

Sociable Robots

Peeping into the Human World

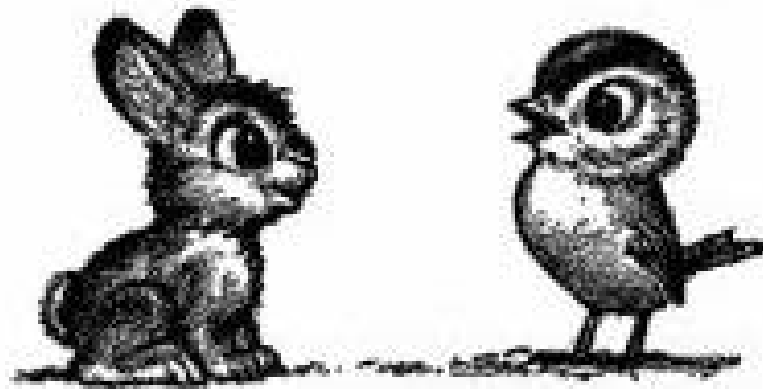


An Infant's Advantages

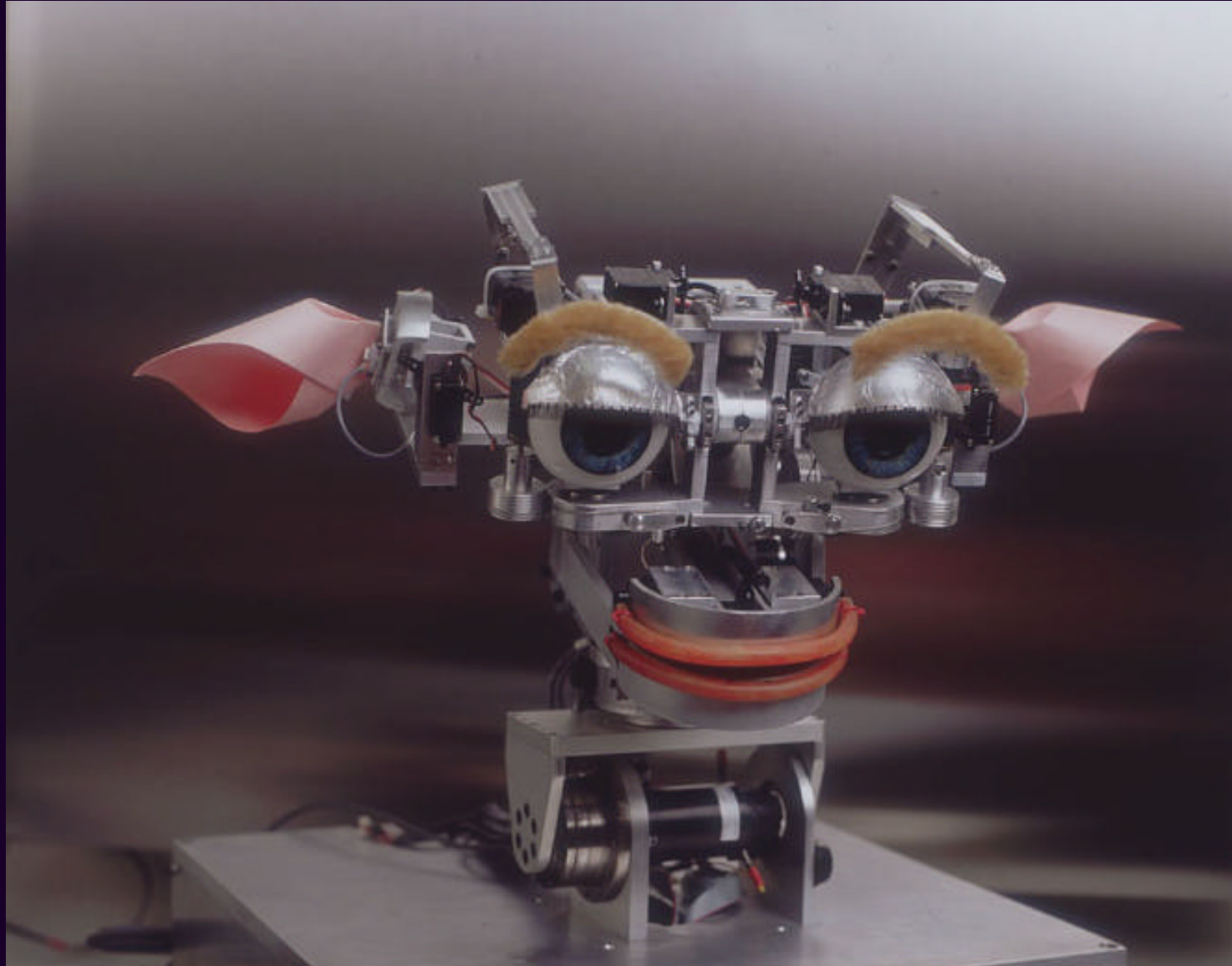
- Non-hostile environment
- Actively benevolent, empathic caregiver
- Co-exists with mature version of self

Baby Scheme

- Physical form can evoke nurturing response
- Caregiver exaggerates voice, gestures



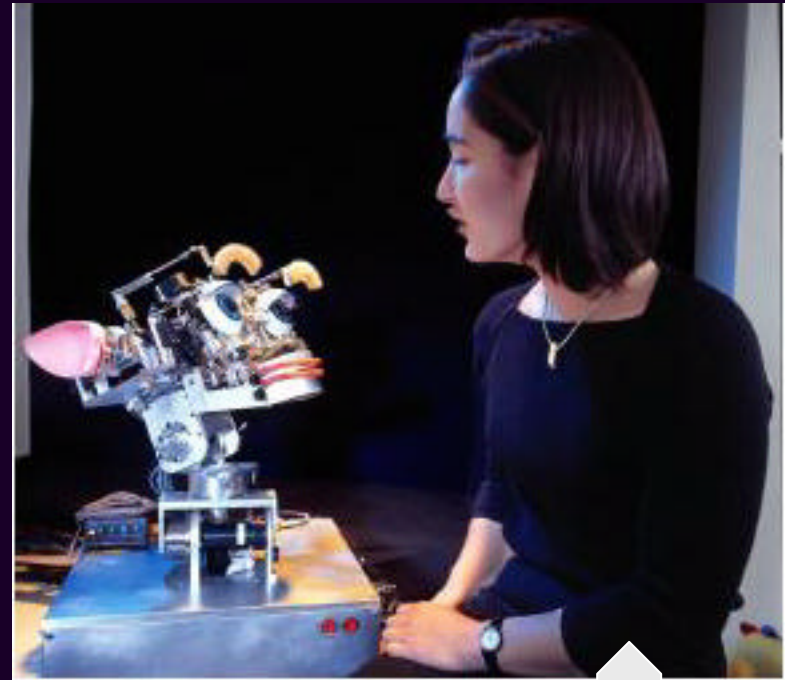
Kismet – a Baby Robot



Requirements

- Robot needs to perceive human state
 - Computer vision, speech processing
- Human needs to perceive robot state
 - Animatronics, speech generation
- Closed loop interaction requires both directions

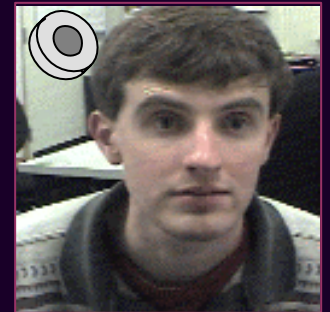
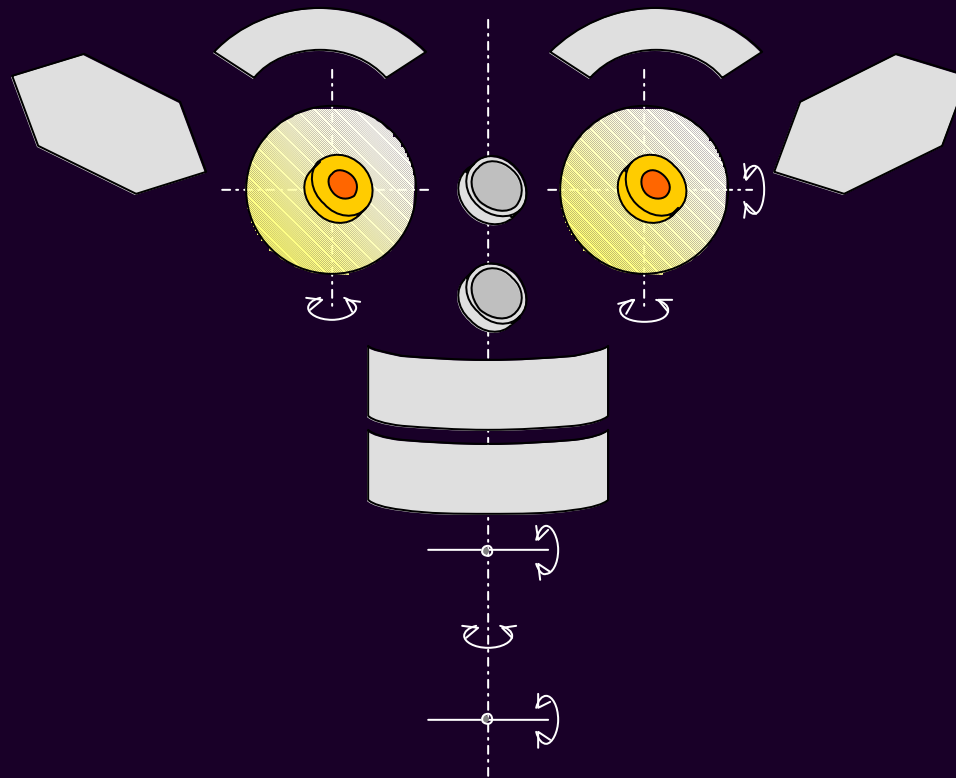
Readable locus of attention



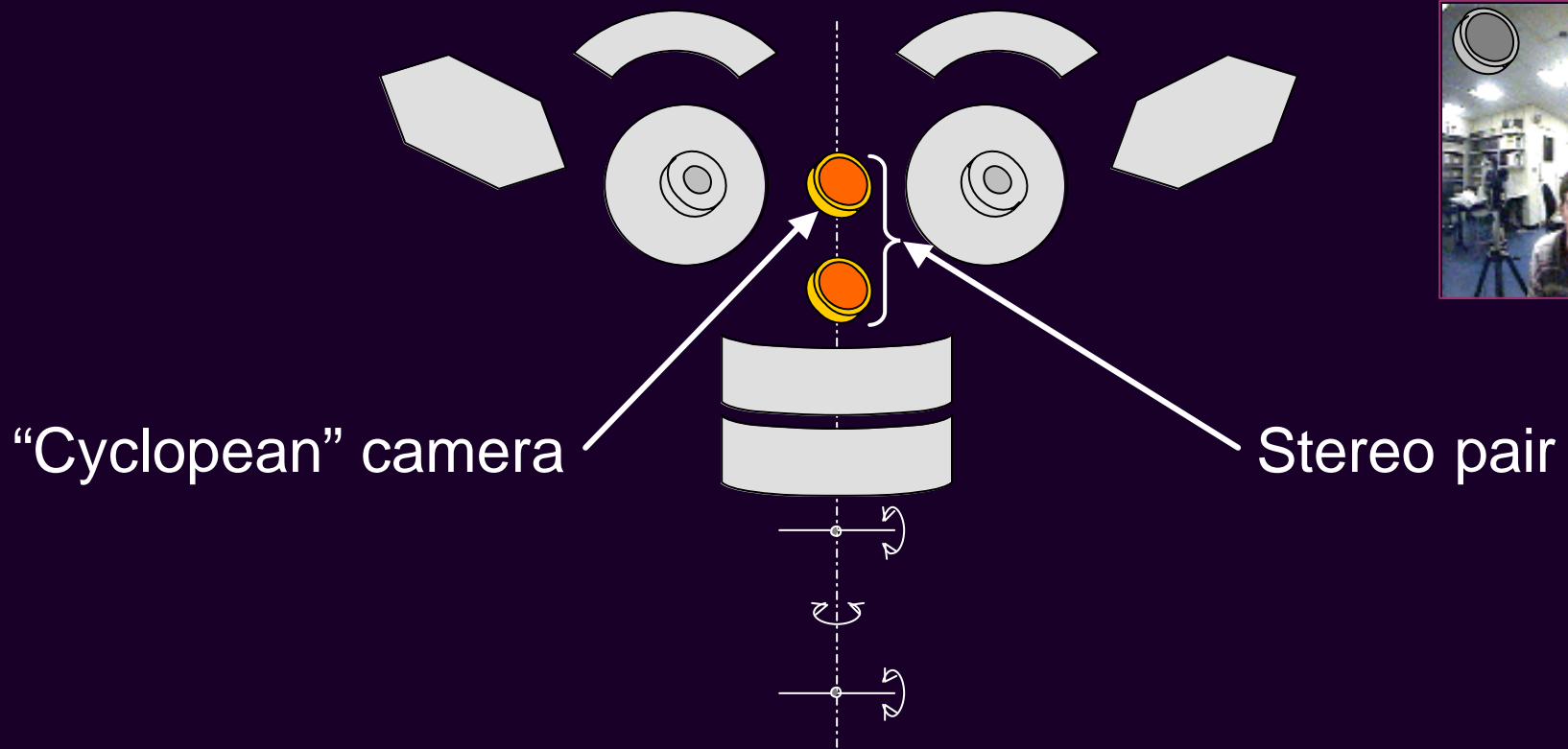
Attention can be deduced from behavior

Or can be expressed more directly

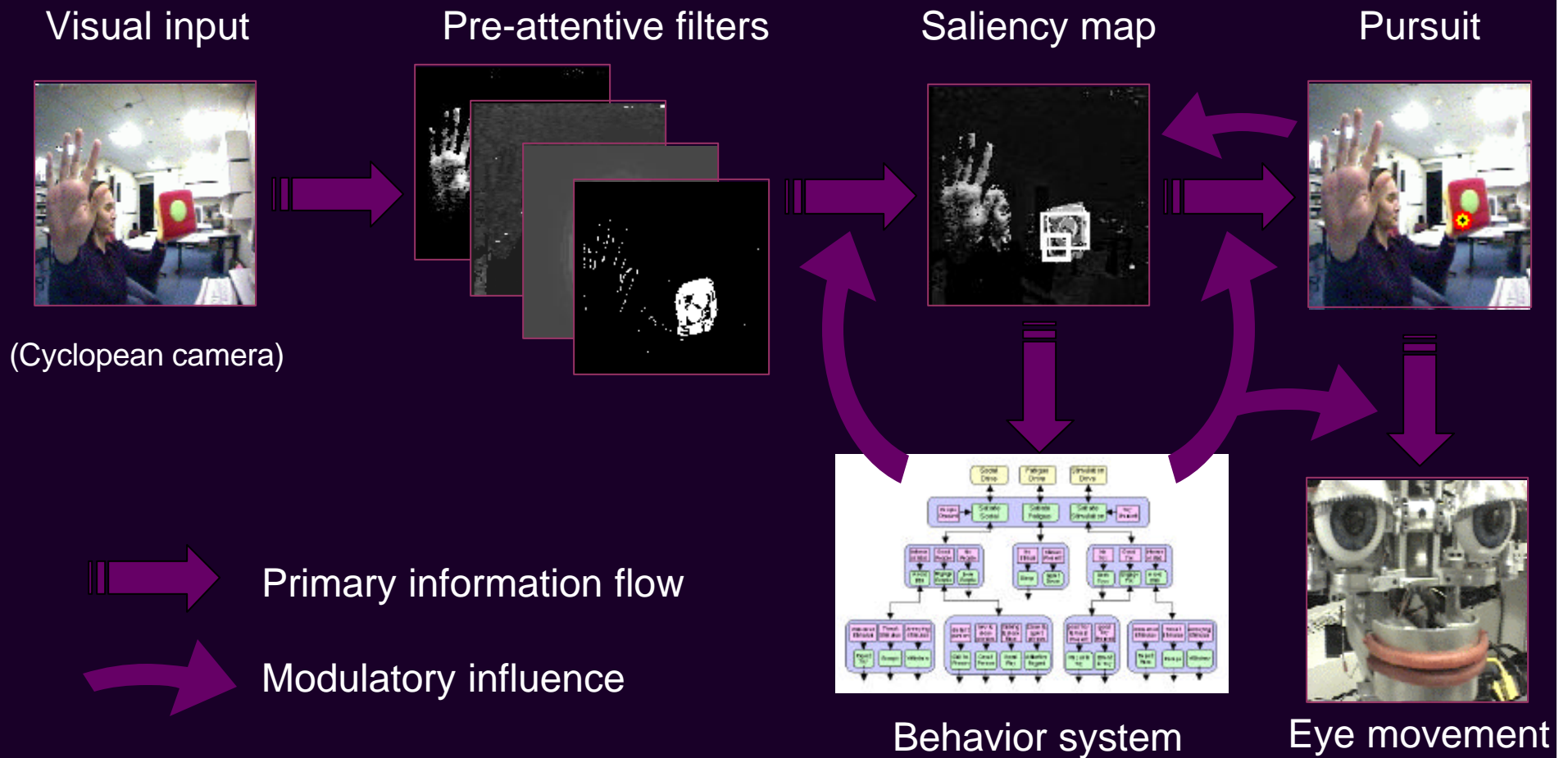
Expressing locus of attention



Computing locus of attention



Computing locus of attention

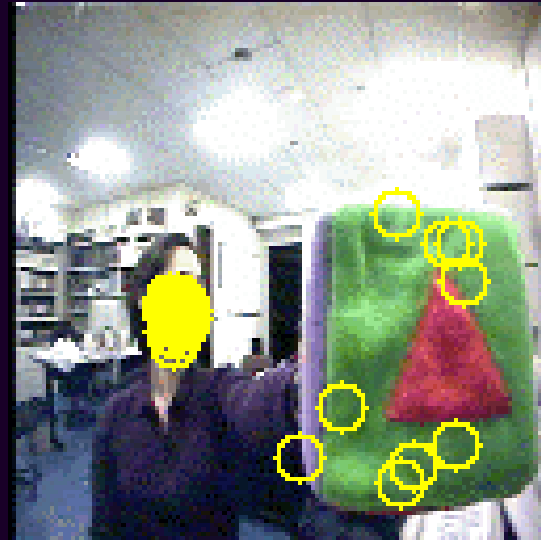


Looking Preference



“Seek toy” –

low skin gain, high saturated-color gain
Looking time 28% face, 72% block

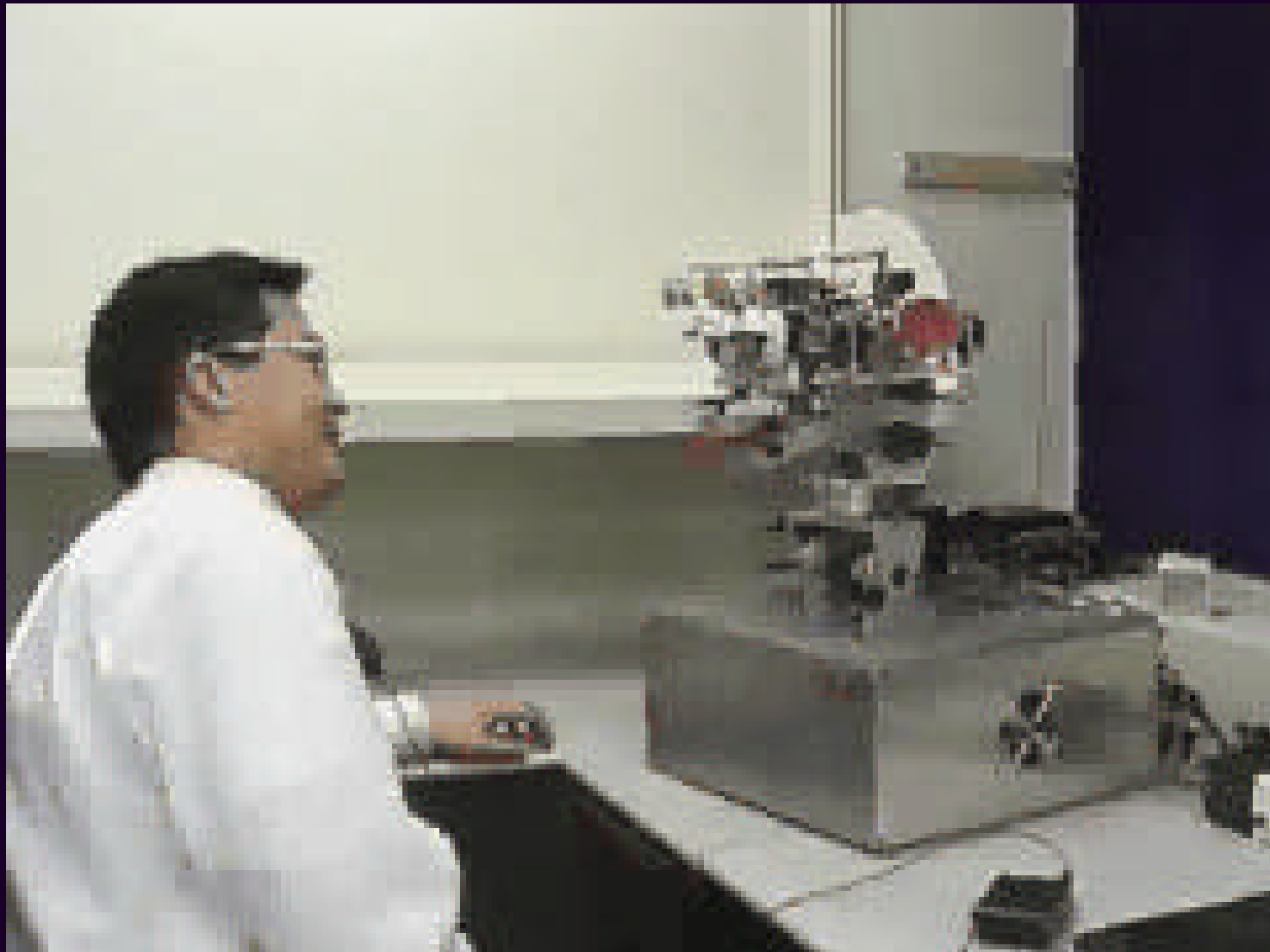


“Seek face” –

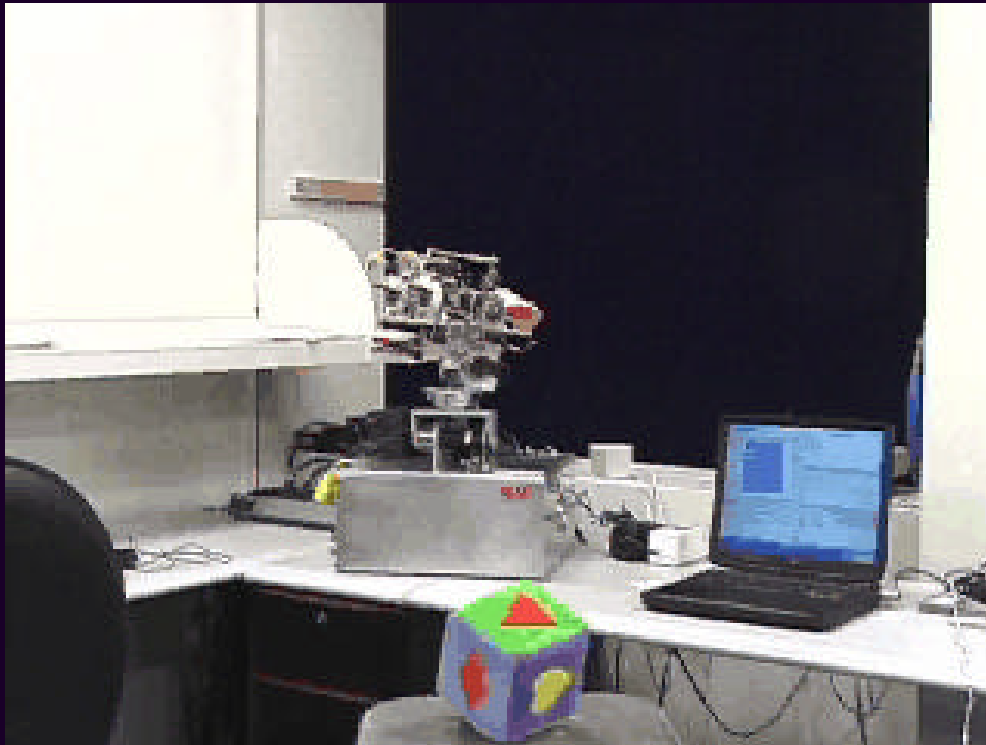
high skin gain, low color saliency gain
Looking time 80% face, 20% block

- Internal influences bias how saliency is measured
- The robot is not a slave to its environment
- Prefers behaviorally relevant stimuli

Example (video)

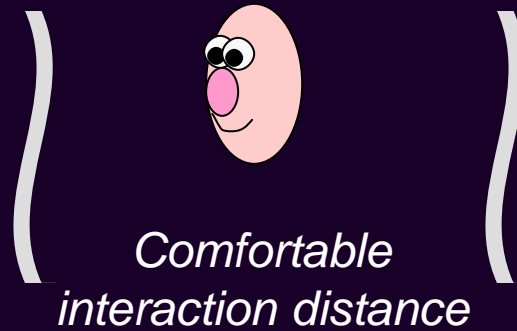
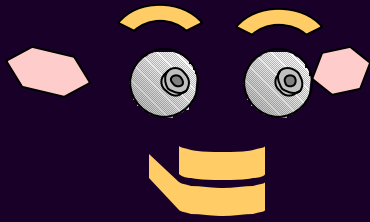


Example (video)



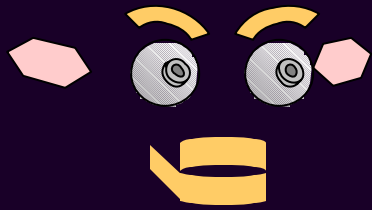
- Robot's search is task-specific
- Still opportunistic when appropriate
- Visual behavior conveys degree of commitment
- Gaze direction, expression conveys interest

Interpersonal Distance

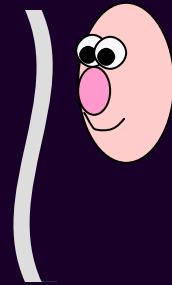
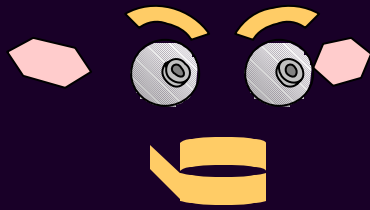


*Comfortable
interaction distance*

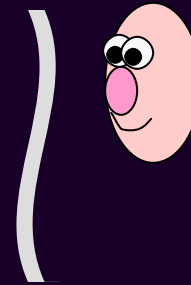
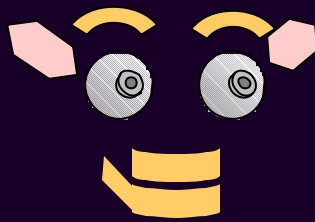
Interpersonal Distance



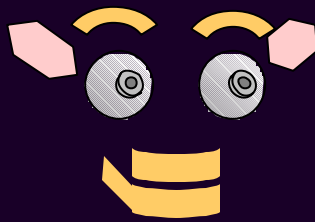
Interpersonal Distance



Interpersonal Distance



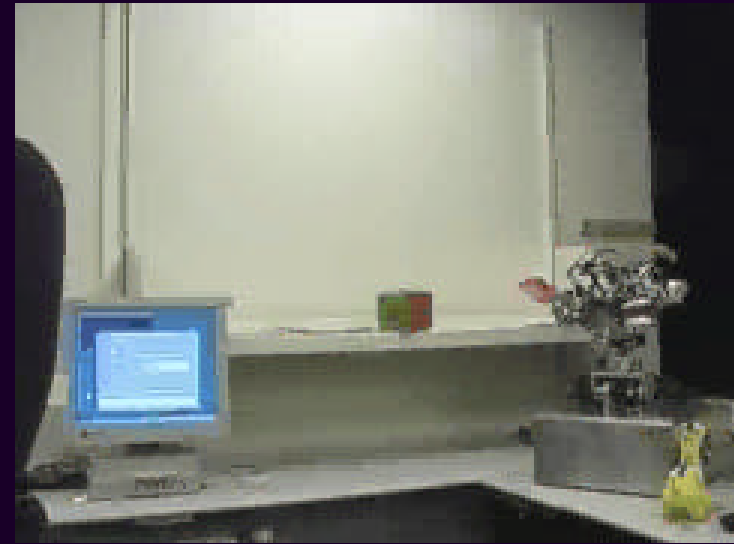
Interpersonal Distance



Examples (video)



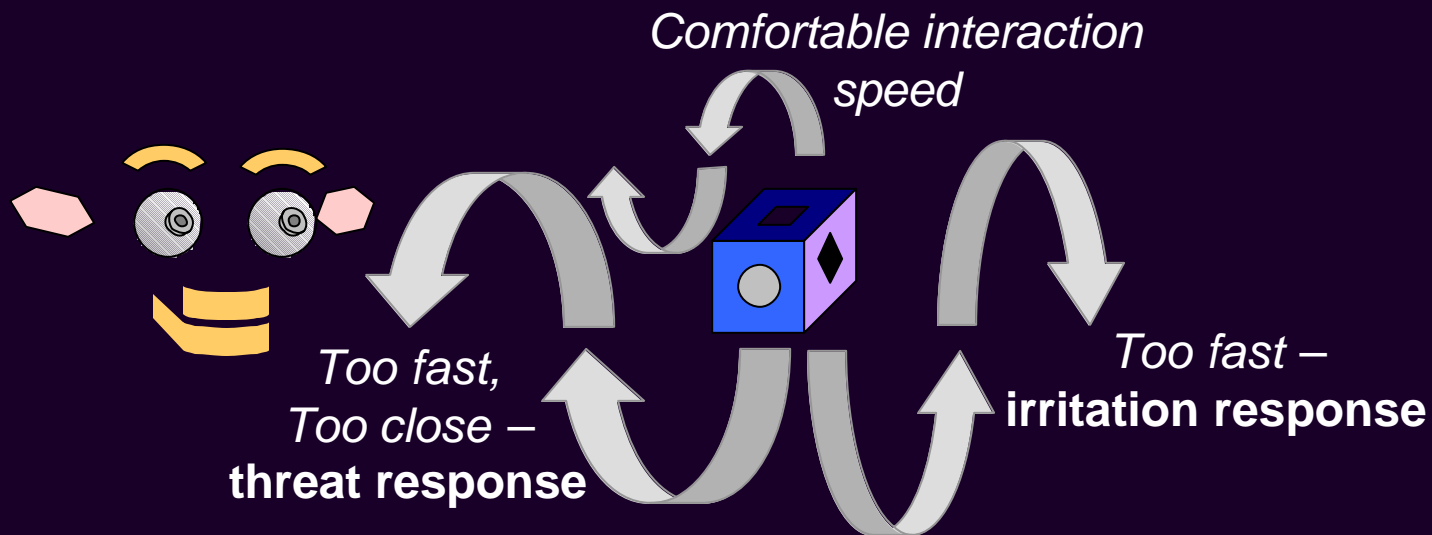
“Back off buster!”



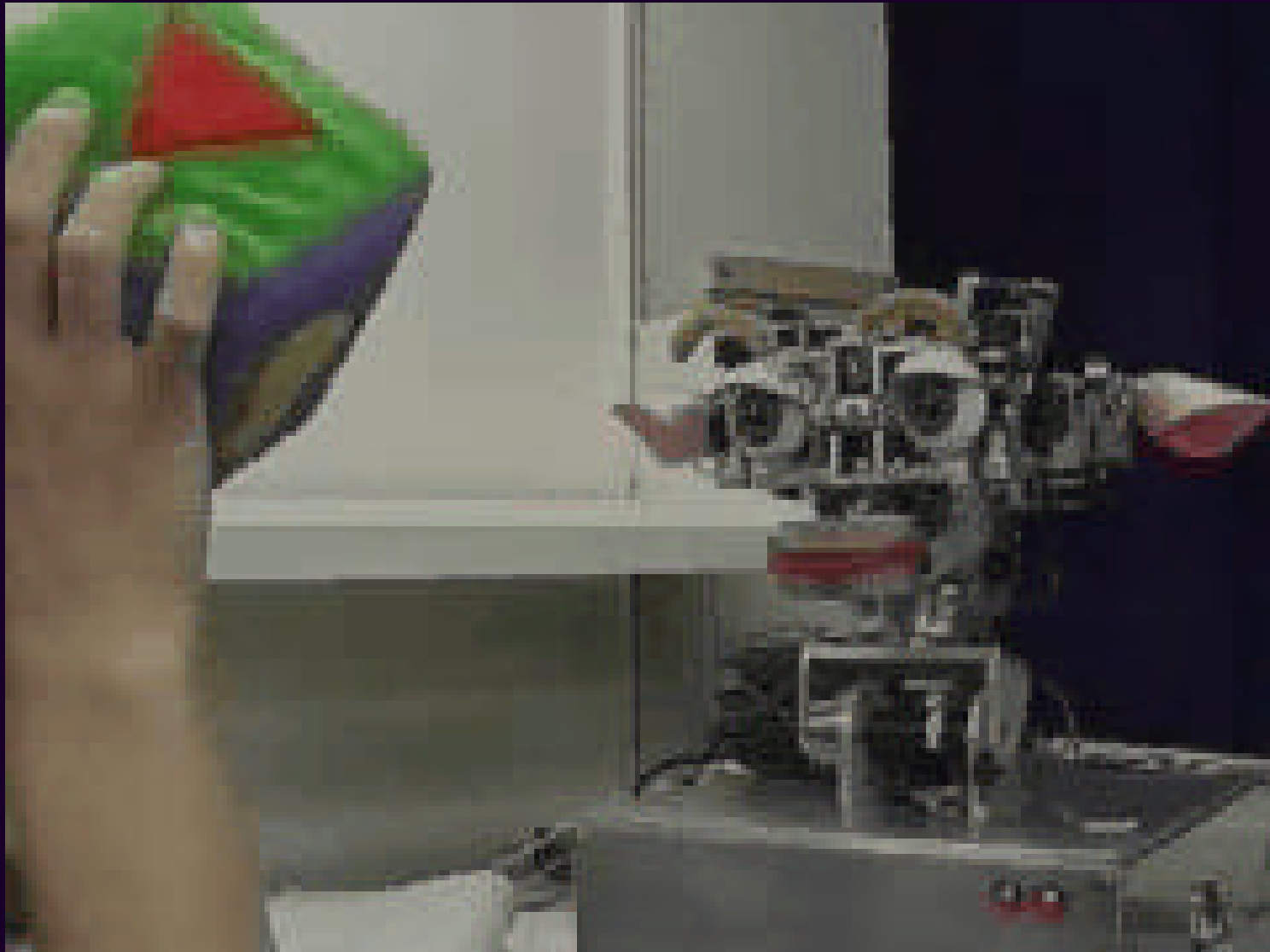
“Come hither, friend”

- Robot backs away if person comes too close
- Cues person to back away too – social amplification
- Robot makes itself salient to call a person closer if too far away

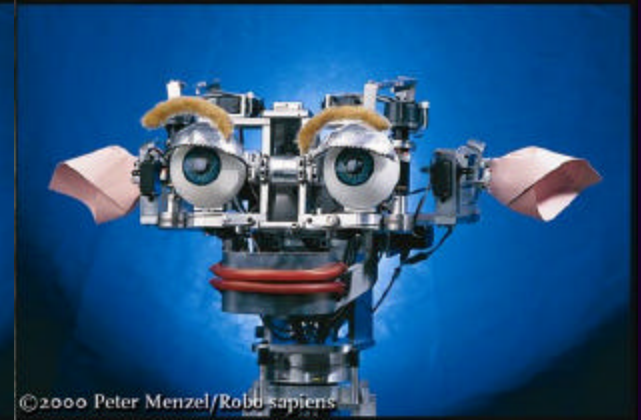
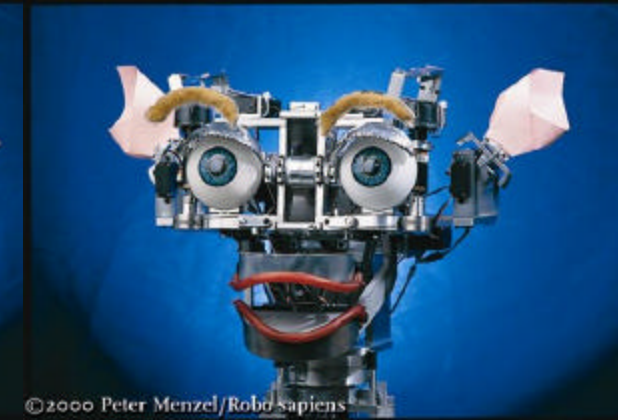
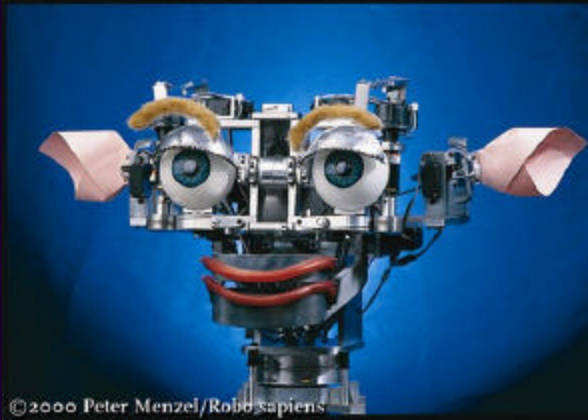
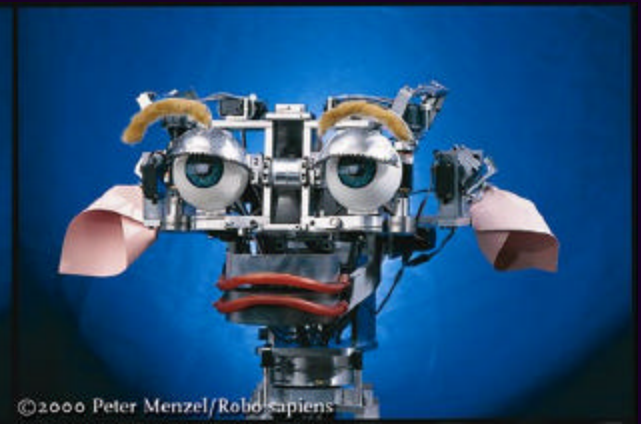
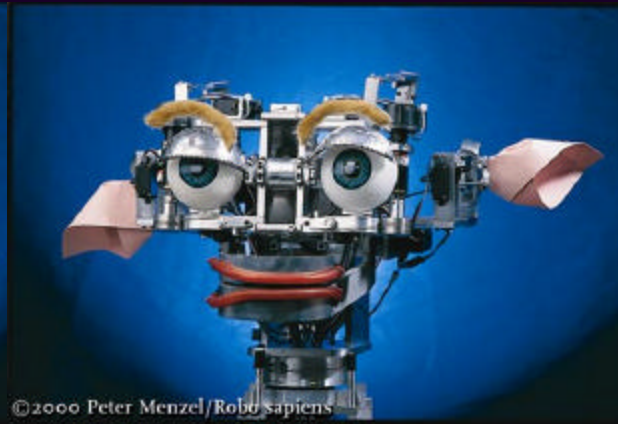
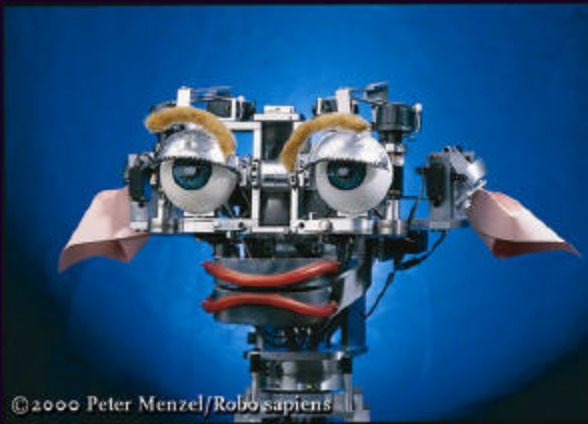
Negotiating object showing



Example (video)

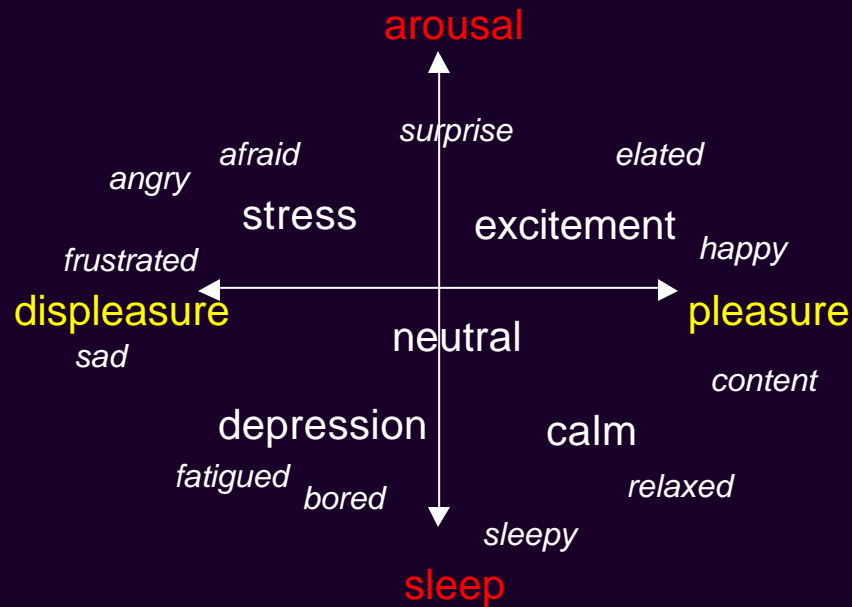


Facial expressions



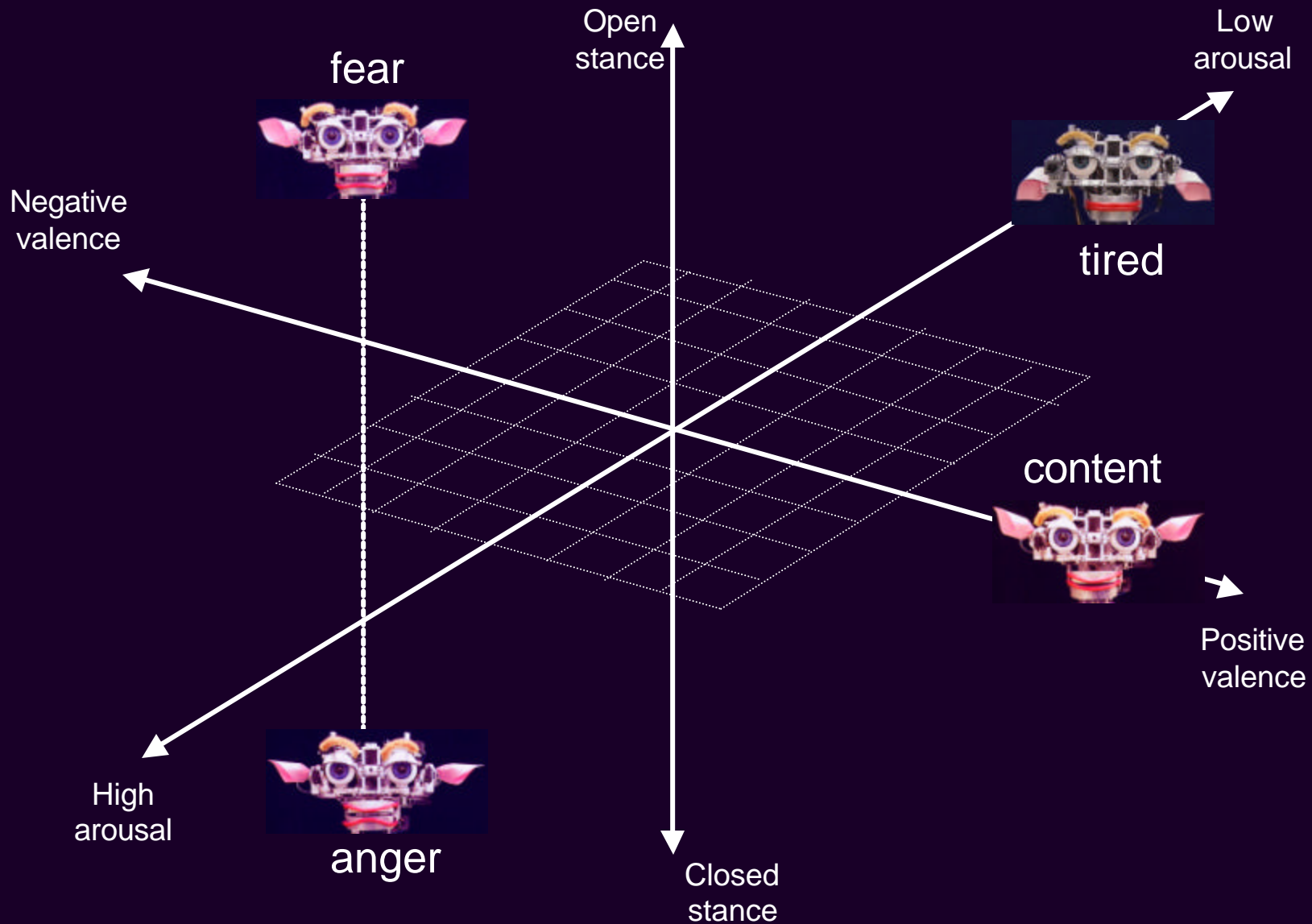
Facial Expressions

(Russell, Scott&Smith)

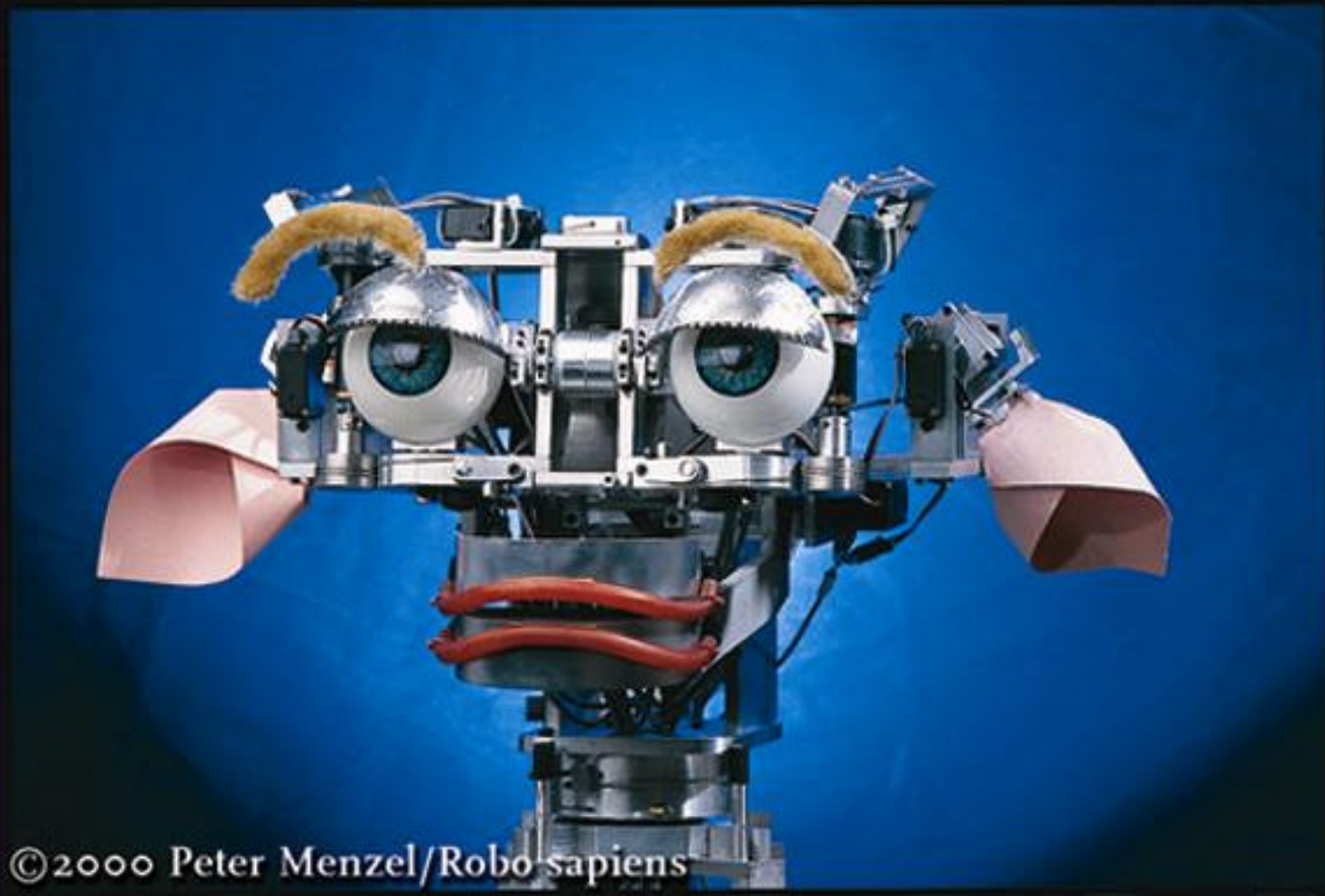


- Emotions can be mapped to affect dimensions (Russell)
- Facial postures are related in a systematic way to these affective dimensions (Smith & Scott)

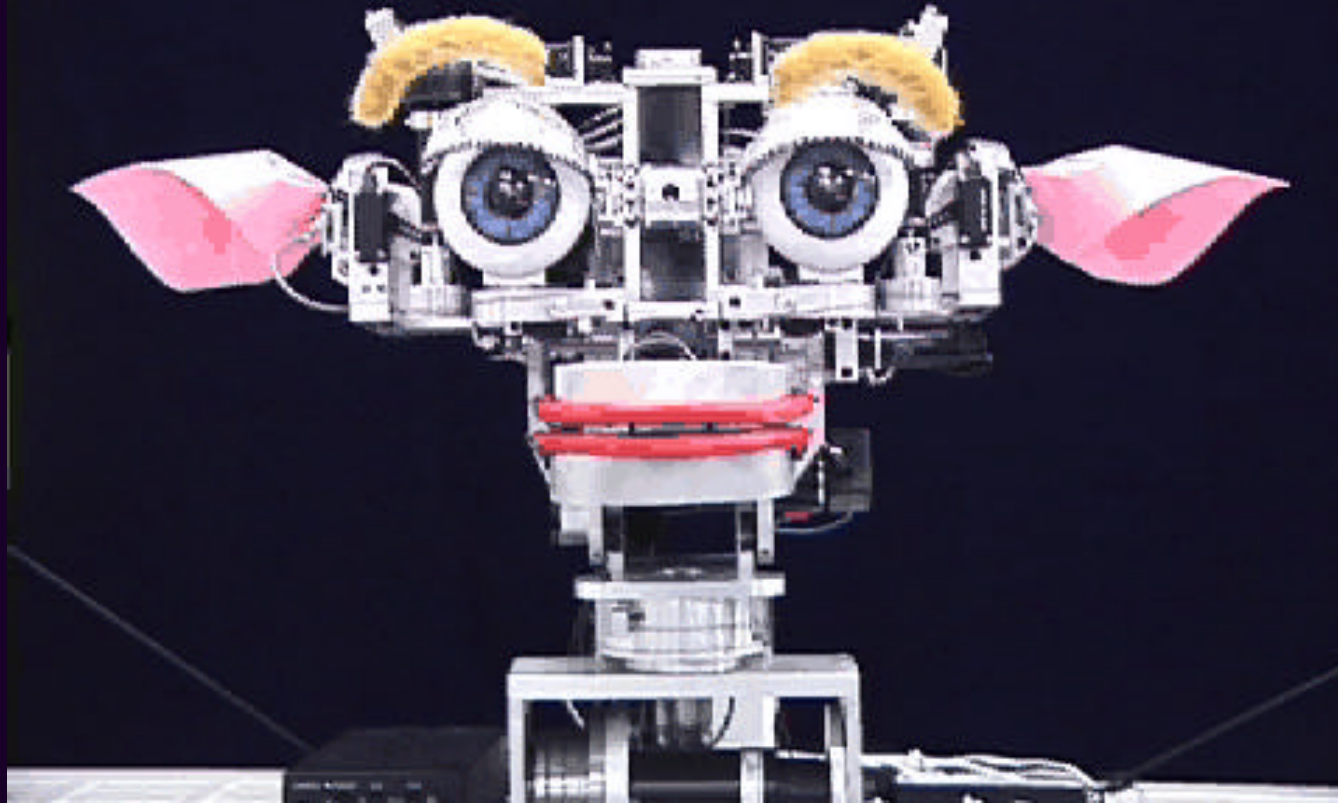
Generating Posture, Expressions



Example Facial Expression



With Posture (video)



Emotive Voice Quality

The effect of emotions on the human voice

	fear	anger	sorrow	joy	disgust	surprise
speech rate	much faster	slightly faster	slightly slower	faster or slower	very much slower	much faster
pitch average	very much higher	very much higher	slightly lower	much higher	very much lower	much higher
pitch range	much wider	much wider	slightly narrower	much wider	slightly wider	
intensity	normal	higher	lower	higher	lower	higher
voice quality	irregular voicing	breathy chest tone	resonant	breathy blaring	grumbled chest tone	
pitch changes	normal	abrupt on stressed syllable	downward inflections	smooth upward inflections	wide downward terminal inflections	rising contour
articulation	precise	tense	slurring	normal	normal	

Synthesized Emotive Speech

The screenshot shows the BabyBox software interface. At the top, there is a title bar with the text "BabyBox" and a close button. Below the title bar, there is an "Initialize" button and a text field containing the following parameters: `[:rate 96 :cp 672 :pp 768]` and `[:dv ap 269 as 65 bf 0 br 47 gf 77 gh 73 gv 65 hr 20 la 0 g5 65 lx 75 pr 222 qu 100 ri 70 sm 2 sr 33]`. To the right of the text field is an "Exit" button.

The interface is divided into several sections:

- Affective State:** A dropdown menu for "Emotion" is set to "Disgust".
- Pitch:** Several sliders are visible: "Accent Shape" (0), "Final Lowering" (0), "Average Pitch" (-8), "Pitch Range" (3), "Contour Slope" (0), and "Reference Line" (0).
- Timing:** "Speech Rate" is set to -8 and "Stress Frequency" is set to 0.
- Voice Quality:** "Breathiness" (0), "Loudness" (0), "Brilliance" (5), "Pause Discontinuity" (0), "Laryngealization" (0), and "Pitch Discontinuity" (10).
- Articulation:** "Precision" is set to 7.

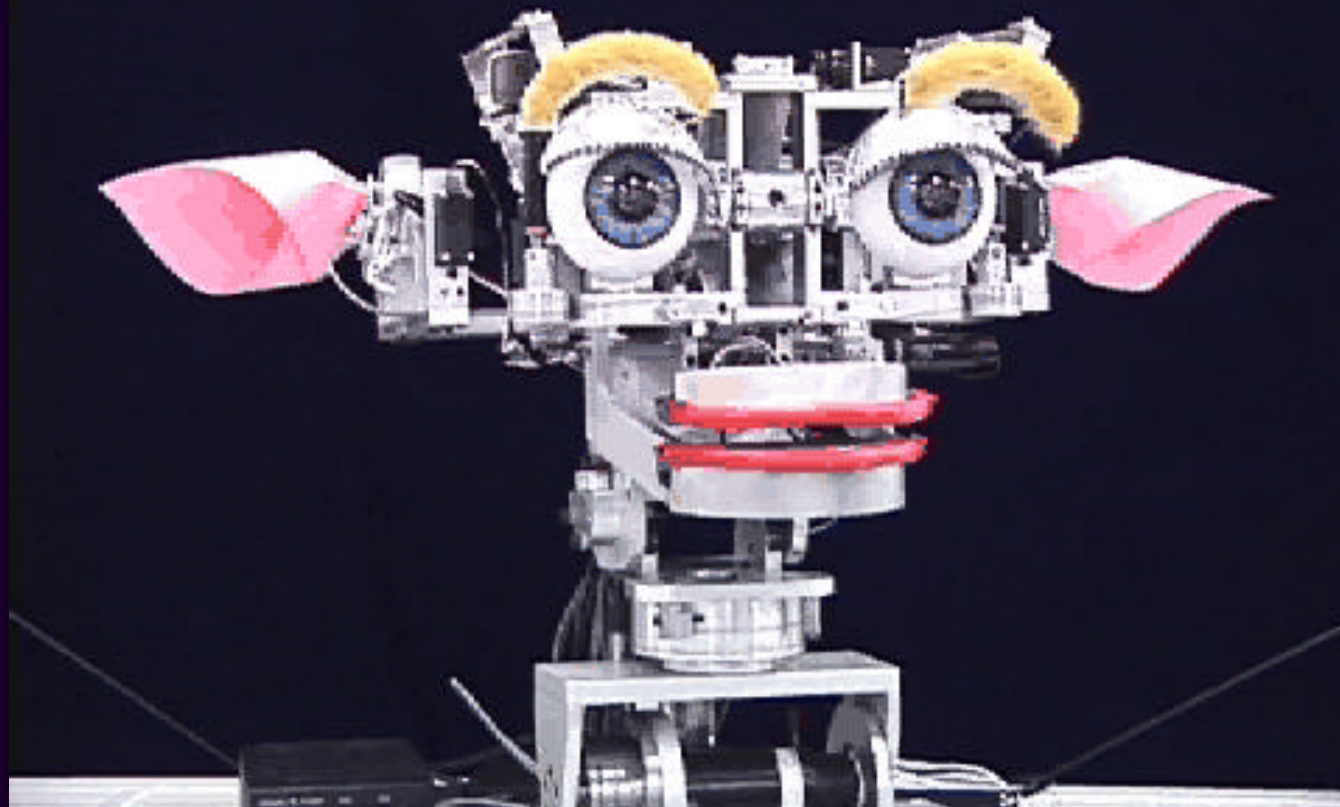
The "Utterance Settings" section on the right contains:

- Manually generated utterance:** A text field with "hello there!" and a "Say This" button.
- Self-generated utterance:** A text field with phonetic transcription `[ey<274>] [] [ch'rr<367>] [\] [ih<202>] !` and a "Baby!" button.
- Past utterances:** A text area showing a list of previously generated utterances and their phonetic transcriptions.

At the bottom, there is a "DecTalk Settings" section with a grid of sliders for various parameters:

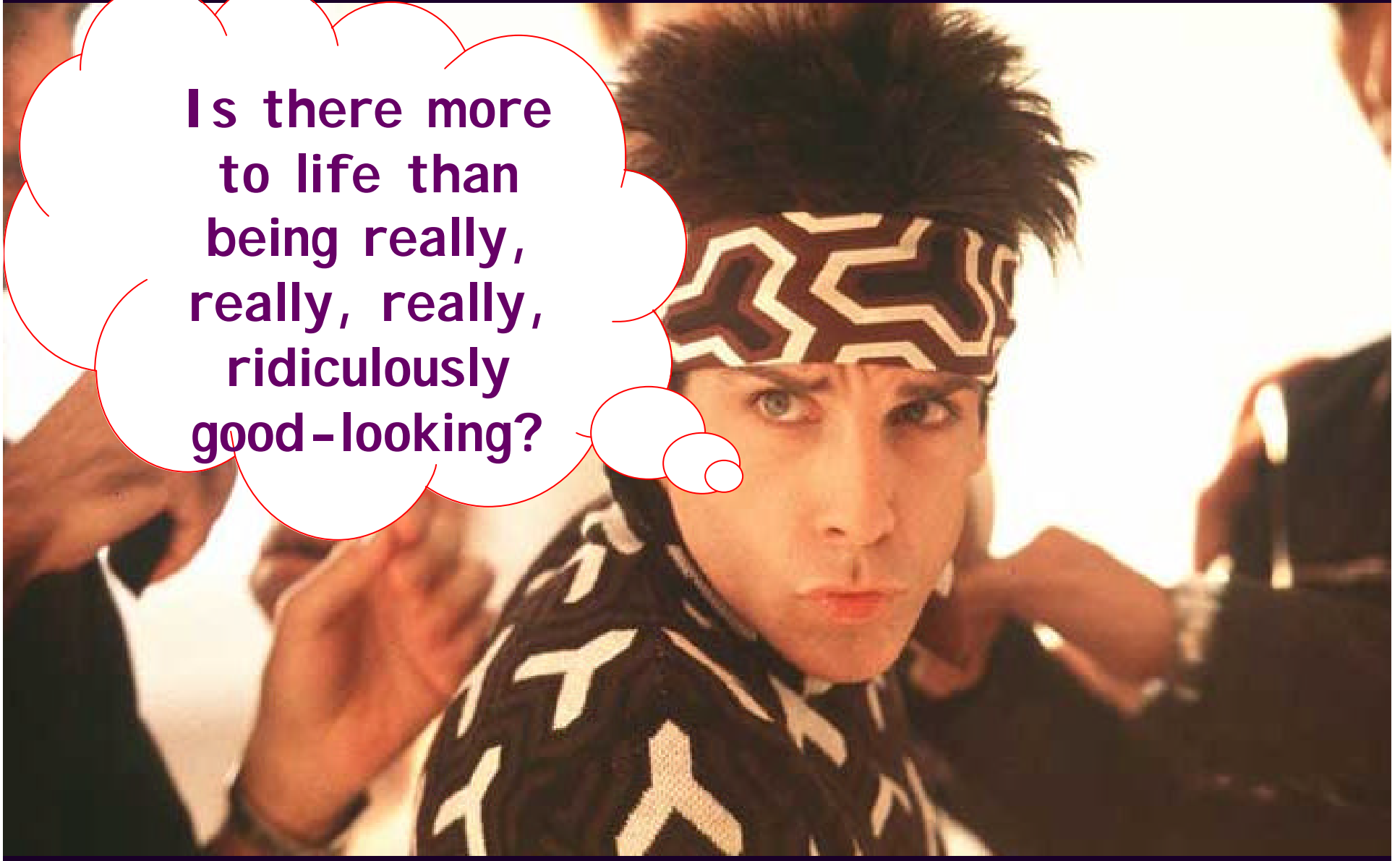
AS	AP	BF	HR	PR	SR	RA	PP	CP	BR	LA	LX	LO	QU	RI	SM	GH	GF	GV	B4	B5
65	269	0	20	222	33	96	768	672	47	0	75	65	100	70	2	73	77	65	2500	2500

With Vocalizations (video)

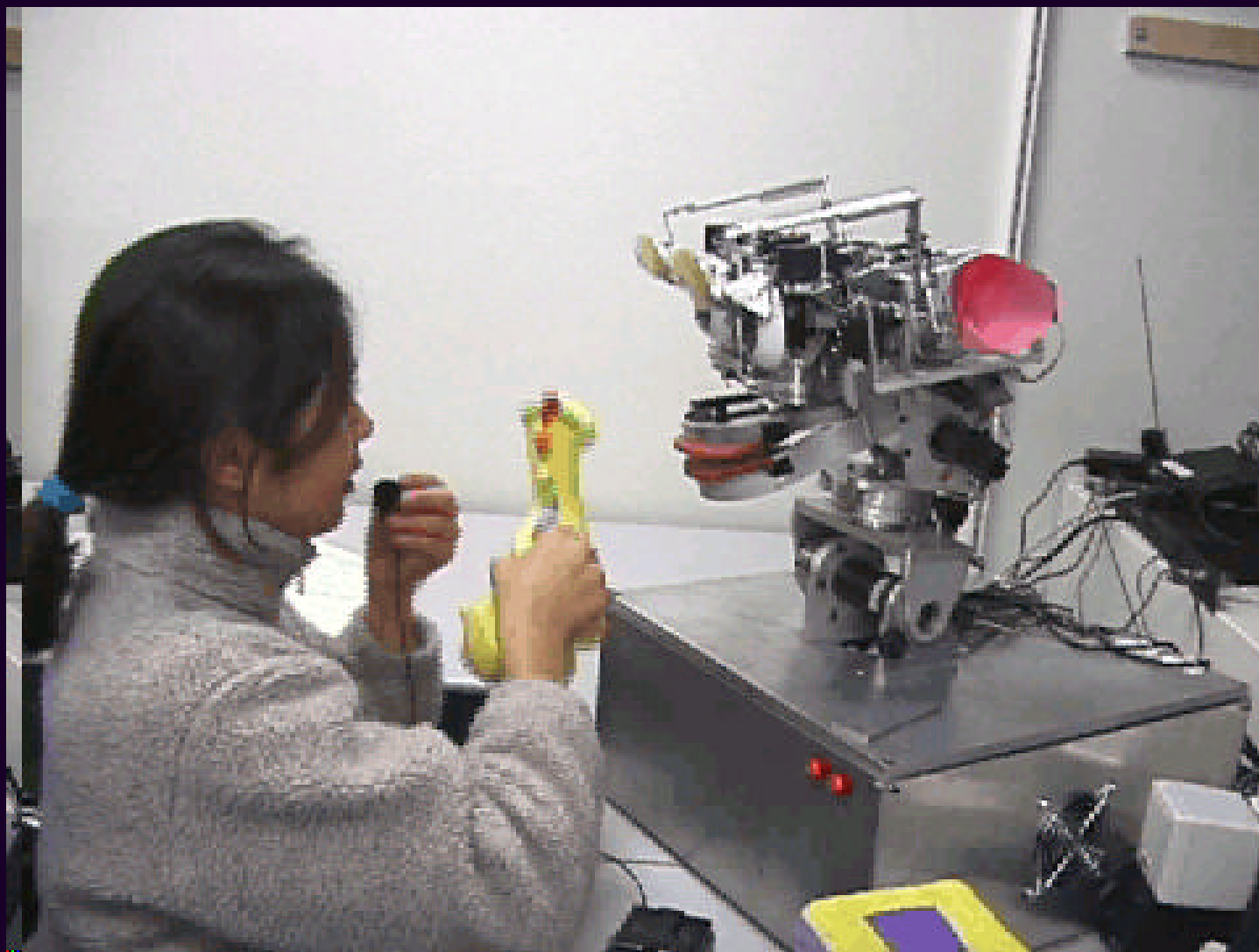


But...

**Is there more
to life than
being really,
really, really,
ridiculously
good-looking?**



Affective Intent (video)



Fernald's Results

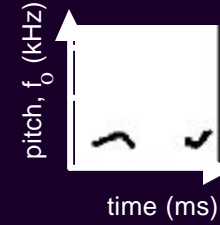
- Four cross-cultural contours of infant-directed speech
- Exaggerated prosody matched to infant's innate responses

That's a good bo-o-y!



approval

No no baby.



prohibition

Can you get it? Can you get it?



attention

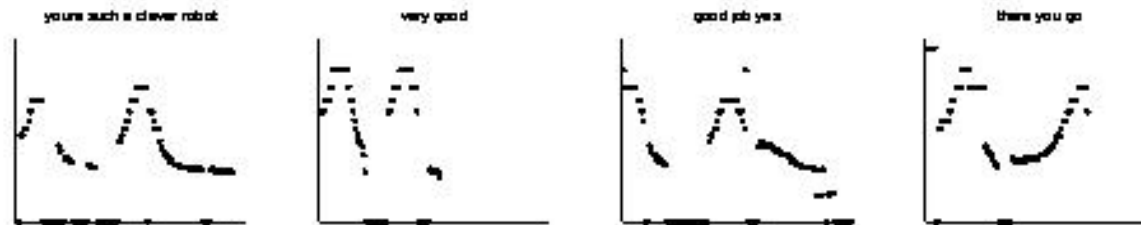
MMMM Oh, honey.



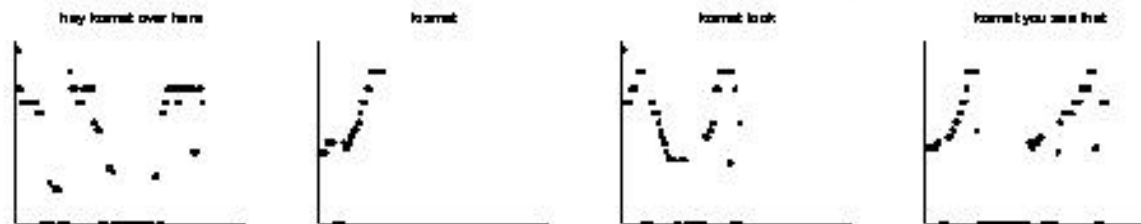
comfort

Evidence for Fernald-like Contours

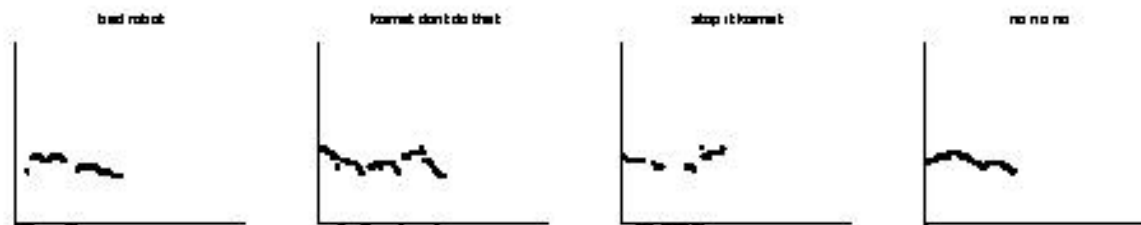
Approval



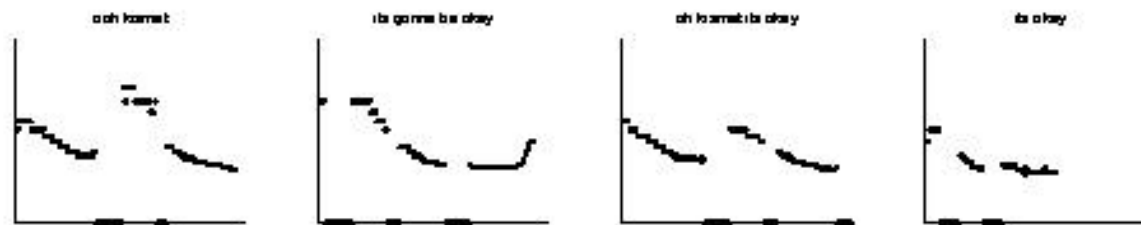
Prohibition



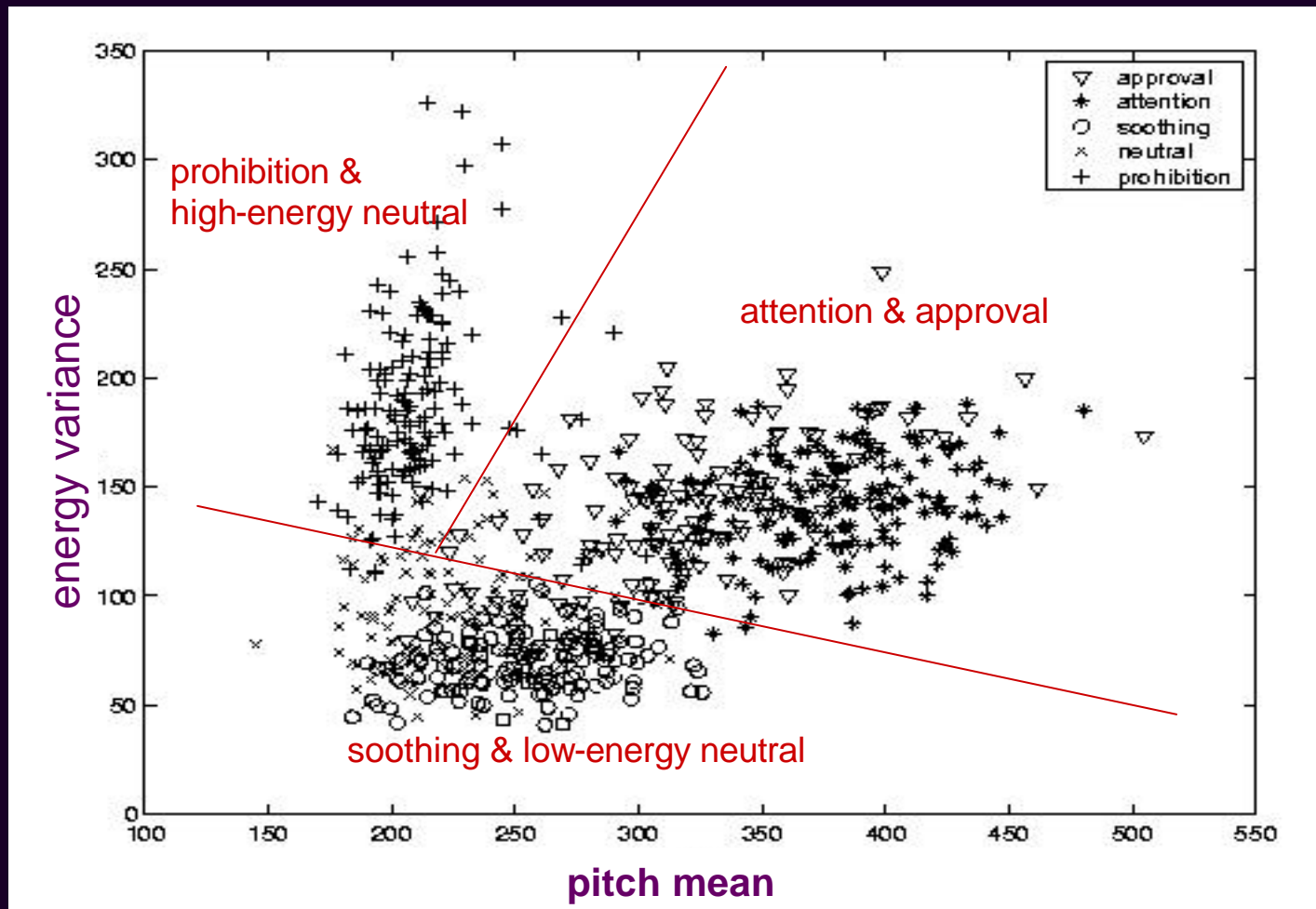
Attention



Soothing

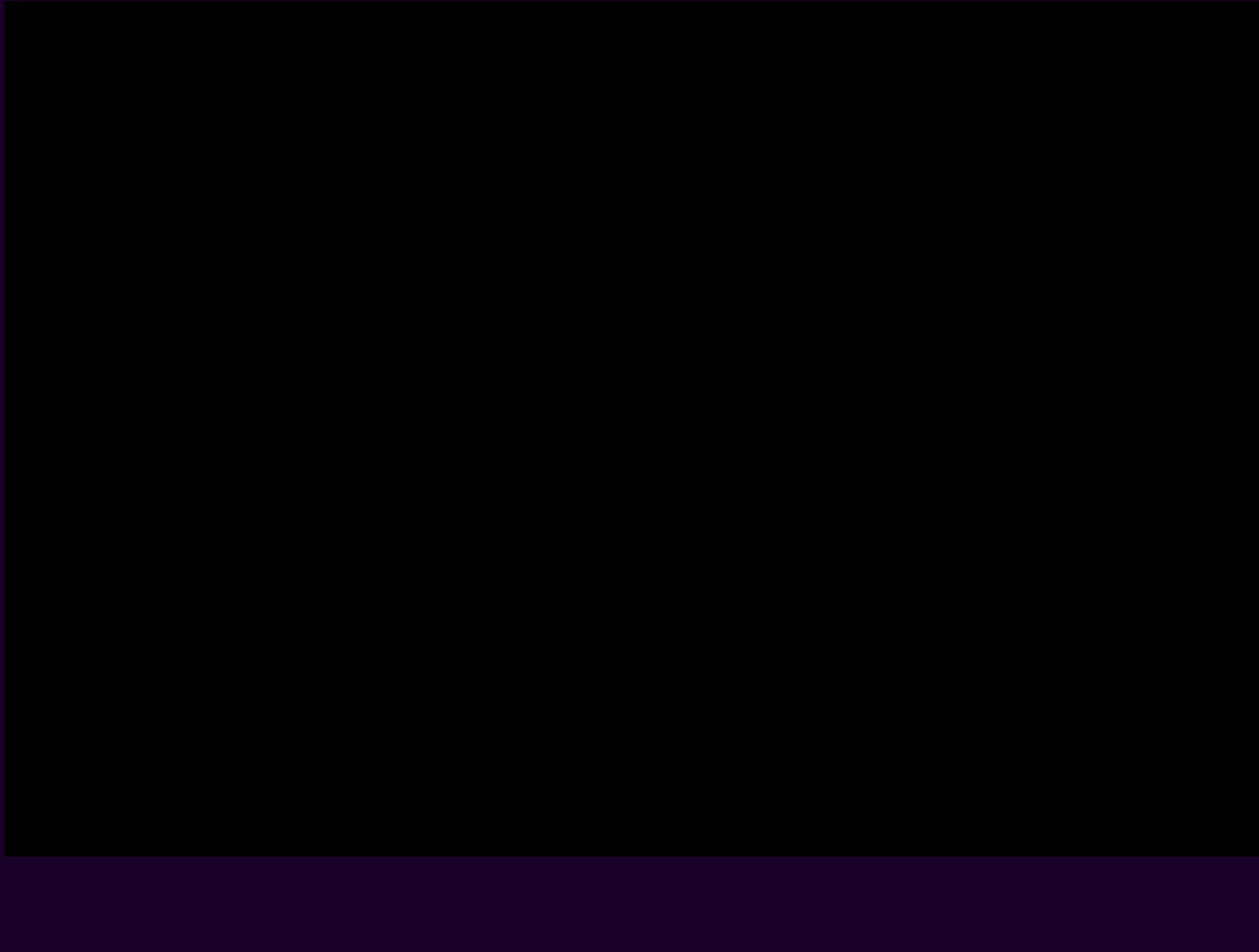


Performing Recognition



Breazeal & Aryananda, Humanoids 2000

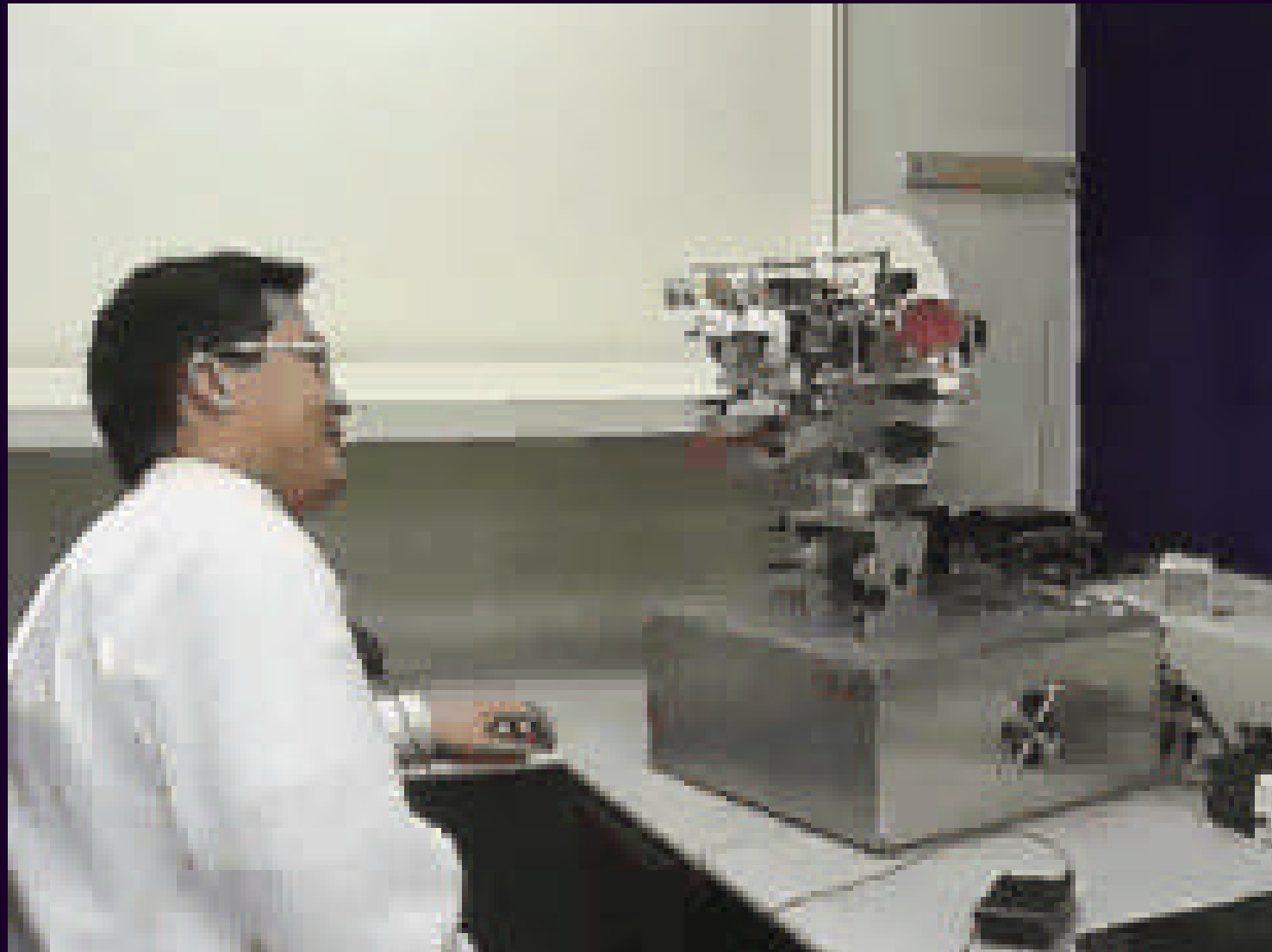
Examples (video)



Turn-Taking

- Cornerstone of human-style communication, learning, and instruction
- Four phases of turn cycle
 - relinquish floor
 - listen to speaker
 - reacquire floor
 - speak
- Integrates
 - visual behavior & attention
 - facial expression & animation
 - body posture
 - vocalization & lip synchronization

Example (video)



Conclusions

- Robots can partake in “infant-caregiver” interactions
- These interactions are rich with scaffolding acts
- Prerequisite for socially situated learning

Kismet’s really, really, ridiculously detailed web-pages:

<http://www.ai.mit.edu/projects/kismet/>