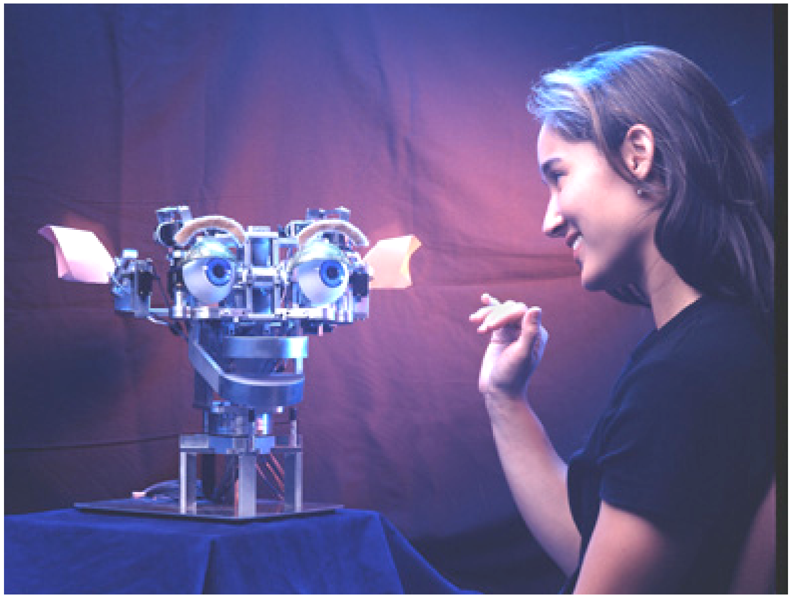


Sociable Machines:

Expressive Social Exchange Between Humans and Robots

Cynthia Breazeal
MIT Artificial Intelligence Lab

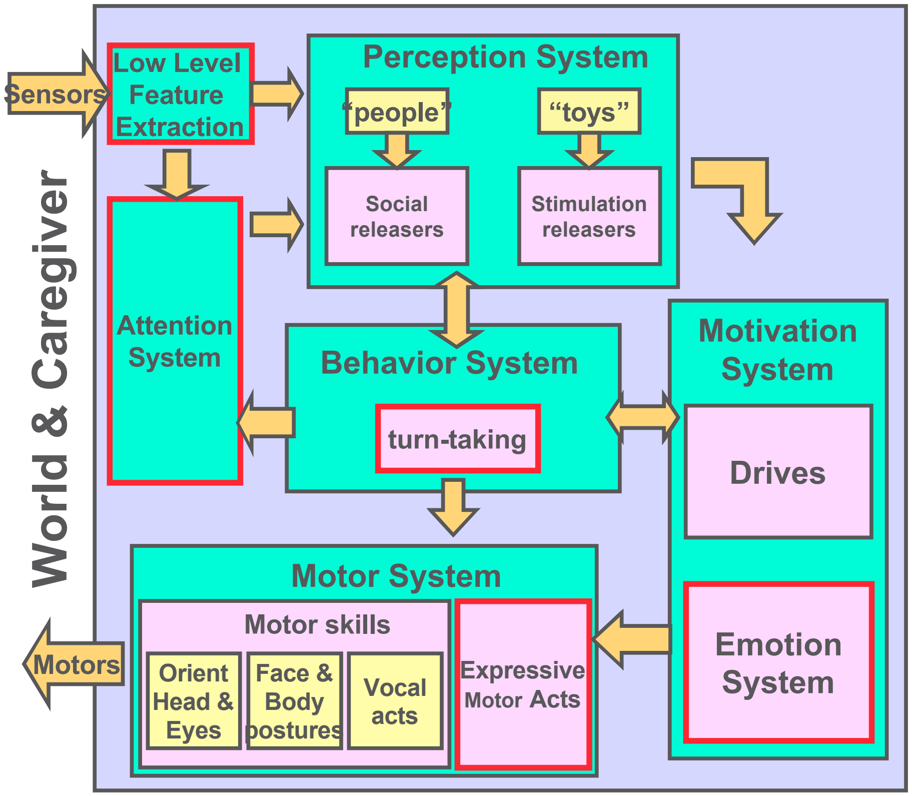
The Game Plan

- Today
  - Present recent advancements in building, Kismet, an autonomous robot that can engage people in human-style social exchange
  - Robot supports multiple communication modalities
    - » visual and auditory
    - » facial expression, gaze direction, body posture, vocalizations
  - Designed for infant-caregiver level of interaction
    - » physical, affective, pre-linguistic
  - Implementation is inspired by theories and models of psychology, social development, ethology, and evolutionary perspectives
    - » Not claiming explanatory power
    - » Not claiming equivalence to human psychological aspects
    - » Synthetic analogs that produce interesting behavior
  - Our immediate goal is to achieve natural interaction characteristic of infants with their caregiver
    - » building infrastructure to support socially situated learning
  - Our ultimate motivation is to explore issues in social learning that are characteristic of caregiver-infant scenarios
    - » pre-requisite is natural social interaction between robot and human
    - » Have not done learning yet – that's the topic of future work

The Game Plan

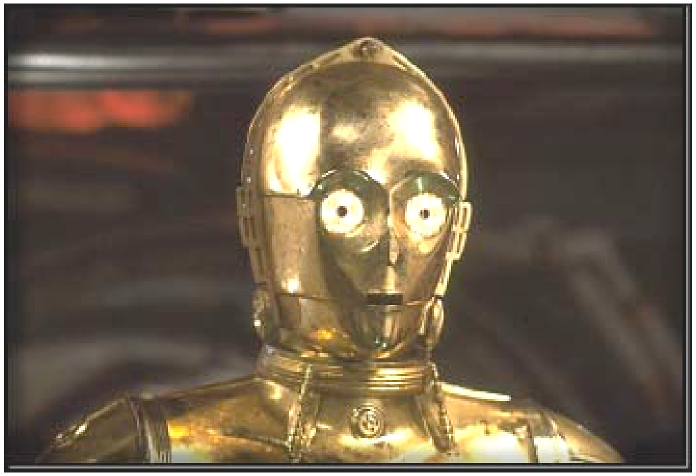
- Highlight
  - Design issues for building sociable robots
    - designing for the “human in the loop”
    - adds special challenges, constraints to the design problem
  - Discuss the implementation of four social skills that are argued to play an important role in social learning for human infants
    - “scaffolding” from developmental psychology
    - refers to various forms of assistance parents provide to their children to foster, shape, and facilitate learning
    - much of this assistance is social in nature
  - Evaluate Kismet’s performance with respect to four criteria of human-robot interaction
    - concerns both human’s performance as well as robot’s performance
  - Summarize results of several studies with naive subjects
    - subjective but quantifiable
    - focus on behavior of human-robot system

Scope of this Talk



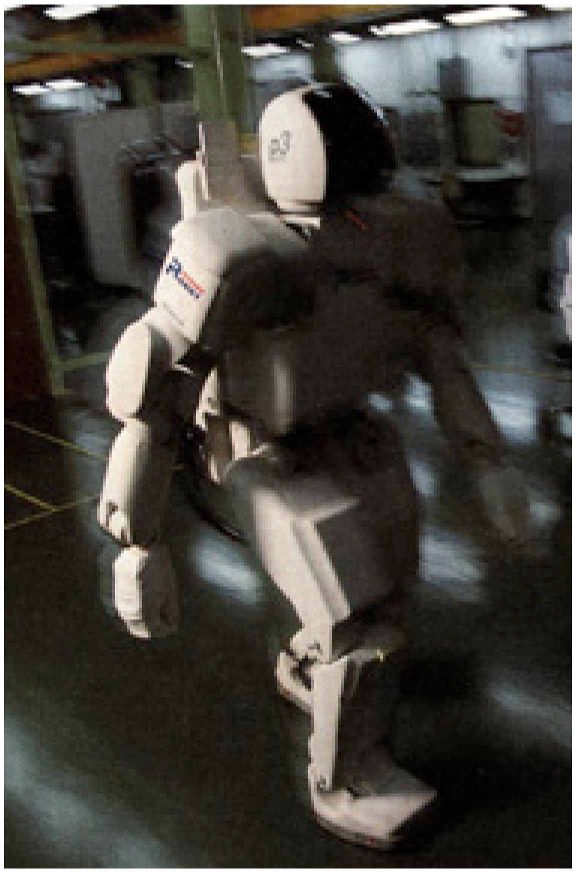
- Only going to present snap-shots of the overall system
- Physical, "affective" social interaction with humans has driven the design of a *highly* integrated system
  - all implemented in the same robot
  - robot situated in a social environment
  - robust, flexible behavior
  - Concentrate on social dimension

What is a “Sociable Machine”



- Sociable machines convey social awareness and intelligence
  - Humans effortlessly explain, predict, and respond to the robot’s observable behavior by attributing internal mental states to the robot (intents, beliefs, desires, feelings)
  - The robot does the same with respect to the human (theory of mind)
- The social dimension is fundamental to human intelligence and human nature
  - evolution, development, society and civilization
  - Personality, emotions, empathy, loyalty, friendship, etc.
- Gain understanding of human nature and social intelligence through engineering sociable machines

Sociable Robots
A Question of Interface



- **Growing research and commercial interest in building "Personal Robots"**
  - **Provide people with assistance in their daily lives**
  - **Adapt to person's preferences**
- **Intuitive, natural interface for untrained users (Reeves&Nass, HCI)**
  - **Humans are already experts in social interactions**
  - **Humans automatically and unconsciously respond socially and naturally to media/technology. (Takes effort to not do so)**
  - **If media/technology *adheres to human social expectations*, people find interaction enjoyable and feel empowered and competent**
  - **Holds for specialists and lay people**
- **Humanoid robots are well suited to this hypothesis**
  - **similar morphology**
  - **similar sensing**
  - **possible to send and receive social cues**

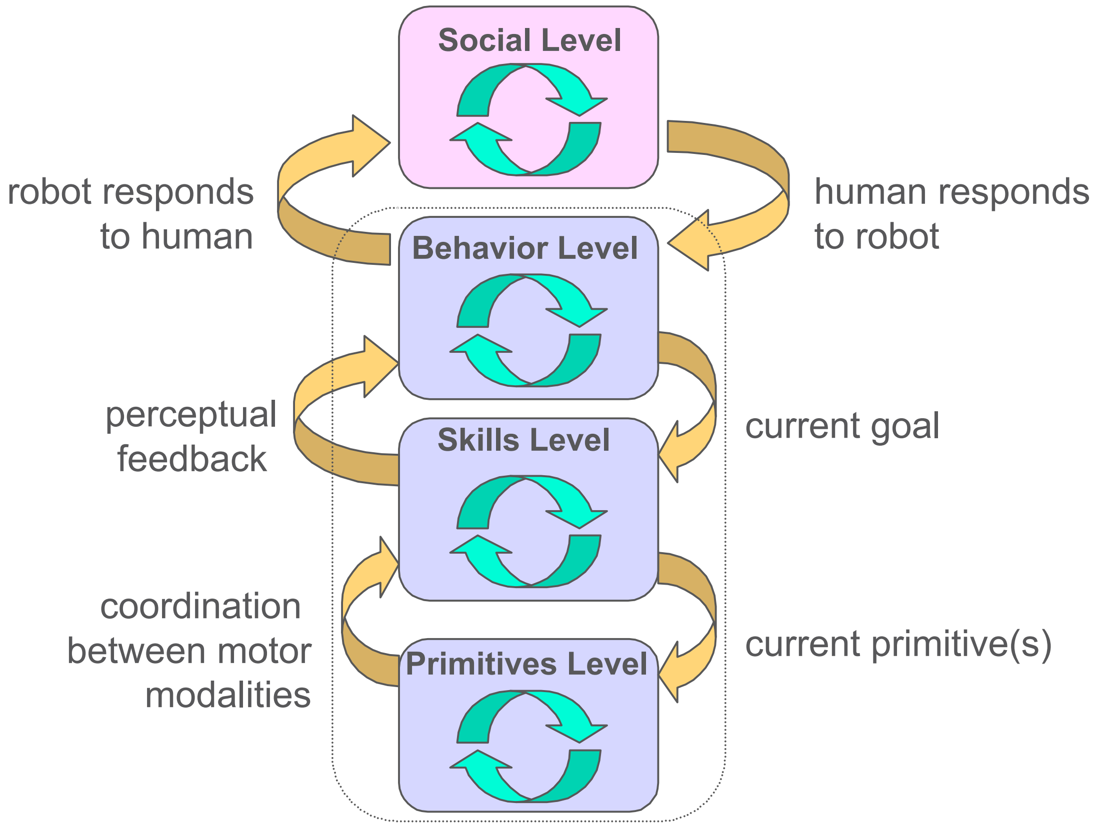
Socially Situated Learning

- “Personal” robots
  - Learning will be important, domain is very unconstrained
  - Robot is situated in a rich learning environment
  - Humans are motivated to help the robot learn

- Robot needs to be natural and intuitive to teach
  - Socially situated -- through social interaction, social cues
  - Assistance provided to robot through social acts
  - Robot can request guidance from human using social acts

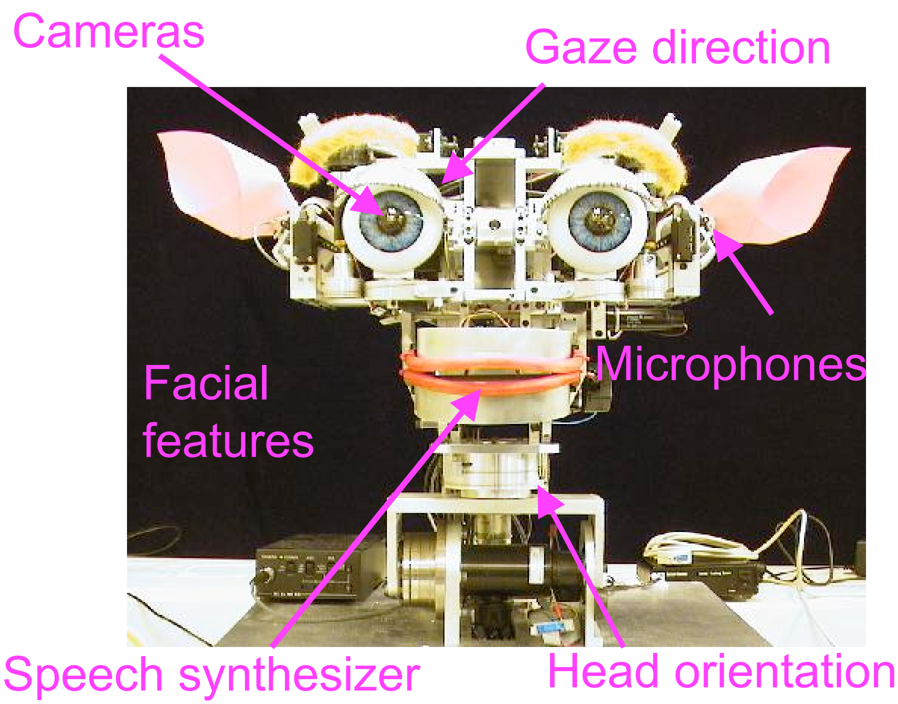
- Many of these social cues and skills could be used by a robot to address several key challenges in robot leaning
  - What’s salient?
  - How to measure progress? Especially for novel situations, tasks.
  - When has success been achieved?
  - How to constrain and structure learning problems?

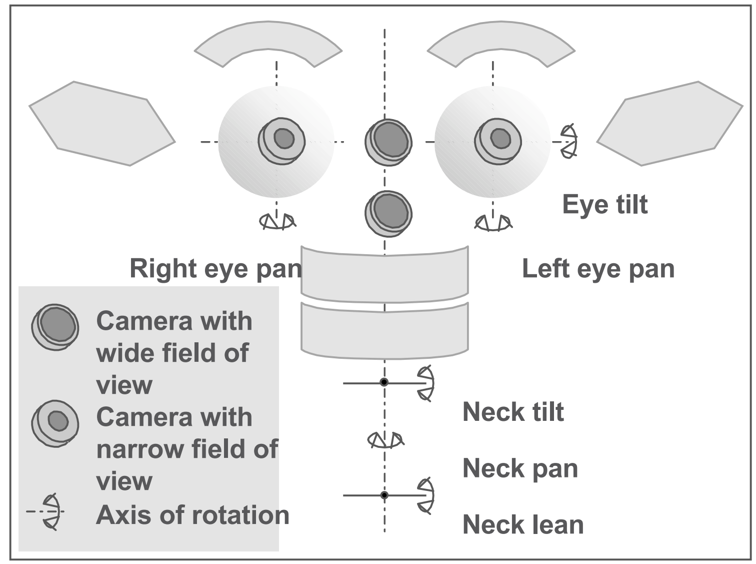
Levels of Control for Sociable Robots



- **Observable behavior generated through multiple levels of interaction**
- **Each level has different constraints**
  - **temporal**
  - **sensory feedback**
  - **information flow**
- **Social level adds significant constraints to control**
  - **Robots actions have communicative value and social significance**
  - **Robot's actions have the ability to manipulate the environment *far beyond their immediate physical consequence***
- **Design to benefit interaction for robot and human**
  - **matched to robots perceptual, physical, and computational limitations**
  - **Interaction is natural, intuitive, and enjoyable for person**

Designing for the Human In the Loop

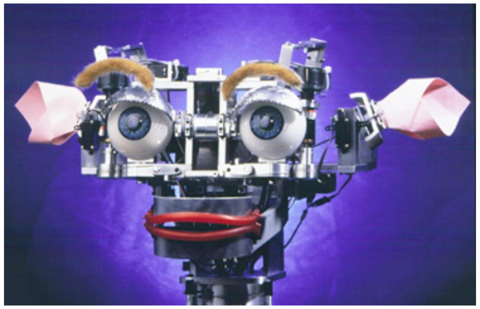




- **Social interaction with person constraints**
  - how robot appears
  - how its sensors are configured
  - its quality of movement
  - how it behaves, interacts
- **Support perception of human social cues**
  - vision, auditory
- **Display human social cues**
  - gaze direction, body posture
  - gesture, facial expression
  - vocal acts

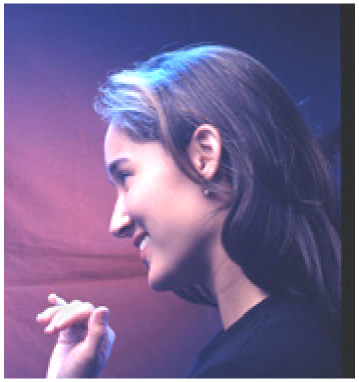
Design for Human-Robot System

perception,
attention,
motivation,
behavior,
expression



tune robot
to human

tune human
to robot

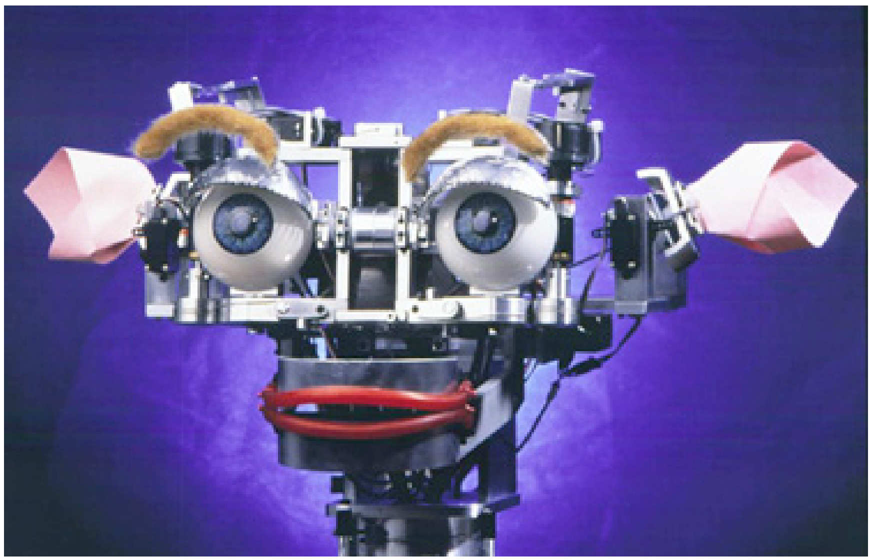


evolution,
development,
culture

- Design to support human nature
  - look at psychology, ethology, evolution, social development of infants
  - Naturally elicit desired complexity of interaction from humans
  - Match to human perceptions, human attention, human social behavior

- Audience perception
  - The design should establish the right social expectations
  - People naturally engage robot at right level of interaction
  - “believability” – treat robot as socially aware
  - “readability” – what person observes and interprets accurately reflects robot’s internal state

Design to support "simple" social interaction



- Design for infant-level social interactions
  - Eibl-Eiblsfeldt "baby scheme"
  - physical size
  - stature
- But not exactly human infant
  - caricature that is readable

- Naturally elicit scaffolding acts characteristic of parent-infant scenarios
  - directing attention
  - affective feedback, reinforcement
  - simplified behavior, suggested to make perceptual task easier
  - slow down, go at learner's pace
- Interaction based in
  - "affect"
  - physical
  - pre-linguistic

Four Competencies to Support Social Interaction and Learning for Kismet

- In this talk, focus on four social competencies:

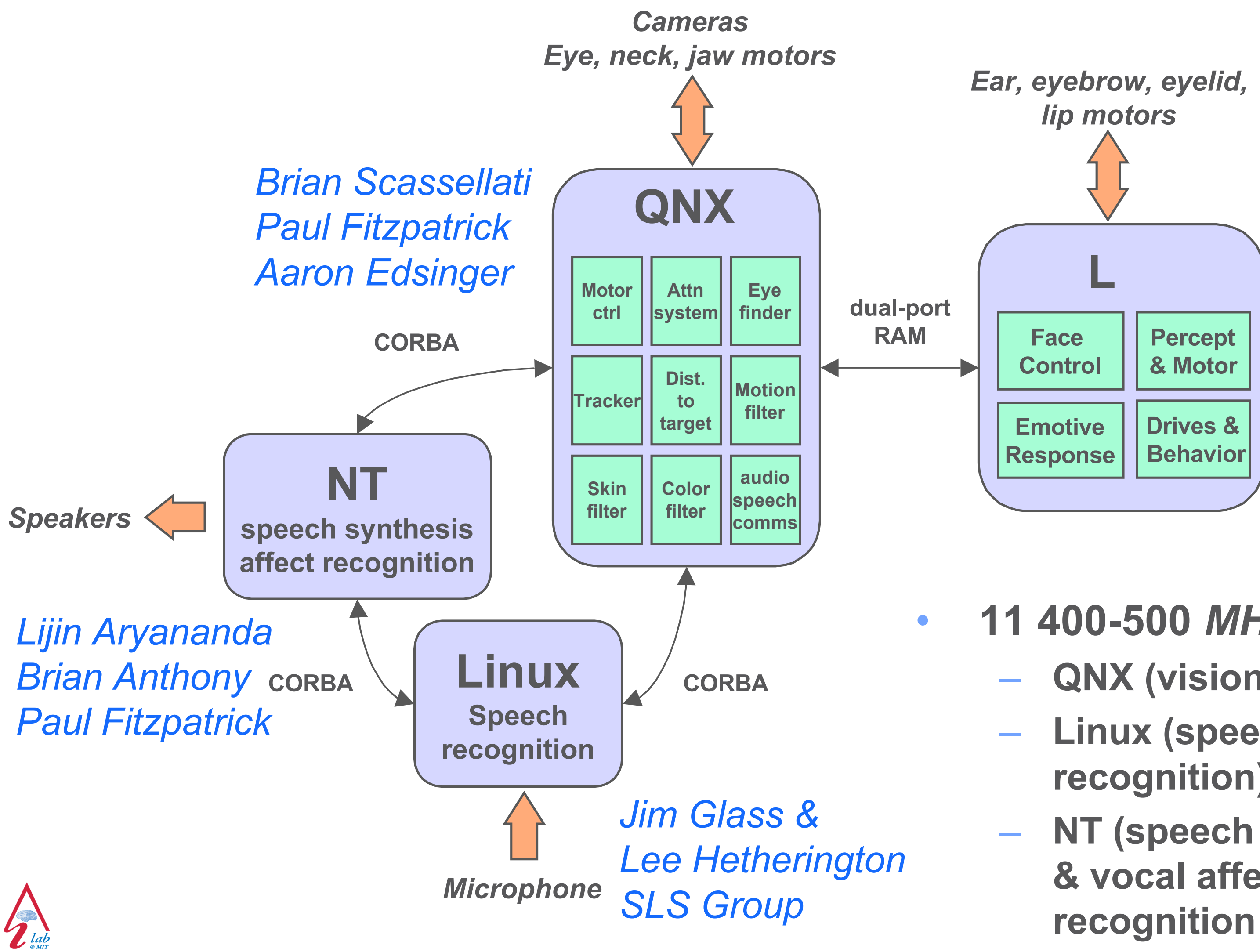
- **Direct Visual Attention**
  - indicate saliency
  - organize behavior about a locus of attention
  - robot should respond to natural attention directing cues of people
  - people should know what Kismet is attending to
  - useful for establishing shared reference

- **Recognize Socially Communicated Reinforcement**
  - voice as a natural instrument to convey praise, prohibition, etc.
  - robot should recognize affective feedback in robot-directed speech
  - people should know when Kismet understands their intent
  - useful as progress estimator, signal for goal attainment

Four Competencies to Support Social Interaction and Learning for Kismet

- **Communicate internal state to human**
  - important cue for human to predict and understand robot's behavior
  - important cue for human to tune their own behavior to robot
  - robot should convey its motivational and behavioral state to human in an intuitive manner
  - can be used to establish better quality instruction
- **Regulation of Social Interaction**
  - turn-taking is cornerstone of social interaction and social instruction
  - keep robot from becoming overwhelmed or under-stimulated.
  - structures learning episodes, interactive games, variations upon a theme

Large Scale Integration
(15 networked computers, many, many collaborators)



- 4 Motorola 68332 micro-controllers
- L, multi-threaded lisp
- higher-level perception, motivation, behavior, motor skill integration & face control

- 11 400-500 *MHz* PCs
  - QNX (vision)
  - Linux (speech recognition)
  - NT (speech synthesis & vocal affect recognition

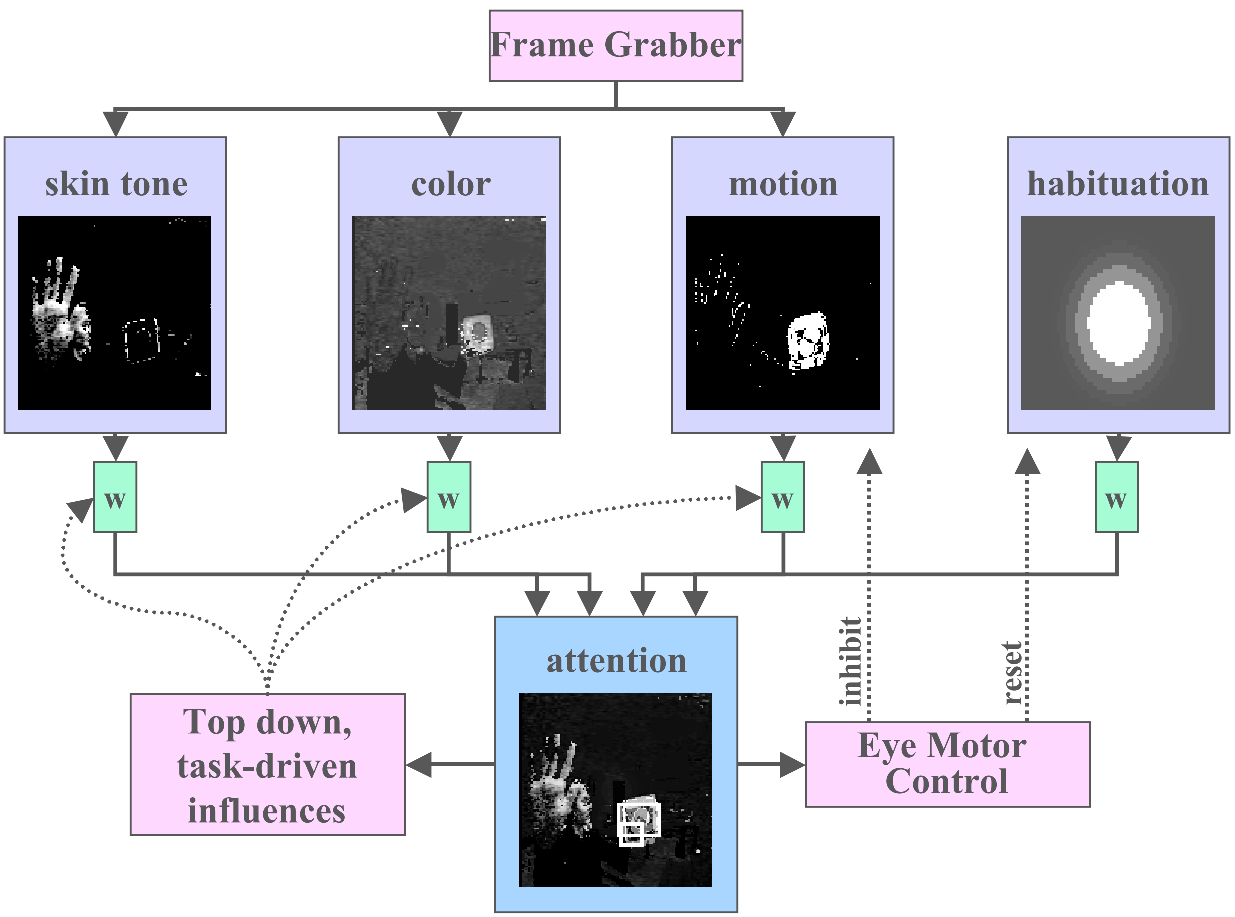
Evaluation Metrics

- Design for natural, intuitive, social interactions
- Must consider human response as well as robot's behavior
  - Subjective but quantifiable
- Four Criteria used to evaluate Kismet:
  - Can people intuitively read and do they naturally respond to Kismet's social cues?
  - Does Kismet readily elicit social cues (scaffolding) from the human that could be used to benefit learning?
  - Can Kismet perceive and appropriately respond to these naturally offered social cues?
  - Does the human adapt to the robot, and robot adapt to the human, in a way that benefits the interaction? (entrainment)

Four Social Competencies

- [ ] **Direct Attention**
- [ ] **Socially Communicated Reinforcement**
- [ ] **Readable Internal State**
- [ ] **Regulate Human-Robot Interaction**

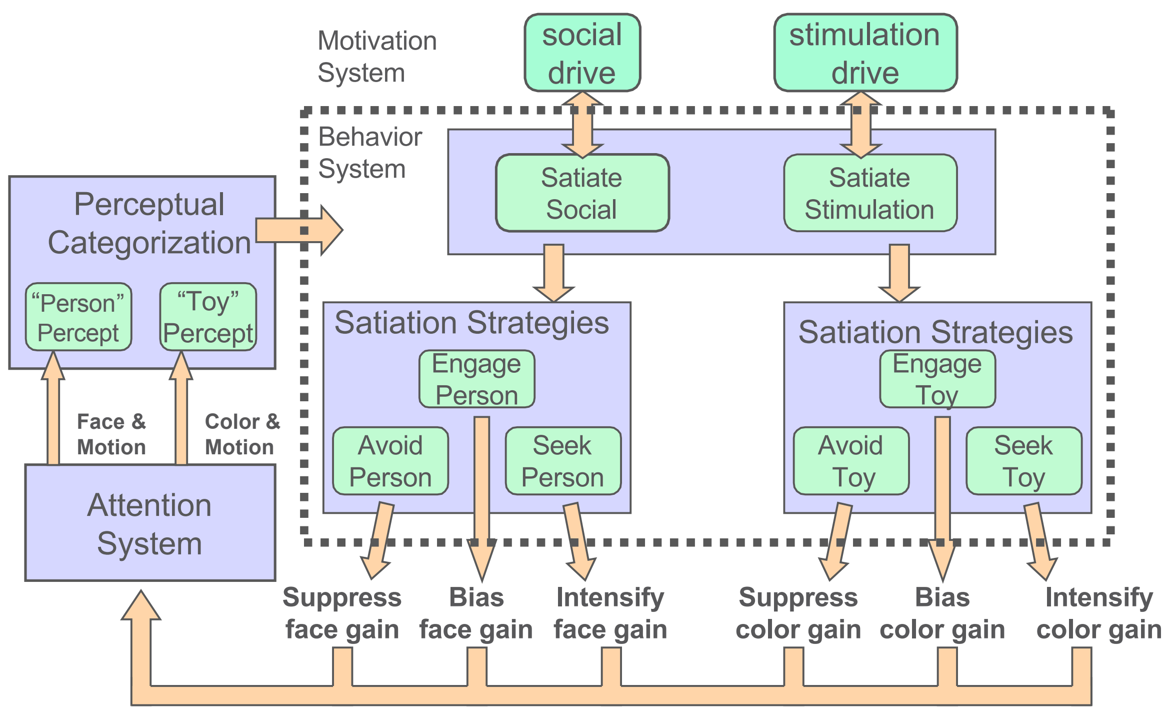
Visual Attention
(collaboration with B. Scassellati, and P. Fitzpatrick)



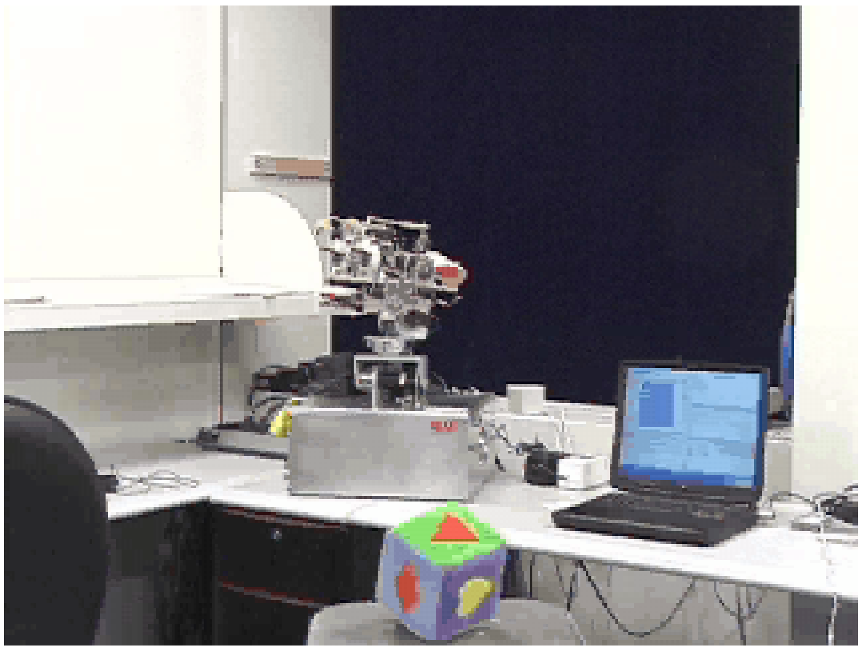
- Adapted from psychological model for human attention and visual search:
  - Wolfe's VGS 2.0
- For robot, extended model:
  - animate vision,
  - habituation,
  - time-varying goals
- Provides locus of attention within complex visual environment to organize behavior
- Human and robot should find same sorts of features salient
- Features need to match motivational requirements
- Robot's gaze direction is an important teaching cue

Dynamically Adjust Gains for Time-Varying Goals

- **Demonstrate "looking preference" for desired stimulus**
- **Motivational state determines what stimuli robot seeks out**
- **Behavioral state (the task) adjusts the gains**
- **Adjusting gains facilitates search**

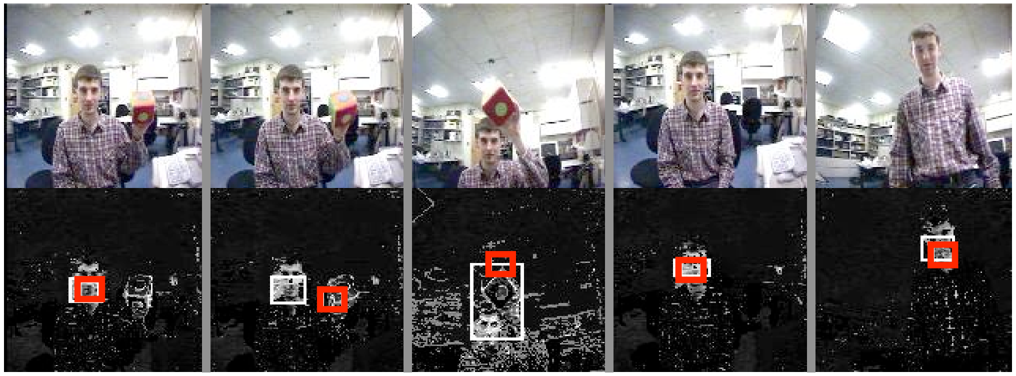


Video of Visual Search for Toy



- Robot's search is task-specific
  - seeks out desired stimuli
  - still opportunistic when appropriate
- Visual behavior conveys degree of commitment
  - fleeting glances
  - smooth pursuit
  - full body orientation
- Gaze direction, facial expression, and body posture convey robot's interest
- The attention and visual systems are always running and responding to external events
  - it's possible to divert attention by layering saliency features

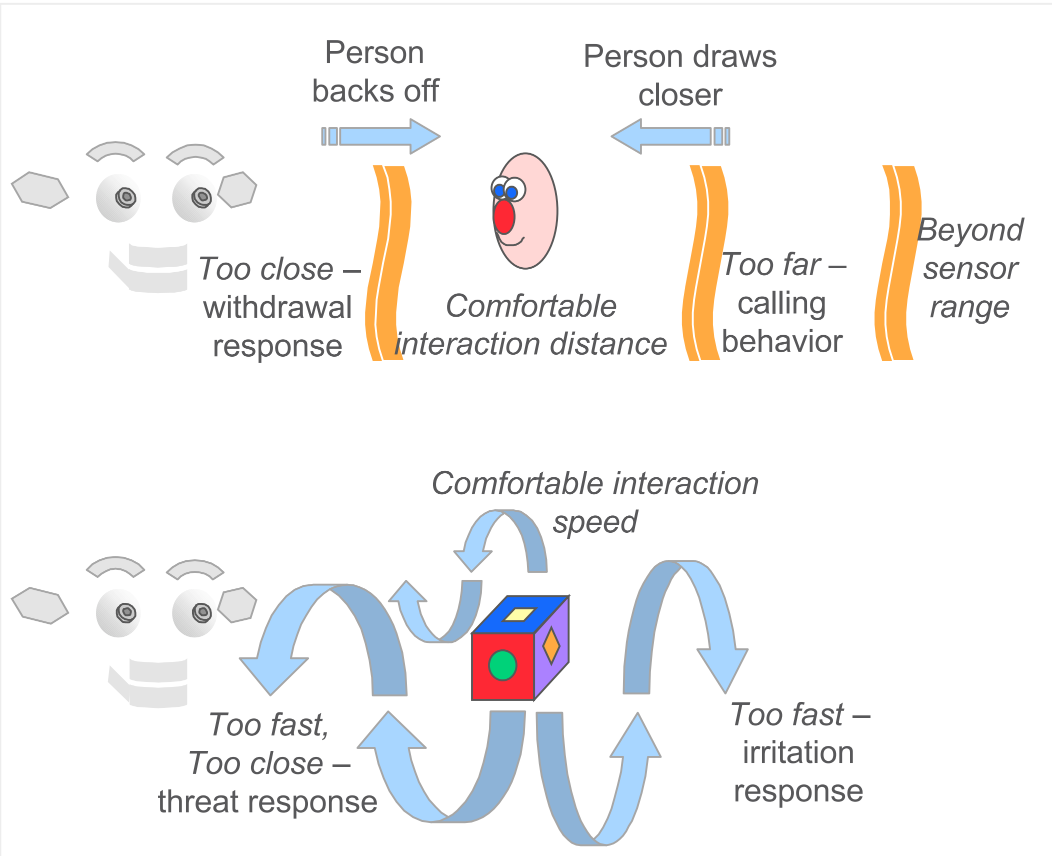
Directing Kismet's Attention and Gaze



stimulus category	stimulus	presentations	average time (s)	commonly used cues	commonly read cues
color & movement	yellow dinosaur	8	8.5	motion across centerline	change in visual behavior
	multi-colored block	8	6.5		
	green cylinder	8	6.0	shaking	face reaction
motion only	b/w cow	8	5.0		
skin-toned & movement	pink cup	8	6.5		
	hand	8	5.0	bringing object close	body posture
	face	8	3.5		
Total		56	5.8		

- Wide field of view tracker follows the most salient target
- Gaze directs foveal cameras to target and eyes smoothly pursue target.
- Can overlay cues to make object more salient to robot
- These cues are naturally the ones people use to direct attention

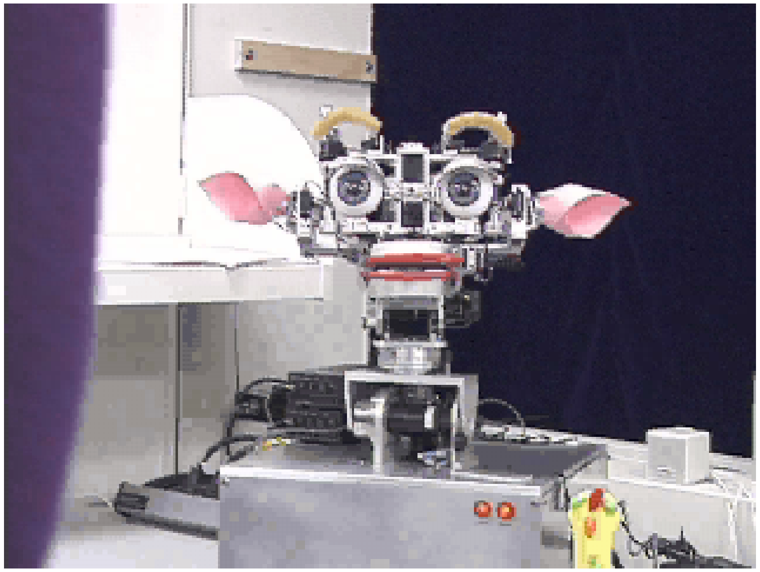
Social Constraints on Animate Vision



- Robot's visual behavior, facial expressions, and body posture have communicative value to the human observer
- The robot's signals can be used to **regulate** the nature of the interaction
  - To benefit robot's behavior
  - To benefit robot's perceptual processing
- These signals must be intuitive to understand for the human observer
  - correlated to emotive responses for coping with unfavorable stimuli
    - fear, disgust, frustration, distress
  - correlated to emotive responses for bringing robot near favorable stimuli
    - interest, joy

Video of Social Amplification

establish personal space



lure people in

- **naive subjects**
  - **robot establishes a "personal space" through expressive cues**
  - **regulates persons interaction to suit its visual processing limitations**

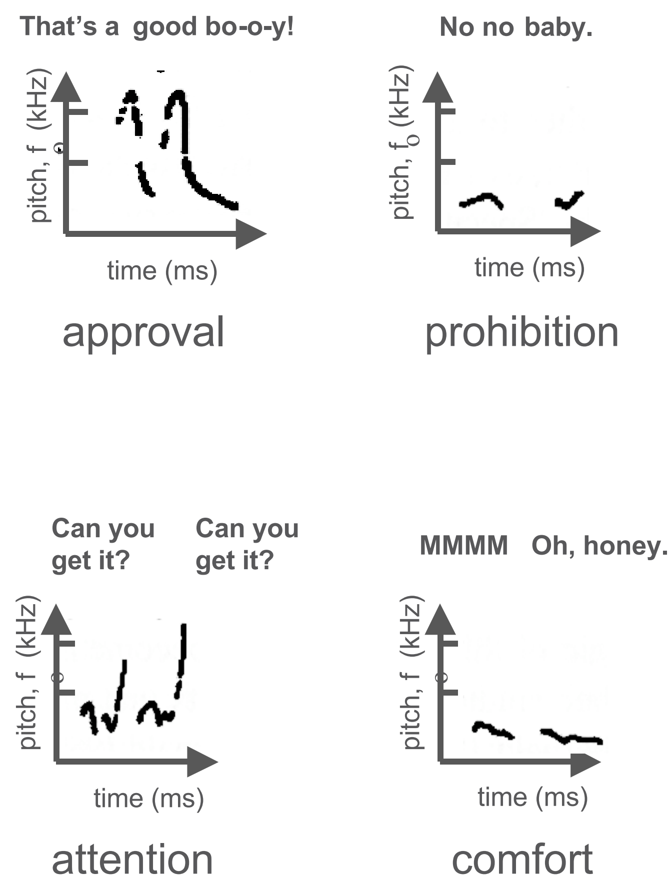
Four Social Competencies

- Direct Attention
- **Socially Communicated Reinforcement**
- Readable Internal State
- Regulate Human-Robot Interaction

Infant Recognition of Vocal Affective Intent
(Anne Fernald)

- **Four cross-cultural contours of infant-directed speech (A. Fernald)**
  - **praise,**
  - **prohibition,**
  - **soothing,**
  - **bids for attention**

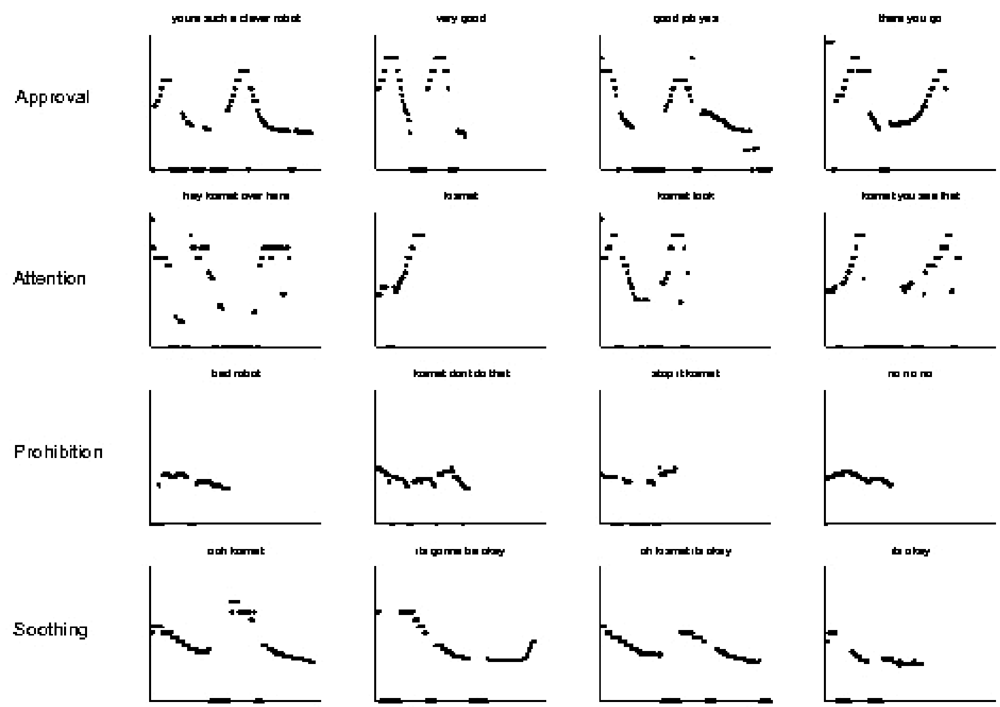
- **Exaggerated prosody is well matched to infant's innate responses to auditory stimuli.**
  - **modulates infant's arousal and attention**
  - **social referencing**



Vocal Affective Intent as Teaching Signal

(collaboration with L. Aryananda)

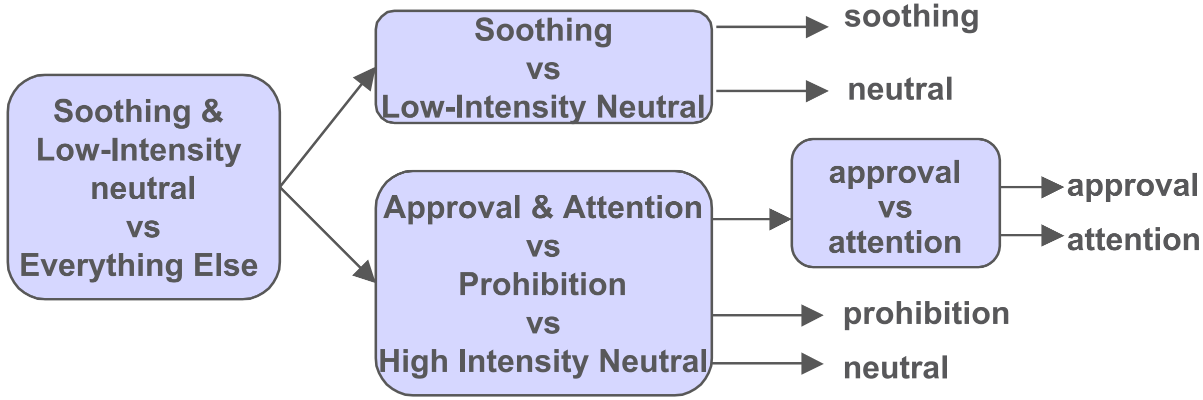
- **Adapt Fernald's research to Kismet**
- **Ultimately, use as training signal (robot does not learn yet)**
  - **Real-time recognition**
  - **Use as affective saliency marker (designed for *strong intent*)**
- **Affective classification modulates robot's affective state**
  - **could be used as reinforcement**
- **Socially situate with expressive feedback**

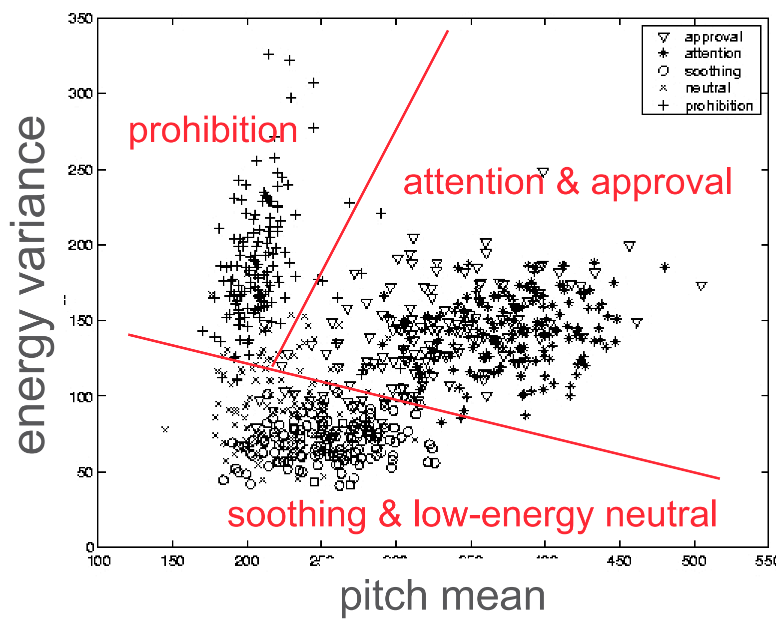


Evidence for 4 contours in Kismet-directed speech

Recognition of Affective Intent in Robot-Directed Speech

- Database of small number of female speakers
  - Recorded in robot's environment
  - Robot-directed speech with strong affect
- Multi-stage model
  - Each stage is simple for real-time performance
  - Some classes easily separated by arousal
    - energy, pitch
  - Off-the-shelf learning mechanism for the stages (Mixture of Gaussian with EM)
- Use Fernald contour information to distinguish *difficult* classes
    - attention vs approval
    - soothing vs low intensity neutral





Recognizing Affective Intent

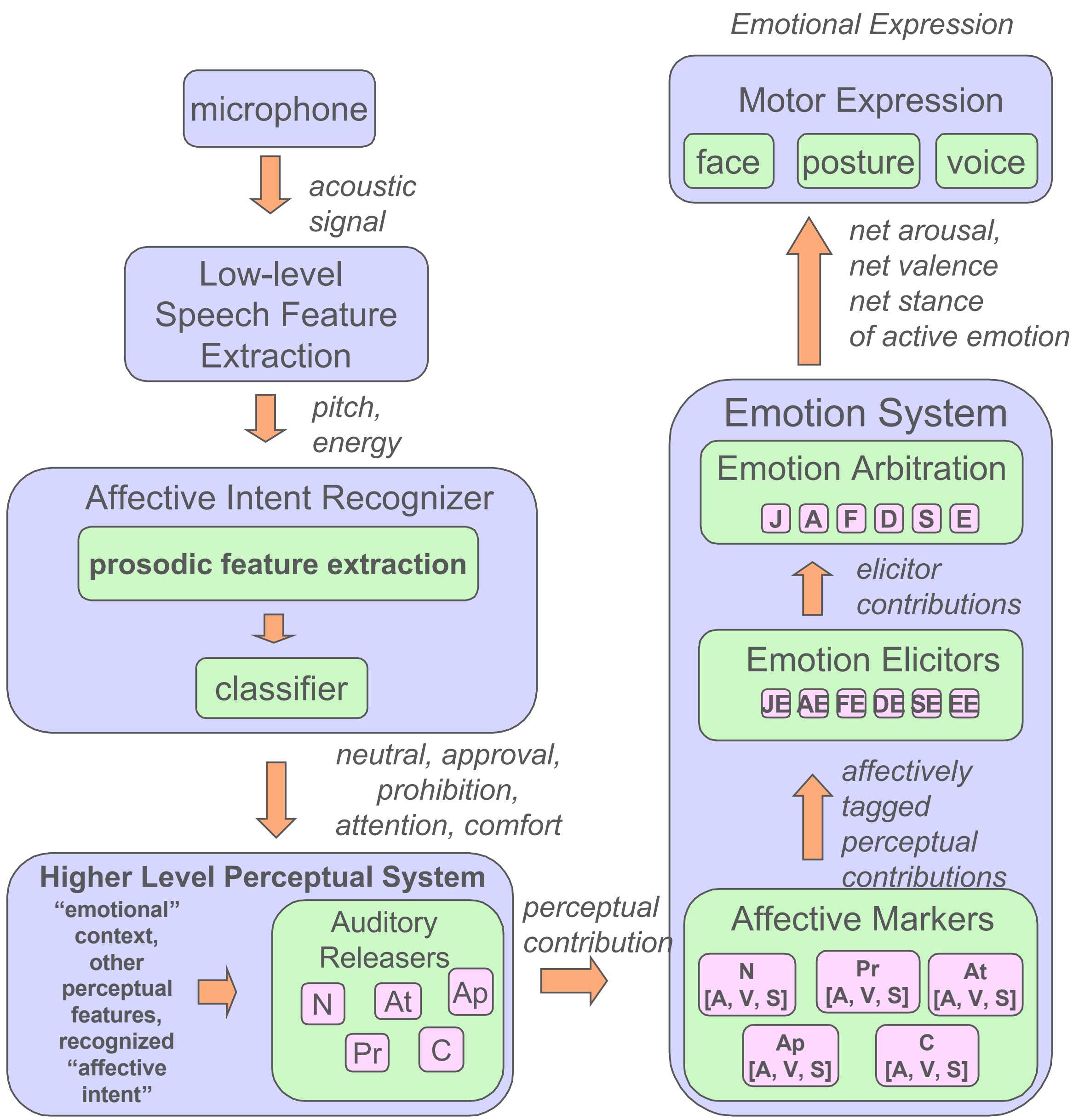
Results with Caregivers and Naive Subjects

Test set	Strength	Class	Test Size	Classification Result					% Correctly Classified
				Approval	Attention	Prohibition	Soothing	Neutral	
Caregivers		Approval	84	64	15	0	5	0	76.19
		Attention	77	21	55	0	0	1	74.32
		Prohibition	80	0	1	78	0	1	97.5
		Soothing	68	0	0	0	55	13	80.88
		Neutral	62	3	4	0	3	52	83.87
Naive speakers	Strong	Approval	18	14	4	0	0	0	72.2
		Attention	20	10	8	1	0	1	40
		Prohibition	23	0	1	20	0	2	86.96
		Soothing	26	0	1	0	16	10	61.54
	Medium	Approval	20	8	6	0	1	5	40
		Attention	24	10	14	0	0	0	58.33
		Prohibition	36	0	5	12	0	18	33.33
		Soothing	16	0	0	0	8	8	50
	Weak	Approval	14	1	3	0	0	10	7.14
		Attention	16	7	7	0	0	2	43.75
		Prohibition	20	0	4	6	0	10	30
		Soothing	4	0	0	0	0	4	0
		Neutral	29	0	1	0	4	24	82.76

- **Five classes of utterances**
  - **neutral speech**
  - **praise, prohibition, attention, soothing**
- **All Female speakers**
- **Multiple languages**
  - **French, German, Indonesian, English, Russian**
- **Objective scorer classifies as strong, medium, weak**
- **Good overall performance for strong instances**
  - **(random perf. = 20%)**
  - **very good for caregivers**
  - **good for naive subjects**
- **Acceptable misclassifications**
  - **minimal confusion of valence**
  - **some confusion of arousal**

Integrate Affective Intent Recognizer with "Emotion" System

- **Lends an additional stage of processing to help minimize misclassifications (soothing and neutral), adds an "inertia term"**
- **Output from recognizer passed through an appraisal phase**
  - **combined with "motivational" and behavioral context**
- **Tagged with "affective" information, passed to "emotion" elicitor processes**
- **If "affect" is strong enough, an "emotive" response is elicited**
- **"Emotive" response**
  - **changes "emotive" facial expression**
  - **changes "emotive" body posture**



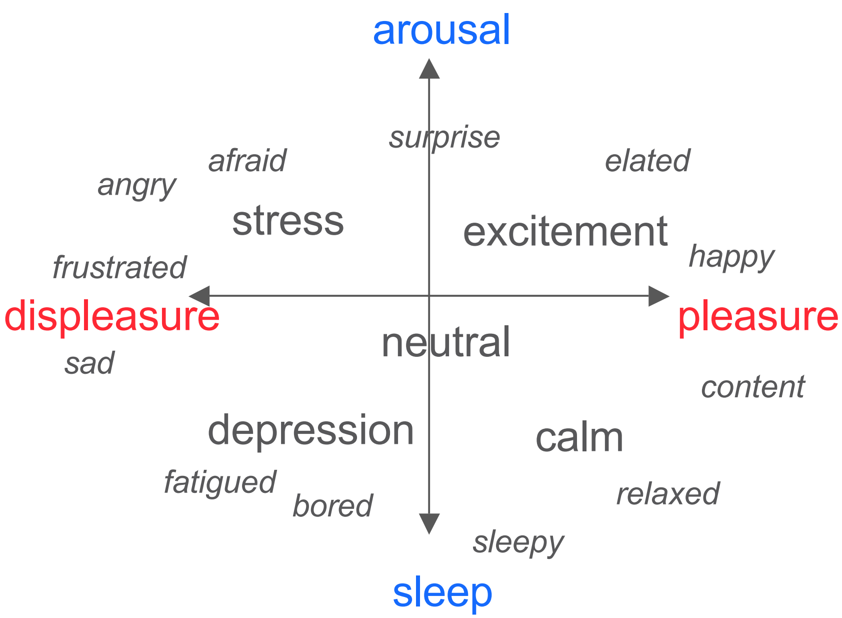
Evaluation with Naive Subjects

- 5 female subjects
  - 4 naive subjects
  - 1 caregiver
- Four contours and neutral speech
  - praise, prohibition, attention, soothing
- Multiple languages
  - French, German, Indonesian, English, Russian
- Driven by Human
- Socially situated with expressive feedback
  - expression easily read as signal of robot's "understanding"
  - Evidence of entrainment
    - affective mirroring
    - affective feedback

Four Social Competencies

- Direct Attention
- Socially Communicated Reinforcement
- **Readable Internal State**
- Regulate Human-Robot Interaction

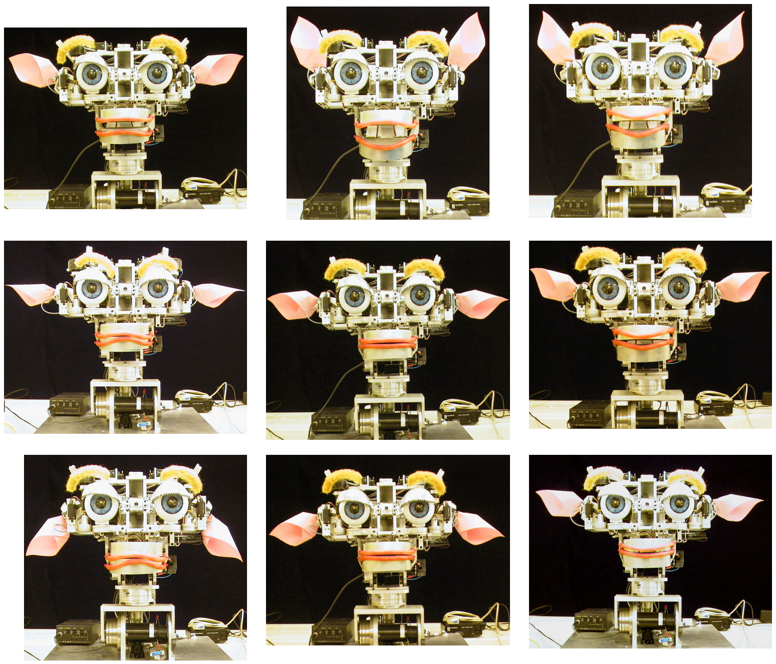
Facial Expressions
(Russell, Scott&Smith)



- **Emotive Facial expressions are a social signal**
- **Is communicated message contained in facial components or entire emotive facial expression?**

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

Generating Affective Facial Expressions

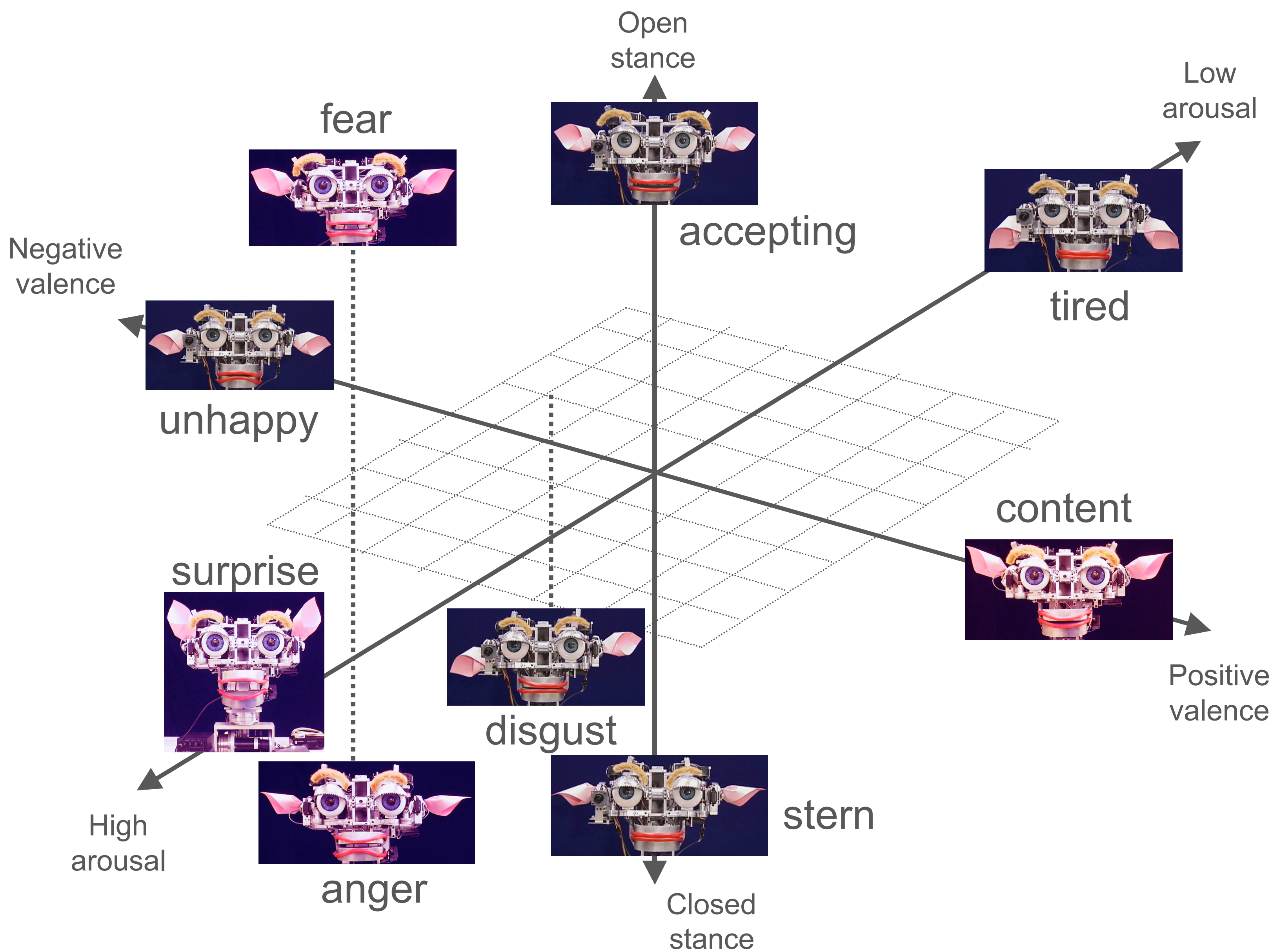


- Emotive Facial expressions are a social signal
- Important Feedback to the human observer
  - predictive of robot's behavior
  - indicative of robot's appraisal (future behavior)

- "Affective tags" of appraisal phase map to affect Kismet's dimensions (A, V, S)

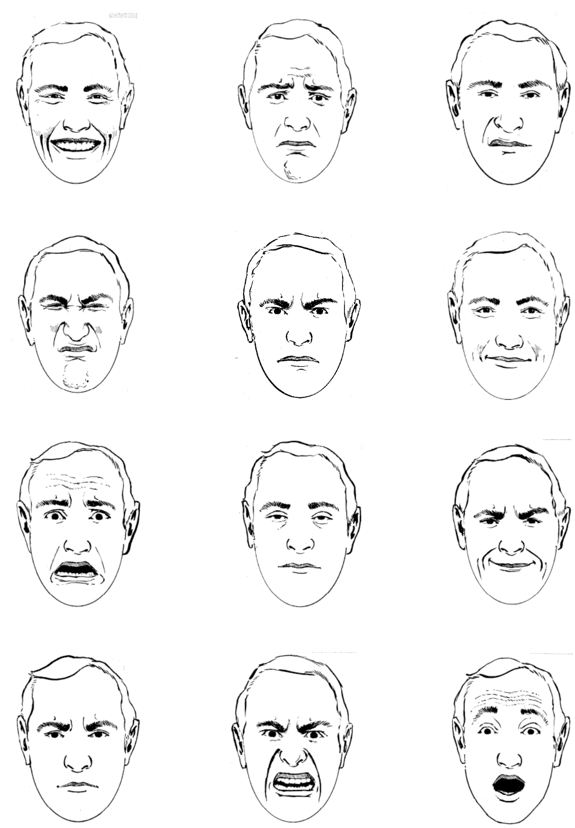
- Blended facial expression reflects the affective state of the robot

Mapping Facial Postures to Affect Space



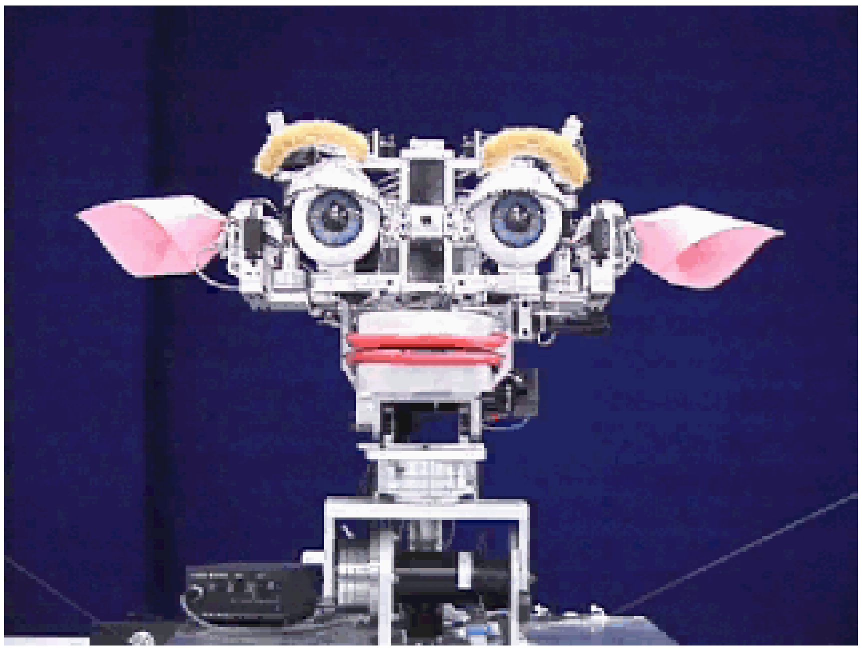
Evaluation of Emotive Expressions

	most similar sketch	percent age	comments
anger	**anger**	**100%**	shape of mouth and eyebrows are strongest reported cues
disgust	**disgust**	**80%**	shape of mouth is strongest reported cue
	sly grin	20%	described as “sneering”
fear	**fear**	**70%**	shape of mouth and eyes are strongest reported cues. Mouth open “aghast”
	surprise	10%	subject associates look of “shock” with sketch of “surprise” over “fear”
	happy	20%	lip mechanics cause lips to turn up at ends, sometimes confused with a weak smile
joy	**happy**	**70%**	report lips and eyes are strongest cues. Ears may provide arousal cue to lend intensity.
	content	10%	report lips used as strongest cue
	repulsion	10%	lip mechanics turn lips up at end, causing shape reminiscent of lips in repulsion sketch
	surprise	10%	perked ears, wide eyes lend high arousal. sometimes associated with a pleasant surprise
sorrow	**sad**	**90%**	lips reported as strongest cue. Low ears may lend to low arousal.
	repulsion	10%	lip mechanics turn lips up and end, causing shape reminiscent of repulsion sketch
surprise	**surprise**	**90%**	reported open mouth, raised brows, wide eyes and elevated ears all lend to high arousal
	happy	10%	subject remarks on similarity of eyes, but not mouth
pleased	**content**	**90%**	reported relaxed smile, ears, and eyes lend low arousal and positive valence
	sly grin	10%	subject reports the robot exhibiting a reserved pleasure. Associated with the “sly grin” sketch
sly grin	**sly grin**	**50%**	lips and eyebrows reported as strongest cues
	content	30%	subjects use robot’s grin as the primary cue
	stern	10%	subject report’s the robot looking “serious”, which is associated with “sly grin” sketch
	repulsion	10%	lip mechanics curve lips up at end. Subject sees similarity with lips in “repulsion” sketch
stern	**stern**	**60%**	lips and eyebrows are reported as strongest cues
	mad	10%	subject reports robot looking “slightly cross”. Cue on robot’s eyebrows and pressed lips.
	tired	20%	subjects may cue in on robot’s pressed lips, low ears, lowered eyelids
	sly grin	10%	subject reports similarity in brows.



random chance = 8%

Video of Emotionally Expressive Face, Voice, Posture



- Evaluated Kismet's expressive cues on naive subjects
  - Ranging in age 12—54
  - 9—17 subjects per study
- Human sketch comparisons
  - 70%--90% for "basic" emotion
- Still images, forced choice
  - 47%--83%
- Video of posture&expression, forced choice
  - 57%--86%
- Pre-recorded vocalizations, forced choice
  - (babbles and English phrases)
  - 50%--70% recognition
  - except "surprise" often confused for "fear"

Four Social Competencies

- Direct Attention
- Socially Communicated Reinforcement
- Readable Internal State
- Regulate Human-Robot Interaction

Turn-Taking and Proto-Dialog

- Already talked about social amplification (slower time scales)
- Cornerstone of human-style communication, learning, and instruction
- Turn-taking is fine grained regulation of human's behavior
  - use of envelope displays
  - shifts of gaze
  - facial expressions
  - shifts of body posture
  - Tightly coupled dynamic of contingent responses to other
- 4 phases of turn cycle
  - relinquish floor
  - listen to speaker
  - reacquire floor
  - speak
- Integrates
  - visual behavior & attention
  - facial expression & animation
  - body posture
  - vocalization & lip synchronization

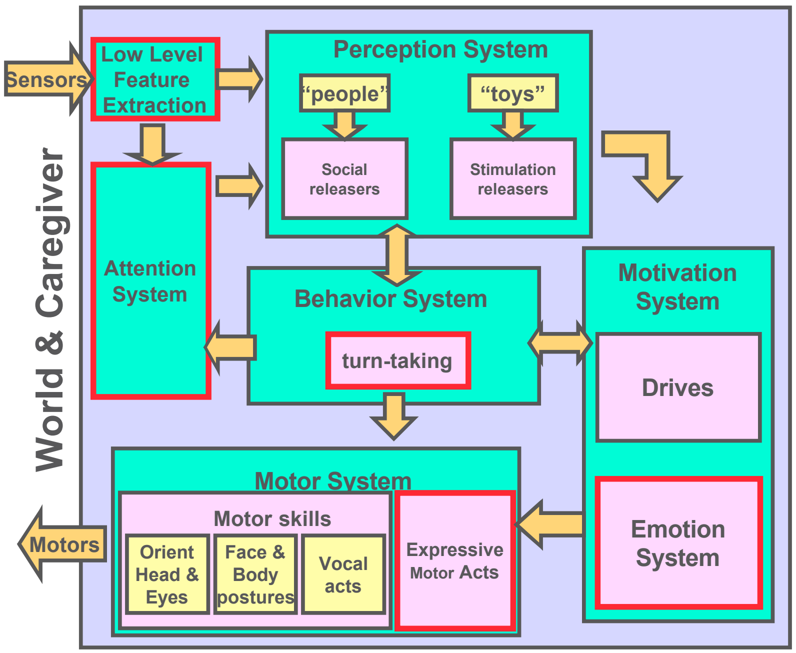
Edited down from long interactions 10-15mins

Evaluation of Performance

- Naive subjects
  - ranging in age from 25 to 28
  - All young professionals.
  - No prior experience with Kismet
  - video recorded
- Turn-taking performance
  - 82% “clean” turn transitions
  - 11% interruptions
  - 7% delays followed by prompting
- Significant flow disturbances
  - tend to occur in clusters
  - 6.5% of the time, but rate diminishes
- Evidence for entrainment
  - shorter phrases
  - wait longer for response
  - read turn-taking cues
  - 0.5—1.5 seconds between turns

		time stamp (min:sec)	time between disturbances (sec)
subject 1	start @ 15:20	15:20 – 15:33	13
		15:37 – 15:54	21
		15:56 – 16:15	19
		16:20 – 17:25	70
	end @ 18:07	17:30 – 18:07	37+
subject 2	start @ 6:43	6:43 – 6:50	7
		6:54 – 7:15	21
		7:18 – 8:02	44
	end @ 8:43	8:06 – 8:43	37+
subject 3	start @ 4:52 min	4:52 – 4:58	10
		5:08 – 5:23	15
		5:30 – 5:54	24
		6:00 – 6:53	53
		6:58 – 7:16	18
		7:18 – 8:16	58
		8:25 – 9:10	45
	end @ 10:40 min	9:20 – 10:40	80+

Contributions



- **Have only presented snap-shots of the overall system**
  - **incorporates diverse theories and concepts from psychology, social development, etholology, evolutionary theory**
- **Social interaction of this kind has driven the design of a *highly* integrated system**
  - **all implemented in the same robot**
  - **robot situated in a social environment**
- **Identified important issues for designing for the human-in-the-loop**
  - **audience perception**
  - **social implications of robot's actions**
  - **mutual influence to benefit interaction**
- **Illustrated them through robotic design**

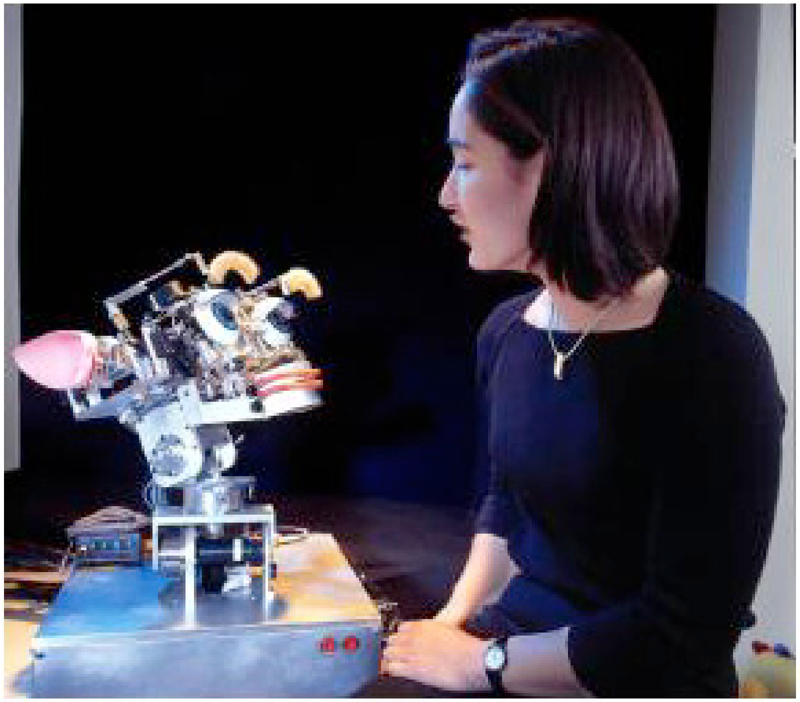
Contributions

- Implemented four social skills:
  - Directing attention
  - Recognition of praise, prohibition, bids for attention, soothing, & neutral robot-directed speech
  - Readable expressive abilities
  - Regulation of human-robot interaction
- Demonstrated how these skills foster robot's social interactions with people
  - benefits people -- natural, intuitive, and enjoyable
  - benefits robot -- can perform within its limitations
- Discussed how they could play an interesting part in socially situated learning
- Evaluated Kismet's social performance with naive subjects
  - robot's ability to elicit social cues
  - robot's ability to interpret and respond to them
  - human's ability to read Kismet's social cues
  - How this can benefit human-robot interaction -- entrainment

Contributions

- Found promising evidence that supports Kismet's ability to elicits the desired behavior from people
  - Evidence for entrainment to Kismet
  - They provide natural scaffolding
  - They initiate simple interactive games
  - They interpret and respond to Kismet's social cues
  - They appear to treat Kismet as a socially aware being
  - They appear to treat Kismet's acts as meaningful
  - They appear to respond to Kismet's cues in a way that Kismet can regulate their interaction

Future Work and Research Directions



- **Now we can begin the learning phase...**
  - **Explore socially situated learning**
  - **Particular interest in social modeling theory**
  - **acquisition of a proto-language**

- **First robotic system of this social complexity.**
  - **We can start to carefully study social human-robot interaction**
  - **promising results from initial studies**

Thanks!

I wouldn't be here, nor would Kismet, without you!

- **Advisor & Committee**
  - Rod, for believing in me and giving me the freedom to pursue numerous outrageously ambitious projects
  - Eric, Justine, Roz
- **Cog Group**
  - Brian Scassellati, Paul Fitzpatrick, Lijin Aryananda
  - past and present group members
- **Media Lab**
  - Bruce Blumberg and synthetic characters group
- **Spoken Language Systems Group of LCS**
  - Jim Glass and Lee Hetherington
- **Friends and family**
  - Mom, Dad, and brother Bill
  - Brian Anthony (more than just a pretty face)
  - The dreamers: Dave Barrett, Anita Flynn
  - The artists: Mike Hlavac, James Alser
  - ...and so many more!