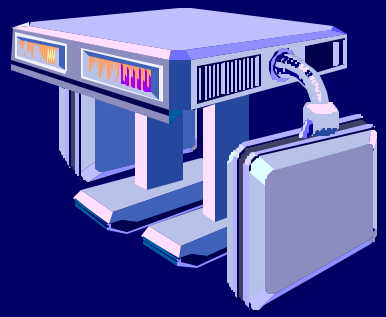
Initiating conversation, turn-taking, interrupting, directing attention, …

Human – robot protocols

Human – computer protocols

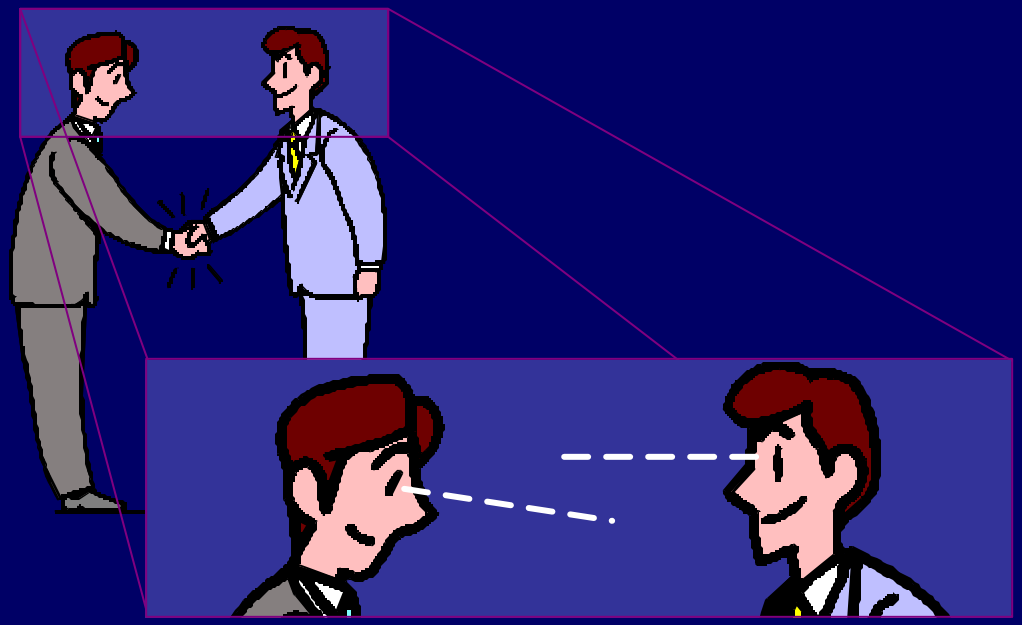
Shell interaction, drag-and-drop, dialog boxes, …

Requirements on robot



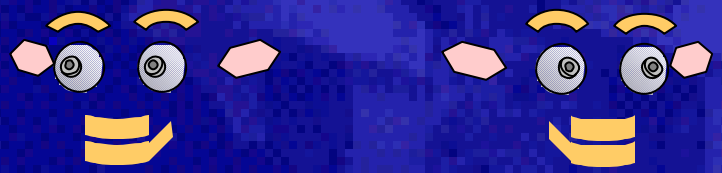
- Human-oriented perception
  - Person detection, tracking
  - Pose estimation
  - Identity recognition
  - Expression classification
  - Speech/prosody recognition
  - Objects of human interest
- Human-readable action
  - Clear locus of attention
  - Express engagement
  - Express confusion, surprise
  - Speech/prosody generation

Example: attention protocol



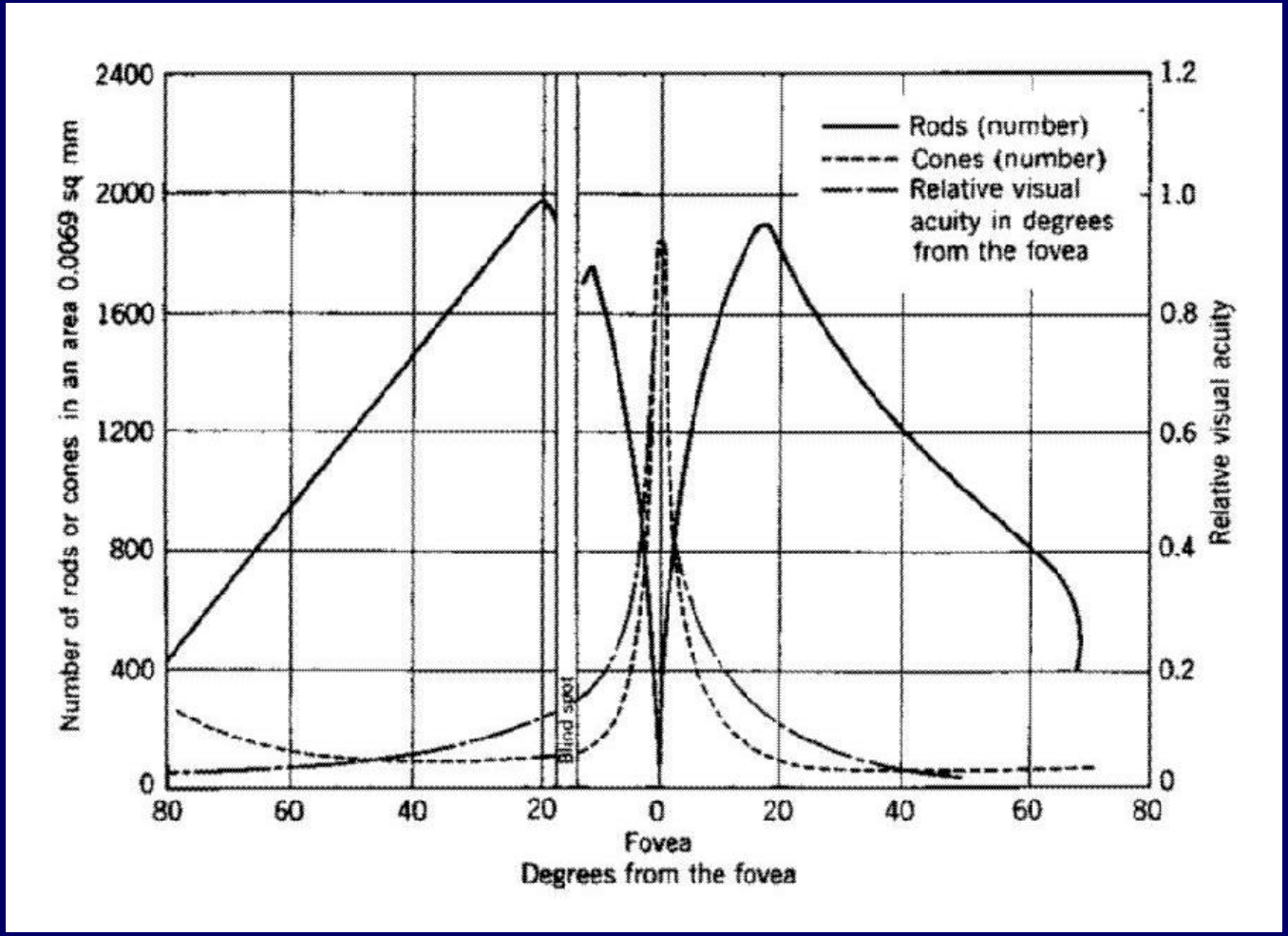
- Expressing attention
- Influencing other's attention
- Reading other's attention

Foveate gaze



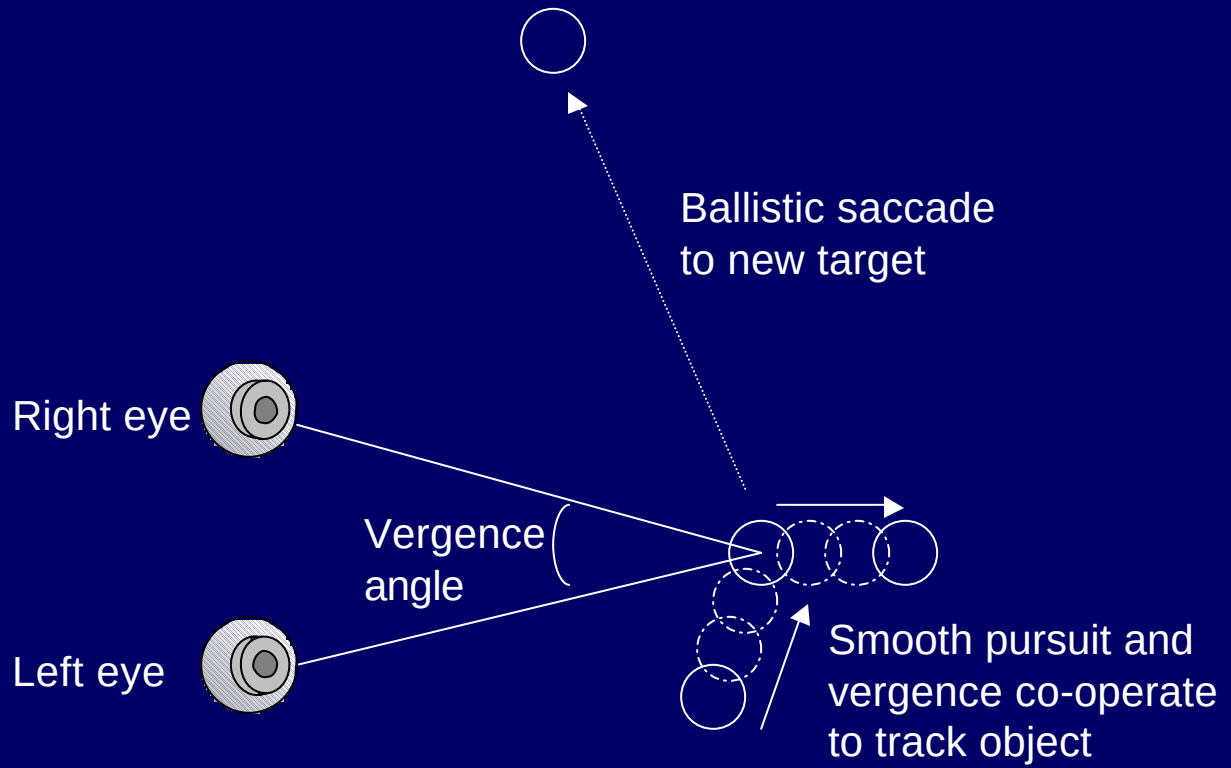
- Motivation for communication
- Human-readable actions
- Reading human actions
- Conclusions

Human gaze reflects attention



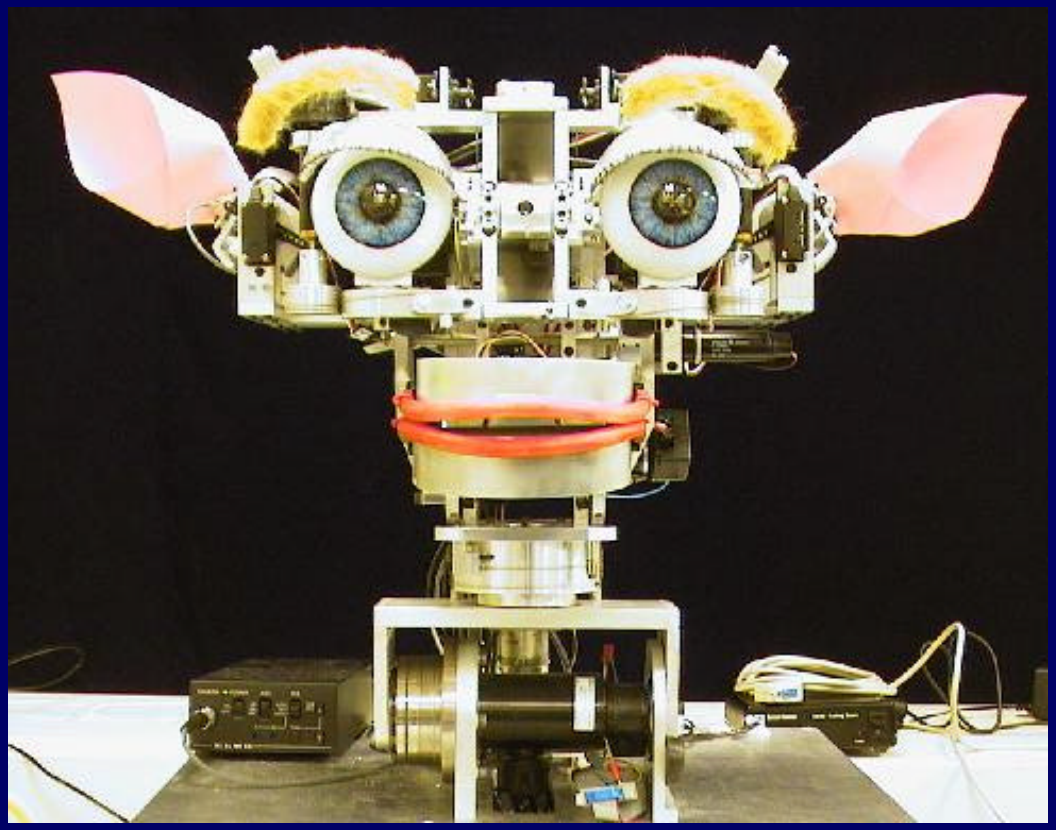
(Taken from C. Graham, "Vision and Visual Perception")

Types of eye movement



(Based on Kandel & Schwartz, "Principles of Neural Science")

Engineering gaze

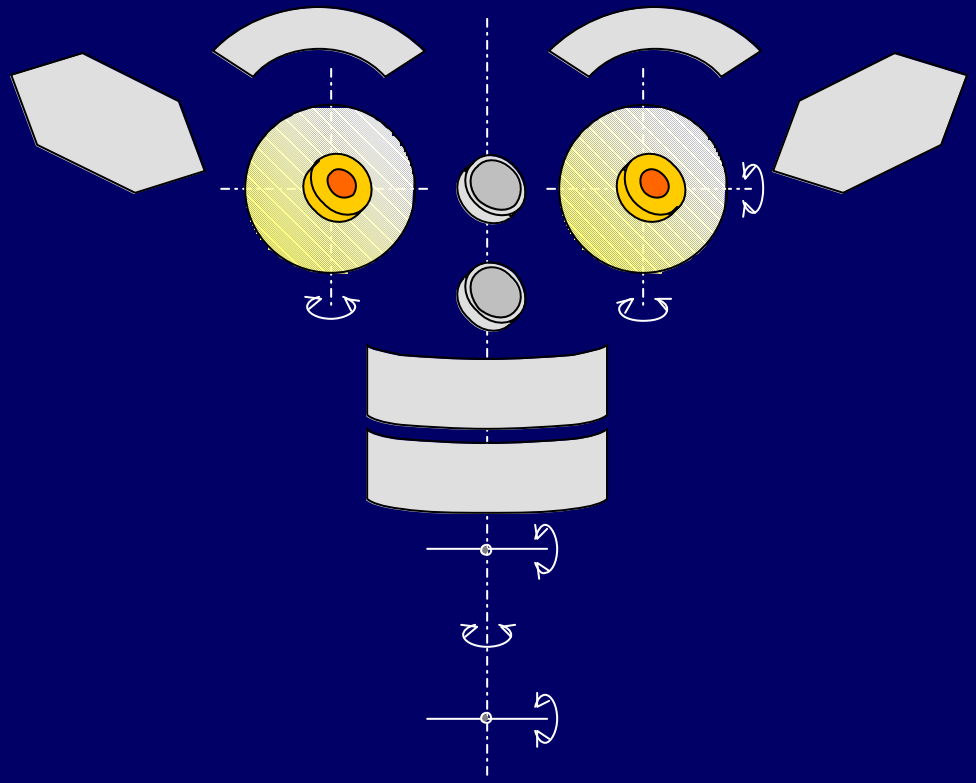


Kismet

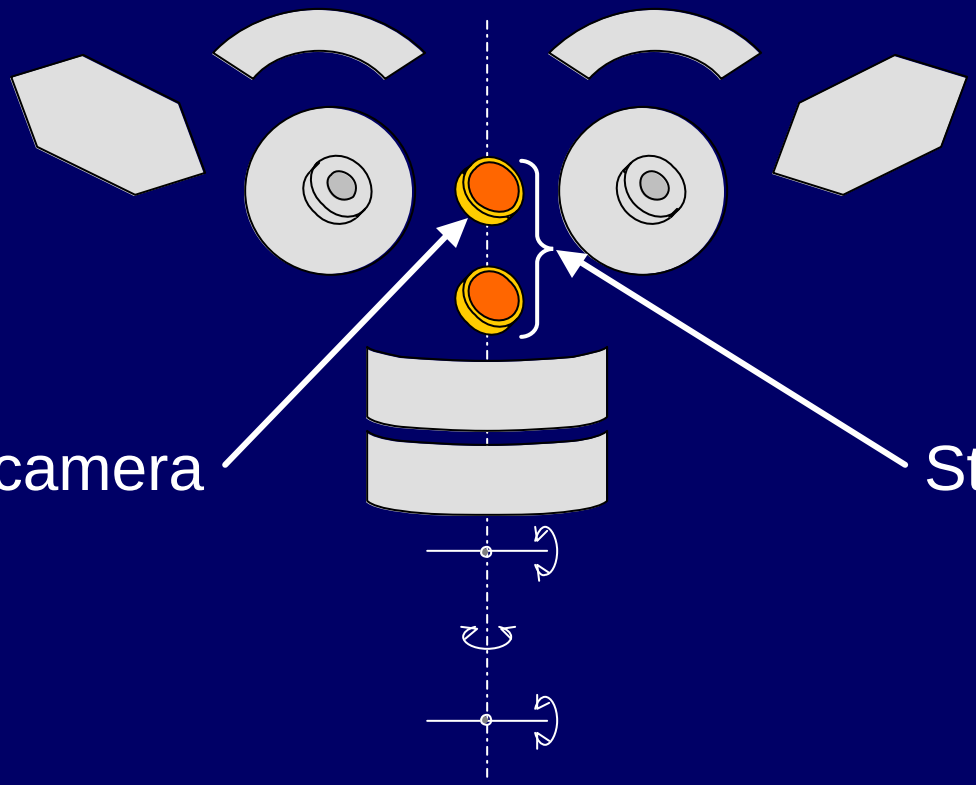
Collaborative effort

- Cynthia Breazeal
- Brian Scassellati
- And others
- Will describe components I’m responsible for

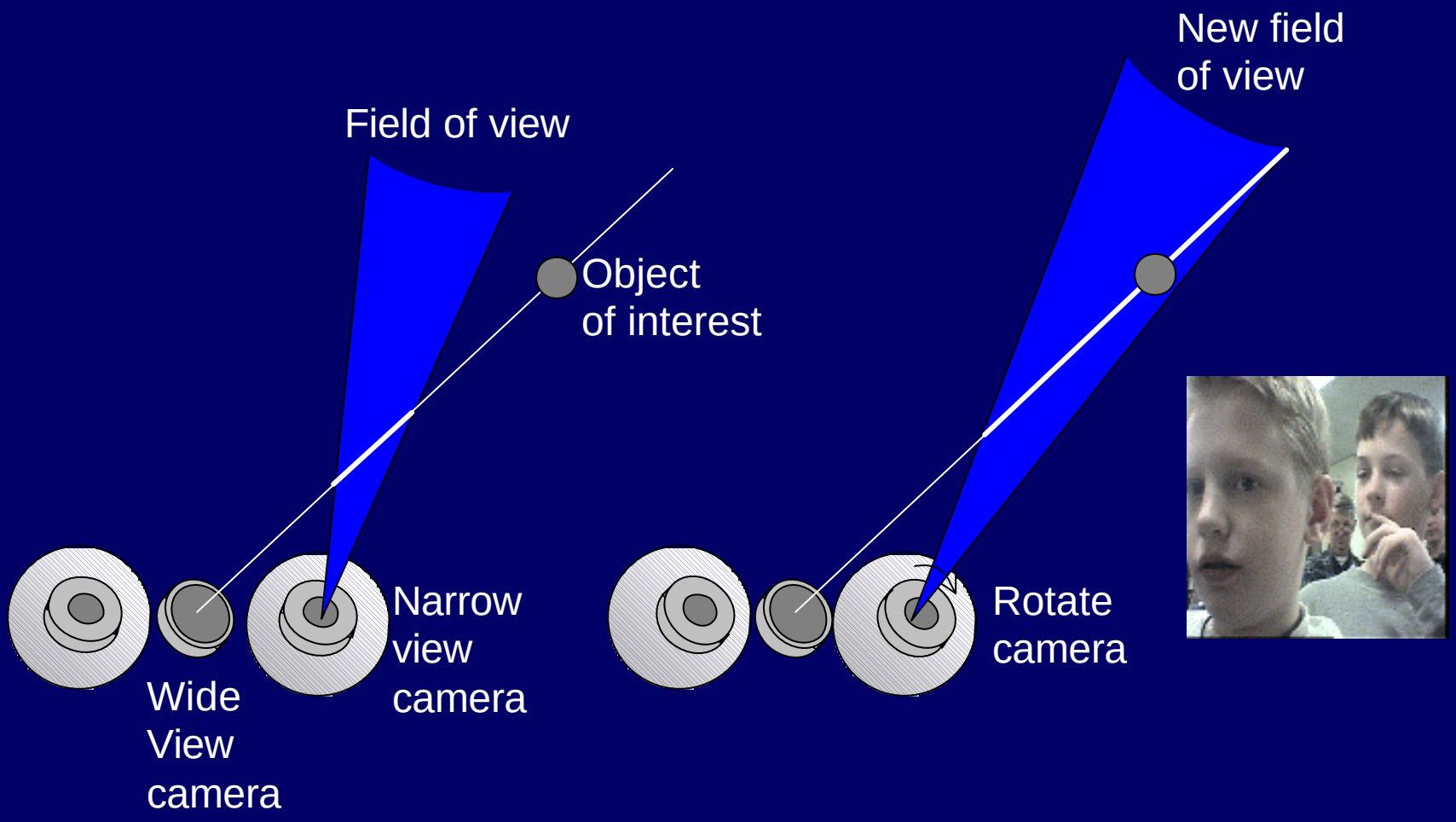
Engineering gaze



Engineering gaze



Tip-toeing around 3D

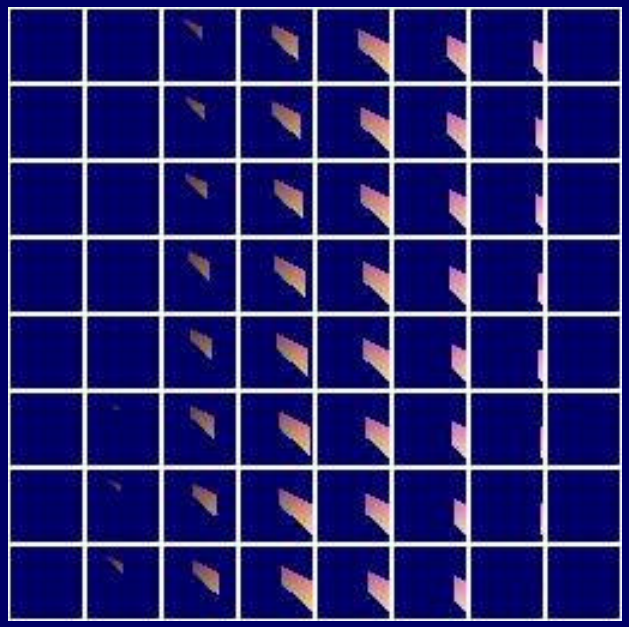


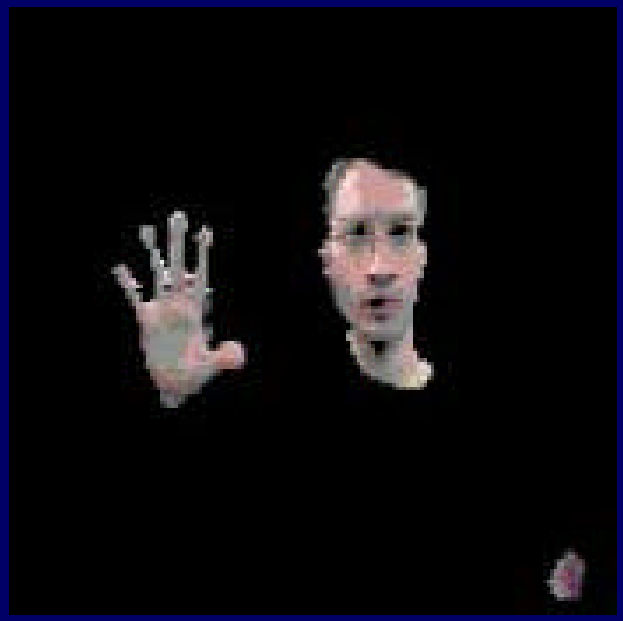
Example

Influences on attention

Influences on attention

- Built in biases





Influences on attention

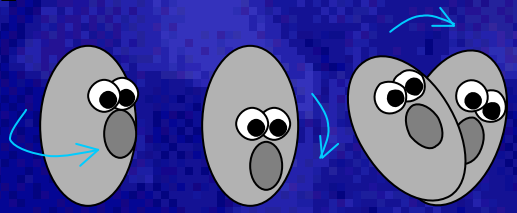
- Built in biases
- Behavioral state

Influences on attention

- Built in biases
- Behavioral state
- Persistence

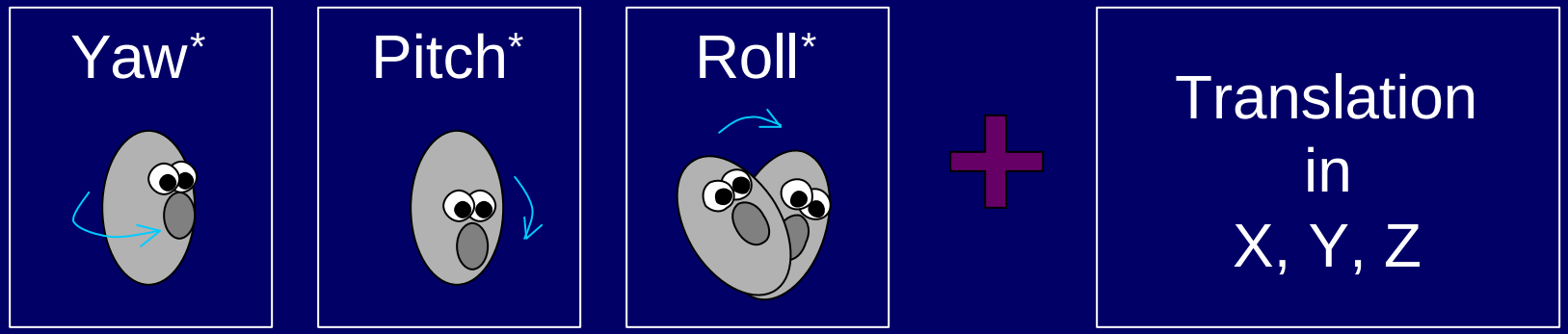
Directing attention

Head pose estimation



- Motivation for communication
- Human-readable actions
- Reading human actions
- Conclusions

Head pose estimation (rigid)



* Nomenclature varies

Head pose literature

Horprasert, Yacoob, Davis ’97

McKenna, Gong ’98

Wang, Brandstein ’98

Basu, Essa, Pentland ’96

Harville, Darrell, et al ’99

Head pose: Anthropometrics

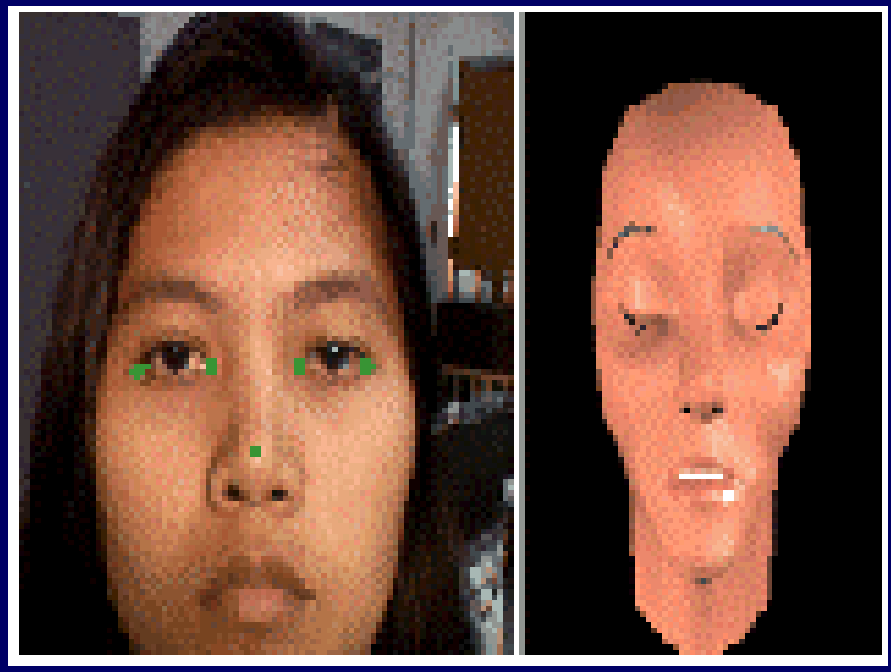
Horprasert, Yacoob, Davis

McKenna, Gong

Wang, Brandstein

Basu, Essa, Pentland

Harville, Darrell, et al



Head pose: Eigenpose

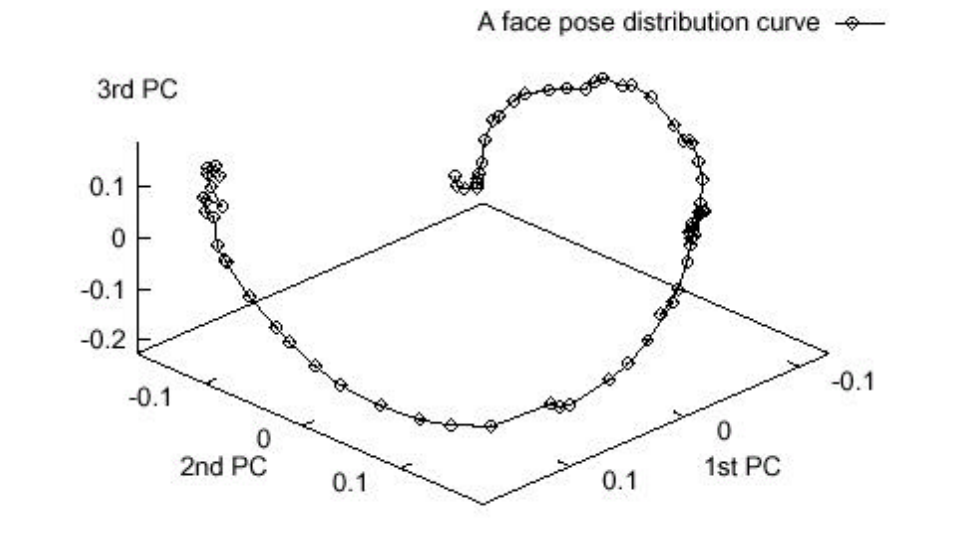
Horprasert, Yacoob, Davis

McKenna, Gong

Wang, Brandstein

Basu, Essa, Pentland

Harville, Darrell, et al



Head pose: Contours

Horprasert, Yacoob, Davis

McKenna, Gong

Wang, Brandstein

Basu, Essa, Pentland

Harville, Darrell, et al

Head pose: mesh model

Horprasert, Yacoob, Davis

McKenna, Gong

Wang, Brandstein

Basu, Essa, Pentland

Harville, Darrell, et al

Head pose: Integration

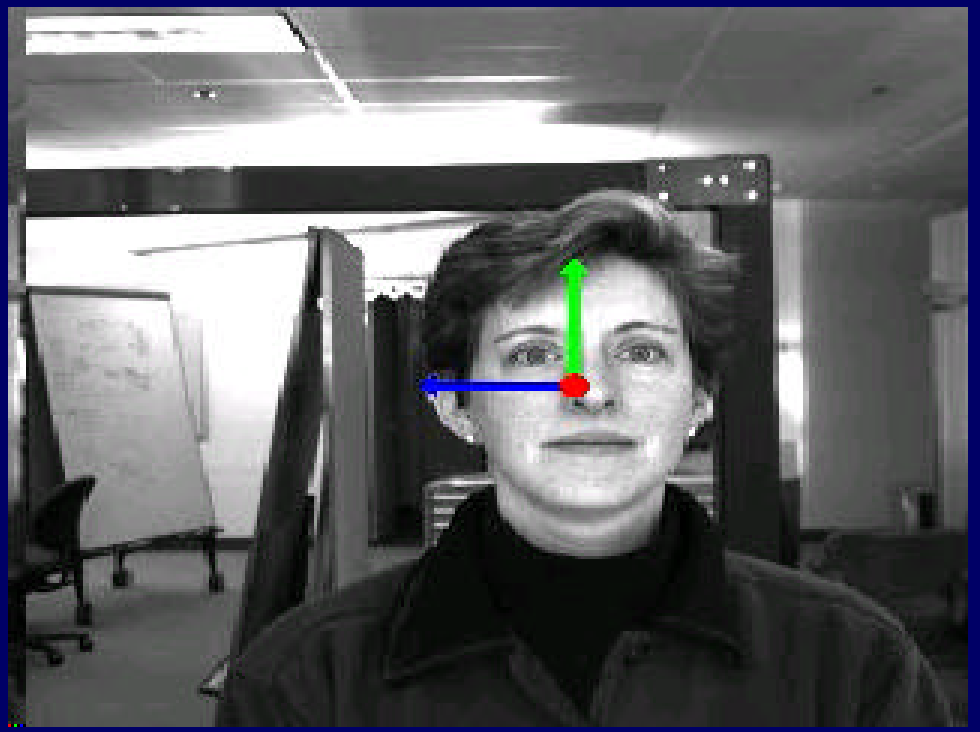
Horprasert, Yacoob, Davis

McKenna, Gong

Wang, Brandstein

Basu, Essa, Pentland

Harville, Darrell, et al



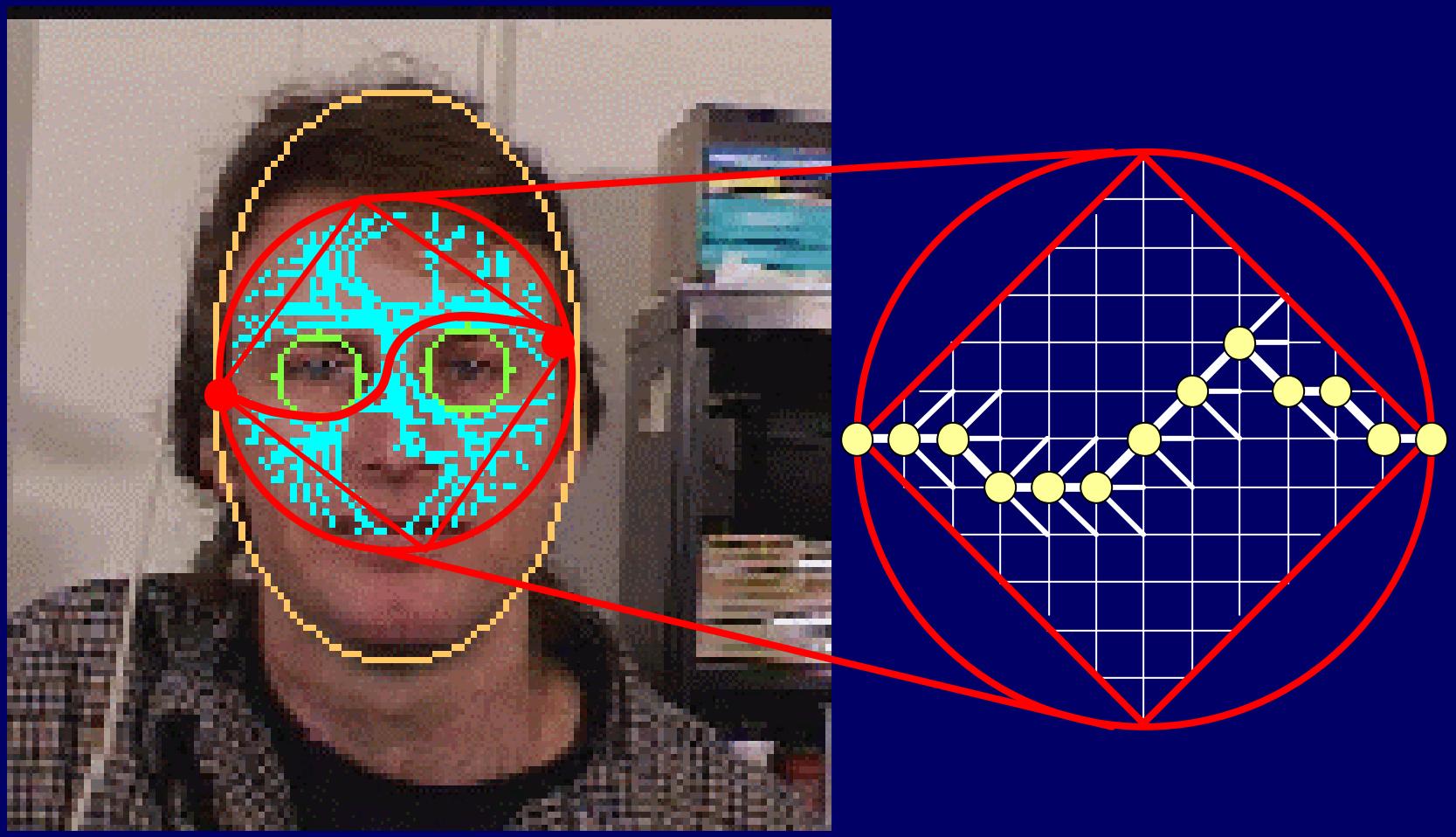
My approach

- Integrate changes in pose (after Harville et al)
- Use mesh model (after Basu et al)
- Need automatic initialization
  - Head detection, tracking, segmentation
  - Reference orientation
  - Head shape parameters
- Initialization drives design

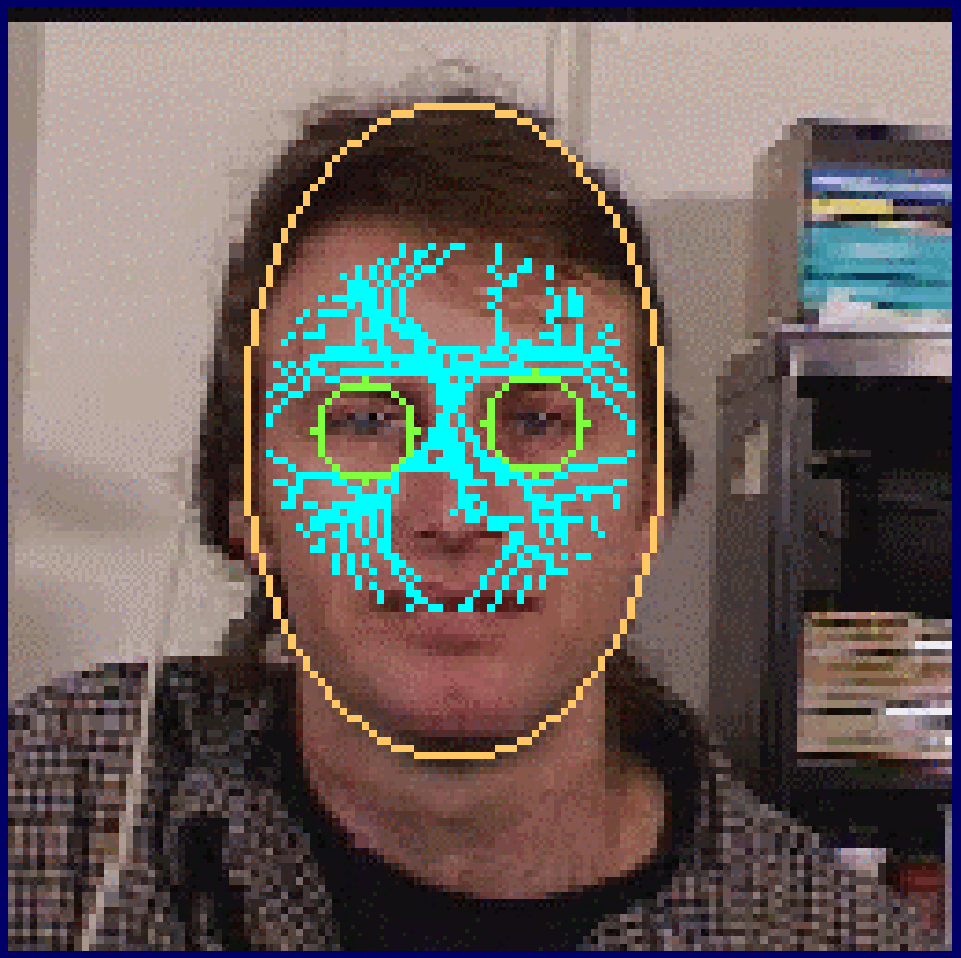
Head tracking, segmentation

- Segment by color histogram, grouped motion
- Match against ellipse model (M. Pilu et al)

Mutual gaze as reference point



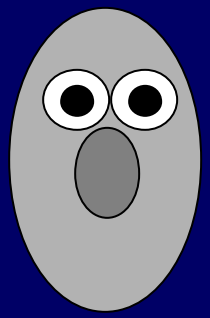
Mutual gaze as reference point



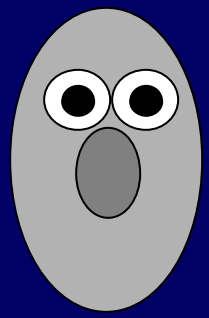
Tracking pose changes

- Choose coordinates to suit tracking
- 4 of 6 degrees of freedom measurable from monocular image
- Independent of shape parameters

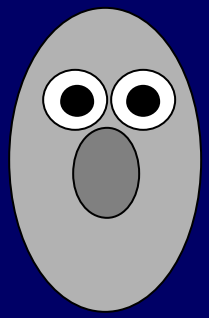
X translation



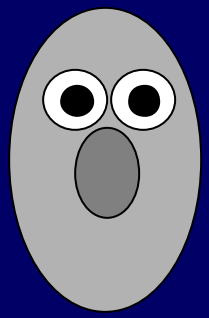
Y translation



Translation in depth

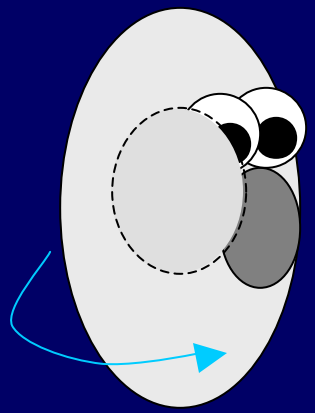


In-plane rotation



Remaining coordinates

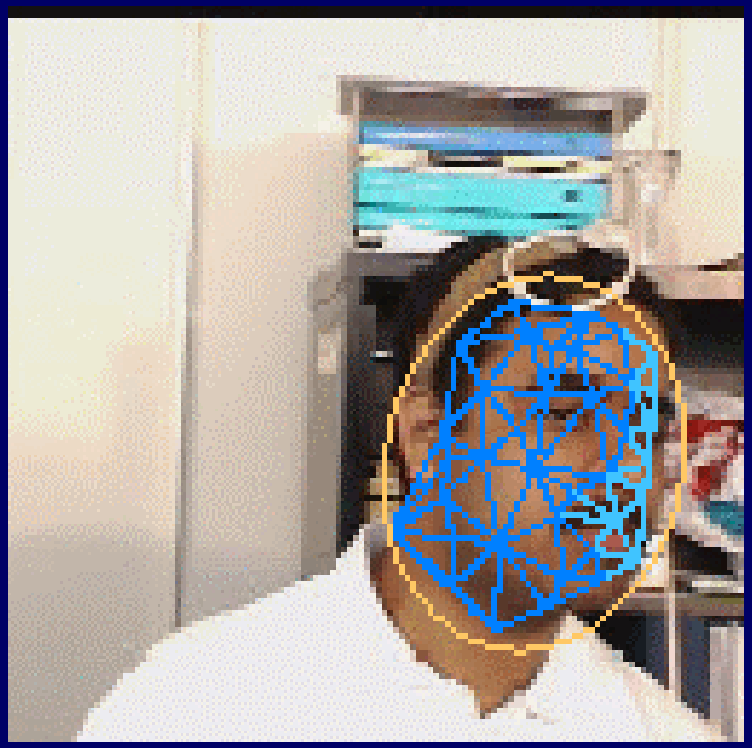
- 2 degrees of freedom remaining
- Choose as surface coordinate on head
- Specify where image plane is tangent to head
- Isolates effect of errors in parameters

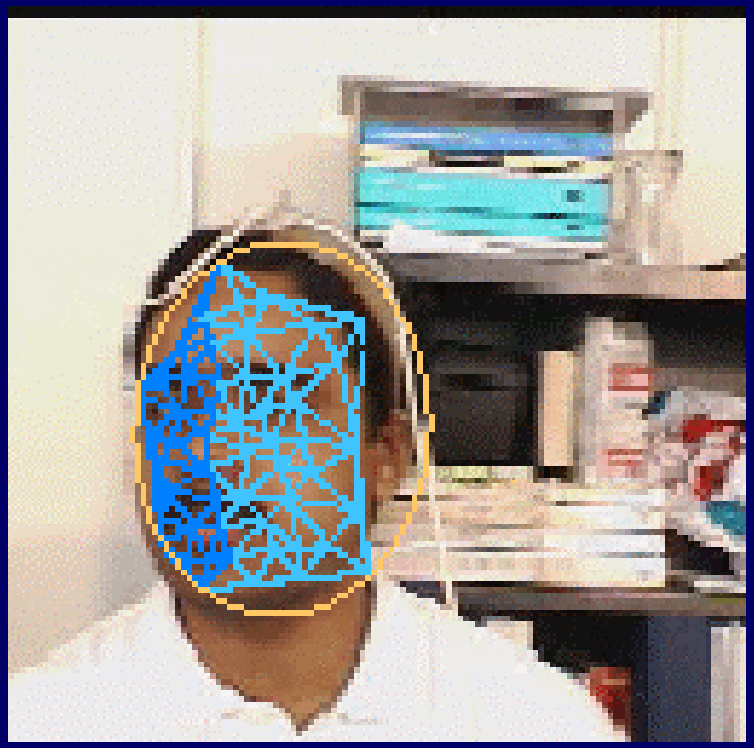


Tangent region shifts when head rotates in depth

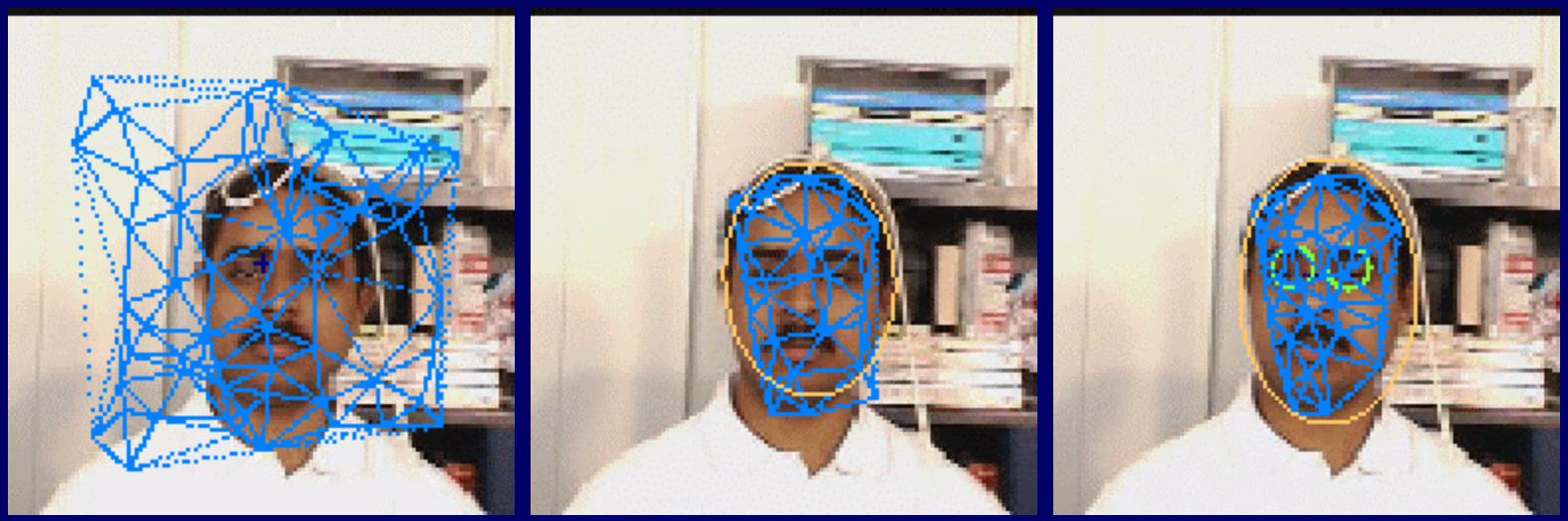
Surface coordinates

- Establish surface coordinate system with mesh



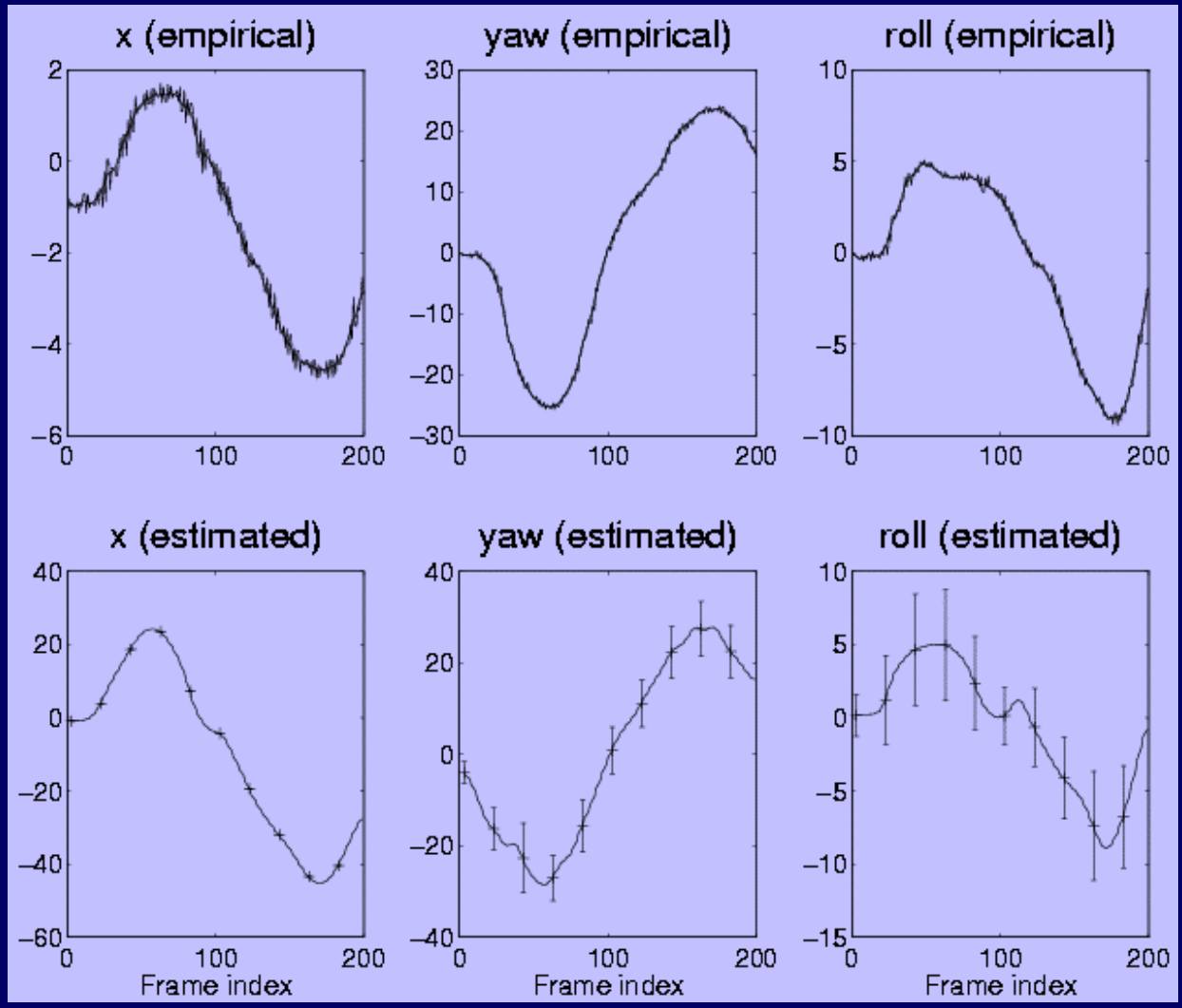


Initializing a surface mesh



Example

Typical results



Ground truth due to Sclaroff et al.

Merits

- No need for any manual initialization
- Capable of running for long periods
- Tracking accuracy is insensitive to model
- User independent
- Real-time

Problems

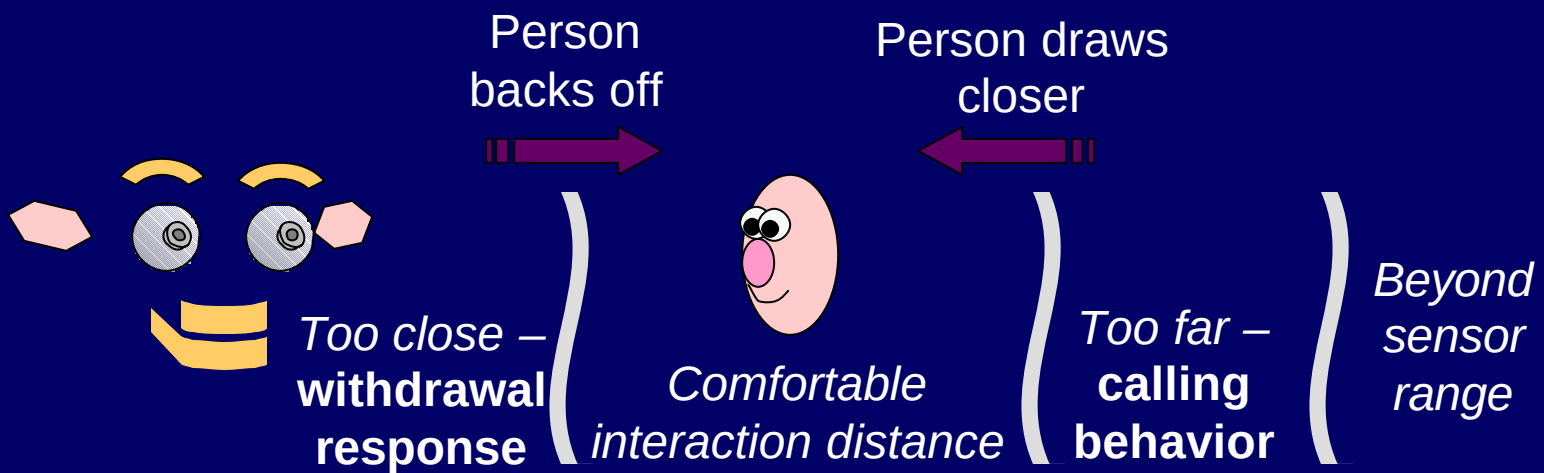
- Greater accuracy possible with manual initialization
- Deals poorly with certain classes of head movement (e.g. 360° rotation)
- Can’t initialize without occasional mutual regard

- Motivation for communication
- Human-readable actions
- Reading human actions
- Conclusions

Other protocols

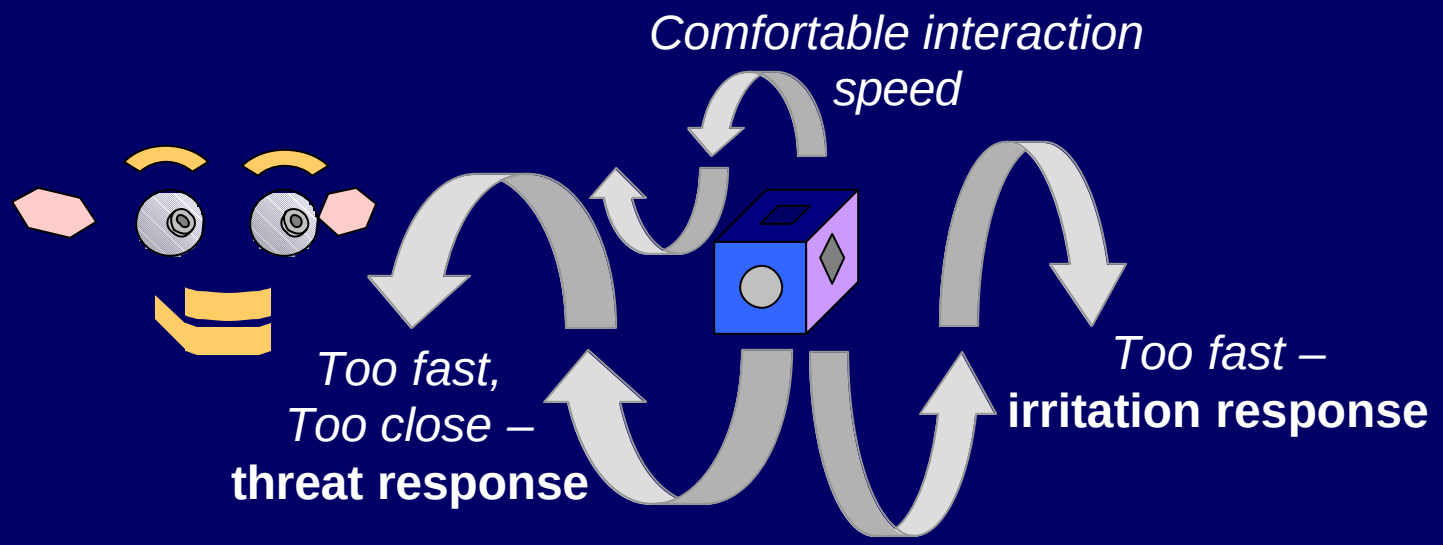
Other protocols

- Protocol for negotiating interpersonal distance



Other protocols

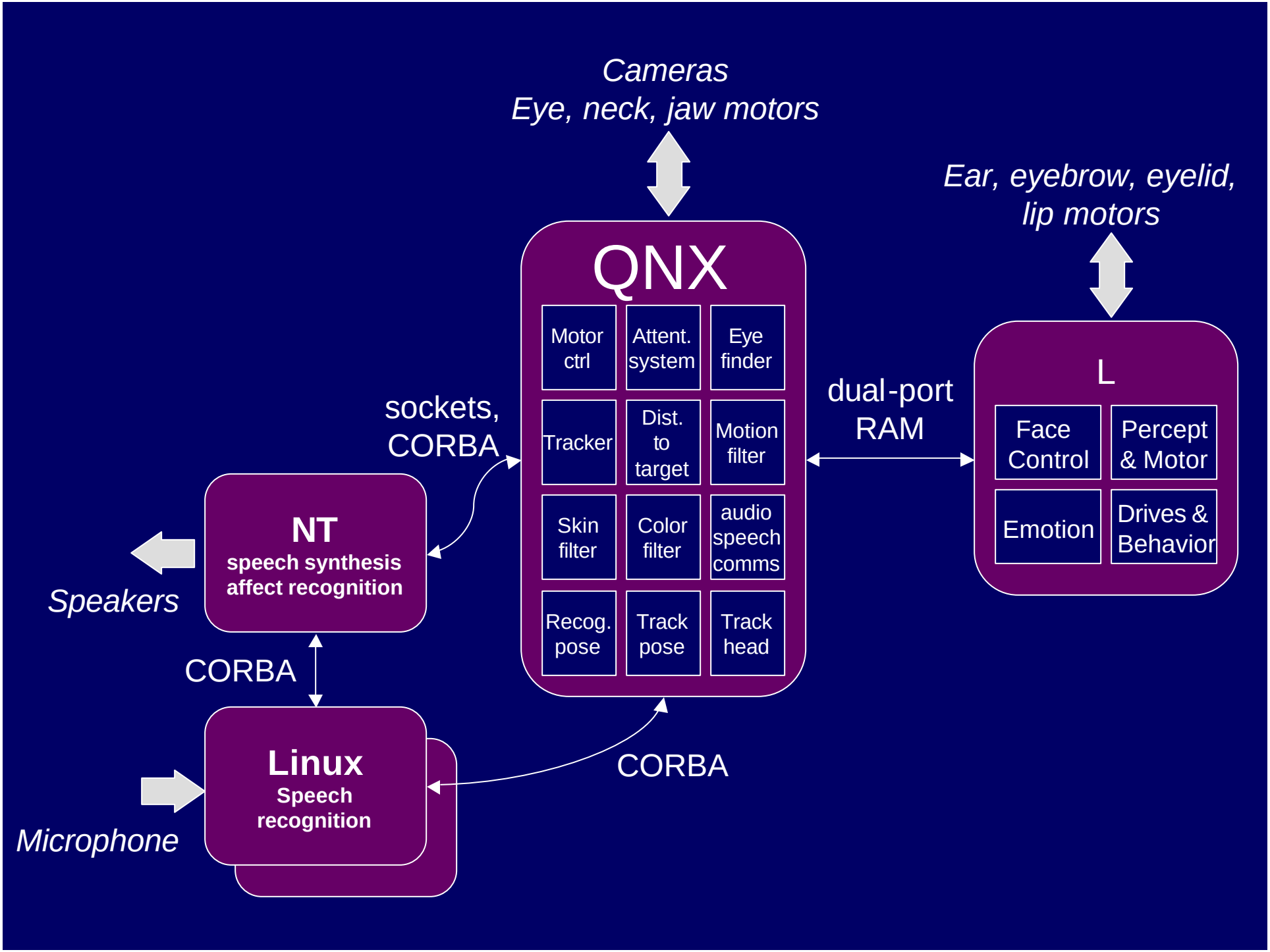
- Protocol for negotiating interpersonal distance
- Protocol for controlling the presentation of objects

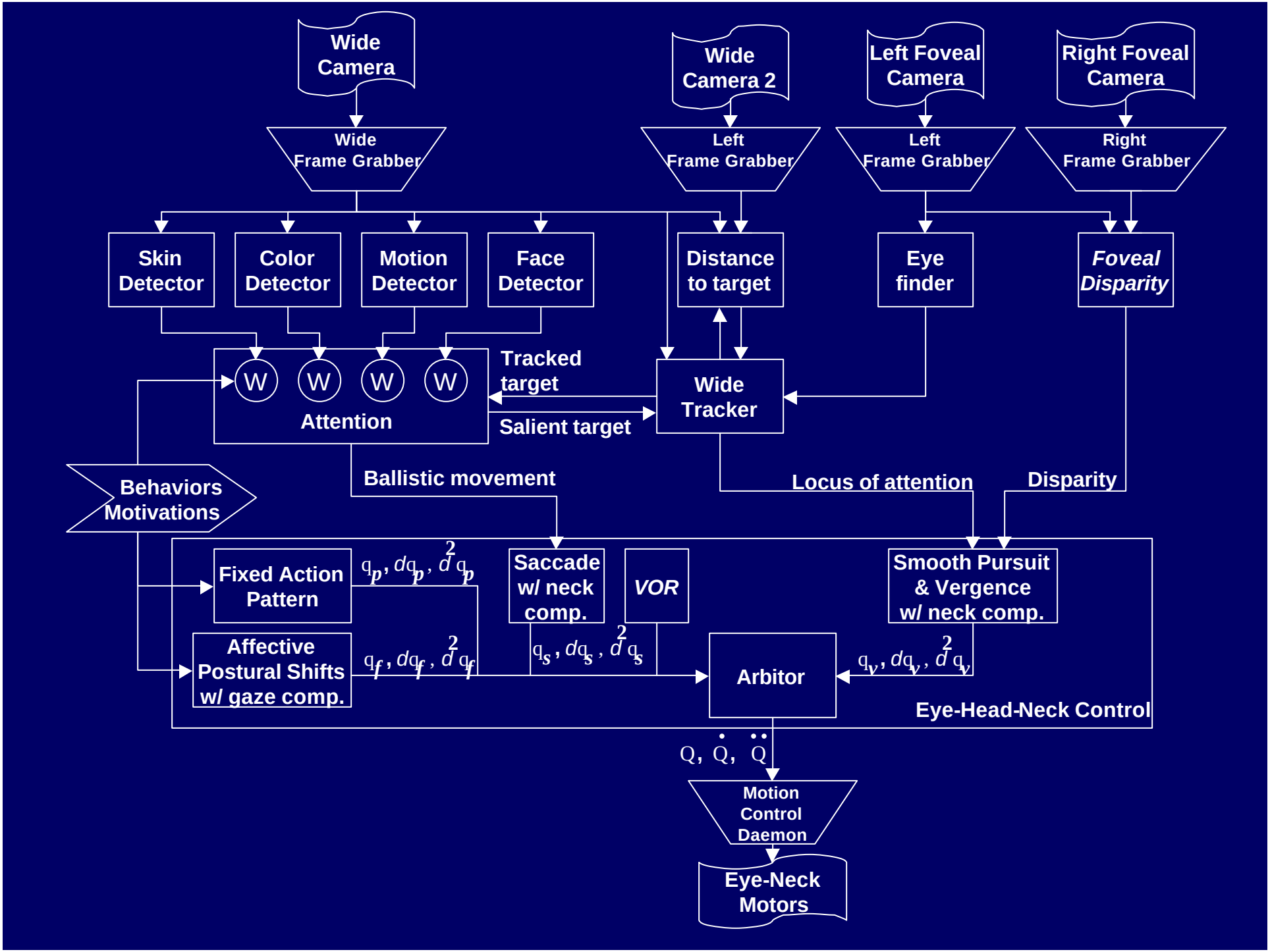


Other protocols

- Protocol for negotiating interpersonal distance
- Protocol for controlling the presentation of objects
- Protocol for conversational turn-taking
- Protocol for introducing vocabulary
- Protocol for communicating processes

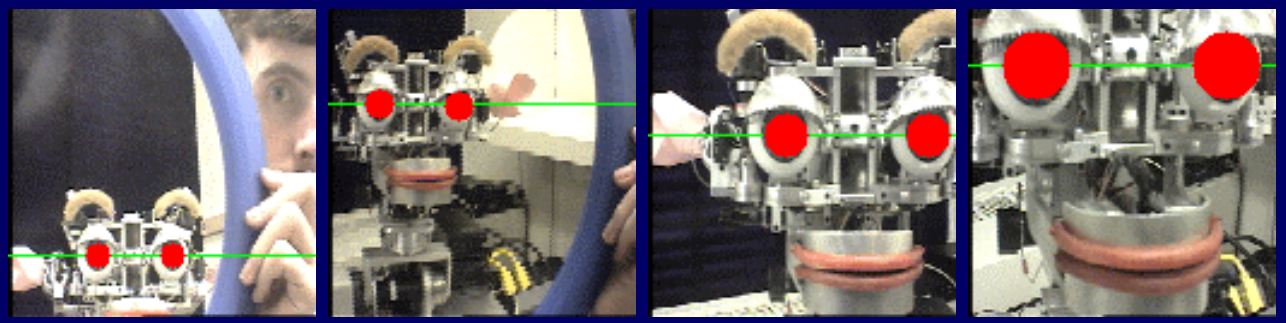
Protocols make good modules


Cameras
Eye, neck, jaw motors
Ear, eyebrow, eyelid, lip motors
QNX
Motor ctrl
Attent. system
Eye finder
Tracker
Dist. to target
Motion filter
Skin filter
Color filter
audio speech comms
Recog. pose
Track pose
Track head
L
Face Control
Percept & Motor
Emotion
Drives & Behavior
dual-port RAM
sockets, CORBA
NT
speech synthesis
affect recognition
Speakers
CORBA
Linux
Speech recognition
Microphone
CORBA


Wide Camera
Wide Camera 2
Left Foveal Camera
Right Foveal Camera
Wide Frame Grabber
Left Frame Grabber
Left Frame Grabber
Right Frame Grabber
Skin Detector
Color Detector
Motion Detector
Face Detector
Distance to target
Eye finder
Foveal Disparity
W
W
W
W
Attention
Tracked target
Salient target
Wide Tracker
Behaviors Motivations
Ballistic movement
Locus of attention
Disparity
Fixed Action Pattern
$q_p, dq_p, d^2 q_p$
Saccade w/ neck comp.
VOR
Smooth Pursuit & Vergence w/ neck comp.
Affective Postural Shifts w/ gaze comp.
$q_f, dq_f, d^2 q_f$
$q_s, dq_s, d^2 q_s$
Arbitor
$q_v, dq_v, d^2 q_v$
Eye-Head-Neck Control
$Q, \dot{Q}, \ddot{Q}$
Motion Control Daemon
Eye-Neck Motors

Other protocols

- What about robot – robot protocol?
- Basically computer – computer
- But physical states may be hard to model
- Borrow human – robot protocol for these



Current, future work

- Protocols for reference
  - Know how to point to an object
  - How to point to an attribute?
  - Or an action?
- Until a better answer comes along:
  - Communicate task/game that depends on attribute/action
  - Pull out number of classes, positive and negative examples for supervised learning

FIN