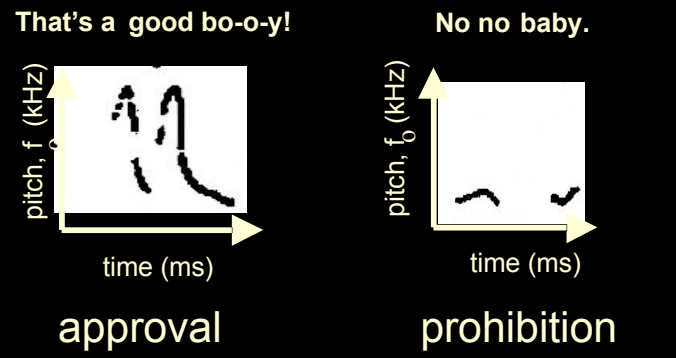
Building Robots with Presence

Cynthia Breazeal
MIT Media Lab
Robotic Life Group

Study 1: Affective Interactions

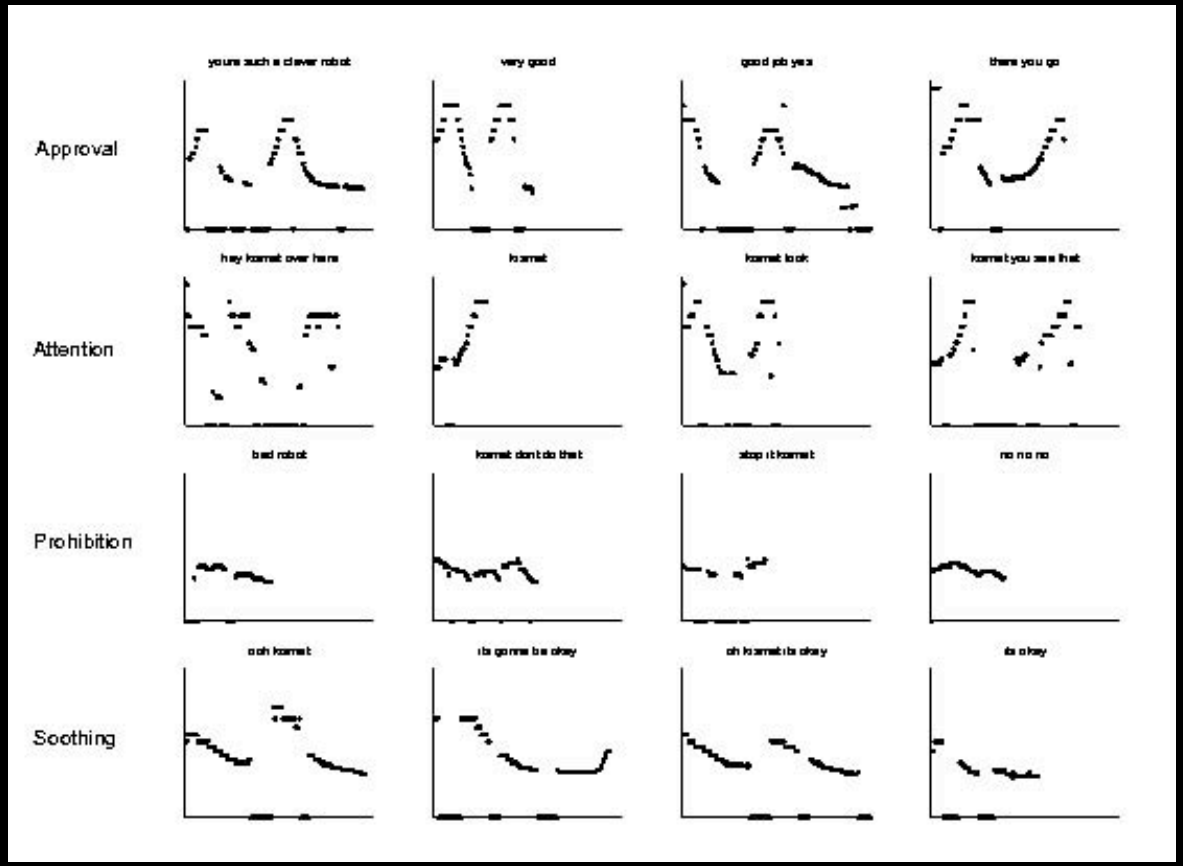
Recognition of Vocal Affective Intent

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

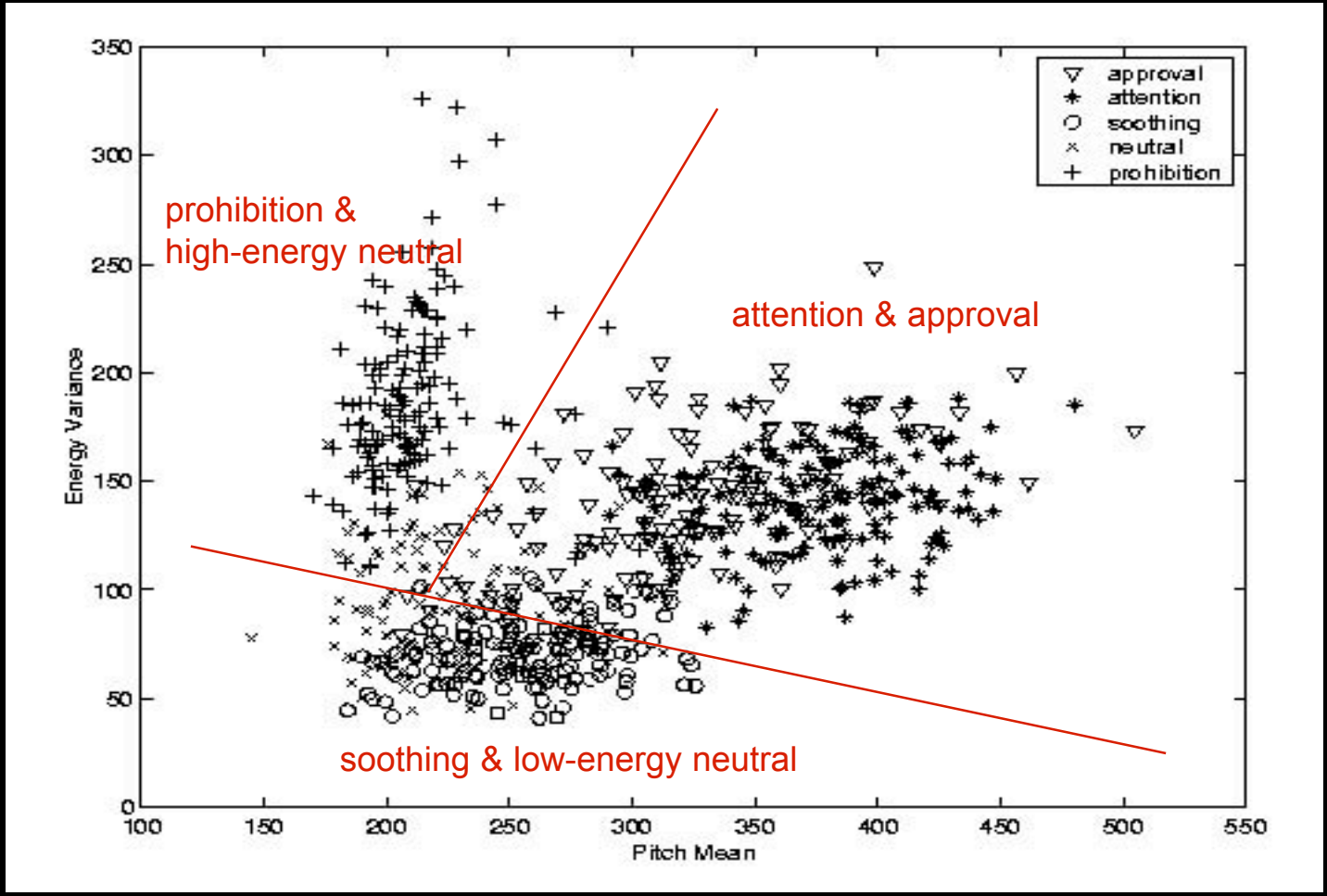


Evidence for Fernald-like Contours

in Kismet-directed speech

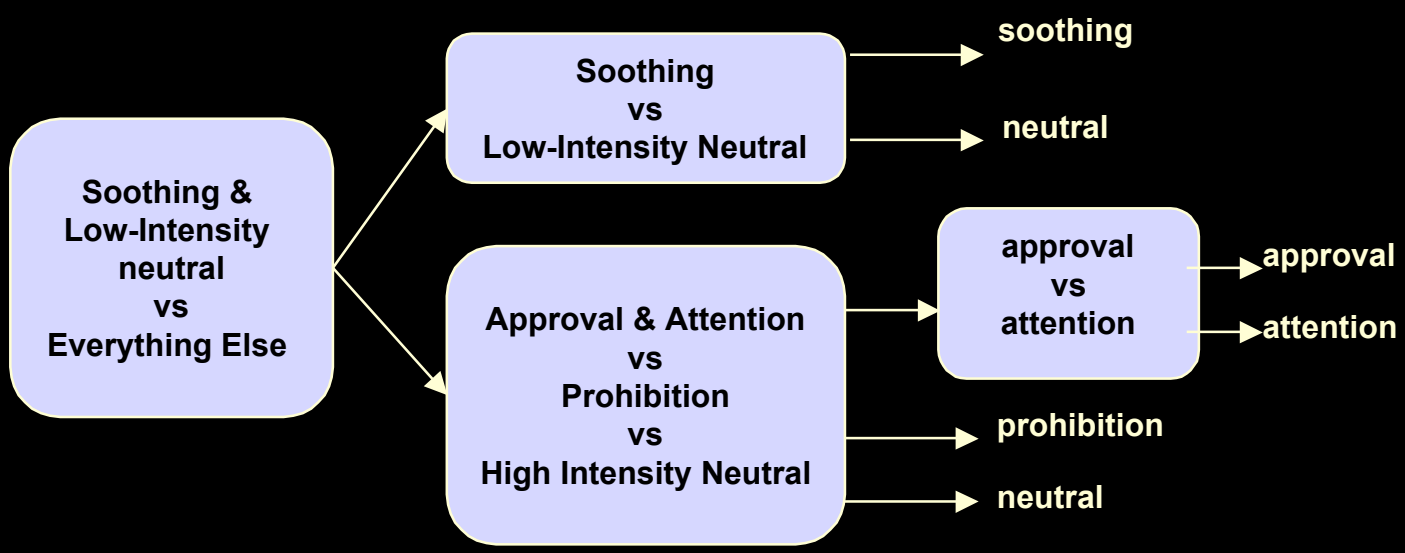


Feature Space has Nice Properties



Breazeal & Aryananda, Autonomous Robots 2001

Multi-Stage Classifier Model



- **Each stage is simple for real-time performance**
- **Later stages use more Fernald contour characteristics**
- **Off-the-shelf learning mechanism for the stages (Mixture of** Gaussian **with EM)**

Performance Evaluation of Recognizer

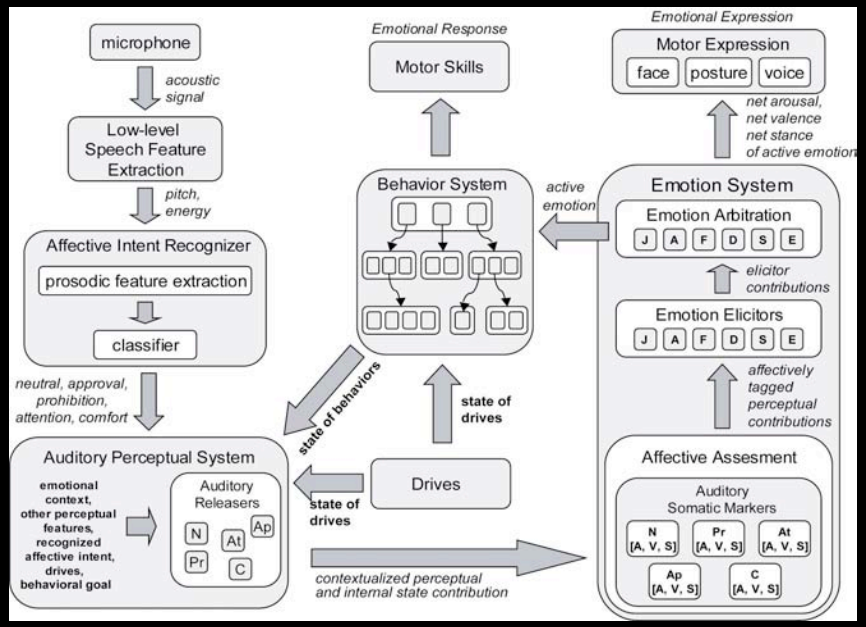
- Five classes of utterances
  - neutral speech
  - praise, prohibition, attention, soothing
- All Female speakers (n=8)
  - 7 Naïve subjects
  - 1 familiar with Kismet
- Multiple languages
  - French
  - German
  - Indonesian
  - English
  - Russian

Results, Multiple Languages

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

- 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

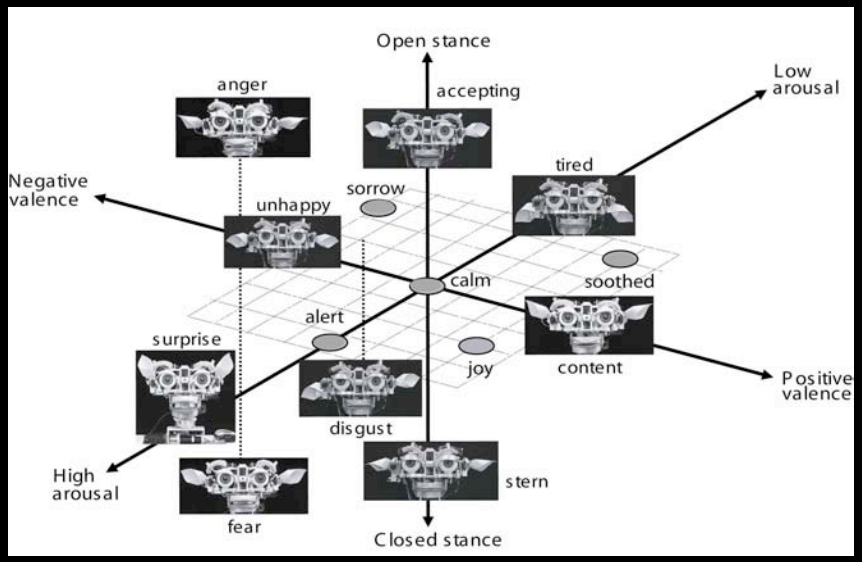
Model of Affect in Robot



Category	*Arousal*	*Valence*	*Stance*	*Typical Expression*
Approval	medium high	high positive	approach	pleased
Prohibition	low	high negative	withdraw	sad
Comfort	low	medium positive	neutral	content
Attention	high	neutral	approach	interest
Neutral	neutral	neutral	neutral	calm

- Support mental model of human
  - Model affect within robot
  - Mental model maps to computational processes
  - Intuitive mapping from tone of voice to resulting affect

Communicate through Facial Expression



- Face is window to robot's internal state
  - Transparency
  - Readable
- Signals to person
  - "I like (or not) how you're interacting with me"
  - "I'm in a corresponding affective state that you are expressing to me"
- Used by human to acknowledge robot understood (or not)

Interaction Study with Subjects

Movie of affective interaction

- All female subjects (n=5)
- 22-54 years of age
- Multiple languages
  - French, German, Indonesian, English, Russian
- Video recorded

Annotation of observable measures

Observable Measures for Communication of Affective Intent		
Cue	*Reading*	*Annotation*
Utterance	*utterance*	*"utter"*
Prosody	*pitch, energy, tempo*	Pr:
Body Posture	*neutral, erect, forward, away*	Bd:
Head Tilt	*neutral, up, down*	Hd:
Gaze Direction	*eye contact, glance/stare-down, glance/stare-up, glance/stare-right, glance/stare-left*	Gz:
Facial Expr	*neutral, relax, happy, sad alert, comforting, other*	Fc:
Ear Pose	*neutral, perk up, droop, fallen*	Er:
Lip Shape	*neutral, rounded, smile, frown*	Lp:
Acknowledge		*ack*
Sequential (across turns)		$\Longrightarrow$, $\Longleftarrow$
Sequential (within turn)		$\rightarrow$
Simultaneous		$\Longleftrightarrow$

Sample annotated interaction

Praising Intent Issued by Subject A

Trial	*Utterance*		*Human*		*Robot*
1	"Great job Kismet"	⟺	Bd:fwd, Fc:happy, Hd:up Gz:eye-ct	⟺	Er:perk-up, Fc:neu, Hd/Bd:neu Gz:eye-ct
2	"Way to go"			⟹	Gz:glance-up, Er:perk-up
3	"You're a smart robot"		Fc:happy	⟹	Gz:eye-ct, Lp:grin → Fc:relax → Fc:happy
4	*ack*				
5	"You're a cute robot"	⟺	Bd:fwd, Fc:happy, Hd:up Gz:eye-ct	⟹	Lp:curve-up Bd/Hd:neu Gz:eye-ct
	(small gasp)	⟺	Bd:erect	⟸	
6	"You're so smart"	⟺	Bd:fwd, Lp:smile Hd:up	⟹	Lp:curve-up, Gz:glance-up Hd:up
7	"What beautiful eyes"	⟺	Fc:happy Hd:up	⟺	Fc:happy, Hd:up
8	*ack*				
9	"Good job"	⟺	Bd:fwd Hd:up, Gz:eye-ct Fc:happy	⟹	Fc:neu, Hd/Bd:neu, Gz:look-right
10	"Good job"	⟹	Gz:stare-down		
11	"That was ingenious"	⟺	Bd:far-fwd, Gz:eye-ct	⟺	Bd:fwd Gz:eye-ct
12	"What are you looking at? Great"	⟺	Body:fwd Hd:up Bd:sit-back	⟺ ⟹	Head:up Er:perk-up Gz:eye-ct
13	"Who's the pretty robot"	⟺	Bd:fwd	⟹	Fc:sad, Hd:down
14	"Oh no"	⟺	Bd:sit-back, Pr:soft,low Fc:neu	⟸	
15	"You're great"	⟺	Pr:excited,high, Bd:far-fwd, Hd:up Bd:sit-back	⟹ ⟸	Hd:up, Lp:smile Er:perk-up

Prohibition Intent Issued by Subject C...continued

Trial	*Utterance*		*Human*		*Robot*
3	"Bad robot"	⟺	Bd:fwd, Hd:far-dwn	⟹	Hd:far-dwn, Gz:hold-dwn, Bd:fwd
4	*ack*				
5	"Bad job"	⟺	Fc:stern, Hd:dwn, Gz:eye-ct, Bd:neu, Bd:sit-back	⟹ ⟸	Bd:lean-back, Fc/Hd:neu, Gz:eye-ct
6	"Don't take over the humans"			⟹	Er:drop, Hd:dwn
7	*ack*				
8	"No Kismet"	⟺	Pr:loud,deep	⟹	Hd:dwn, Gz:dwn, Er:drop → Fc:sad
9	*ack*				
10	"That was so bad of you"	⟺	Bd:fwd, Fc:stern, Gz:eye-ct	⟺	Hd:bit-dwn, Gz:eye-ct, Er:bit-dwn
11	"Did you make potty on the rug"	⟺	Bd:far-fwd, Hd:dwn, Pr:deeper,louder	⟺	Gz:dwn, Hd:dwn→ Ers:fallen,
	("Oh sorry")			⟸	Face:sad
12	*ack*				
13	"Kismet"	⟺	Face:stern, Hd:dwn	⟺	Gz:eye-ct, Hd:dwn, Er:drop
14	"Bad robot"	⟺	Hd:dwn, Pr:deeper, Hd:dwn	⟺	Gz:dwn, Fc:sad, Hd:dwn
15	*ack*				
16	"Dont do that to me"	⟺	Hd:dwn	⟹	Hd:dwn
17	"Dont look at me like that"	⟺	Pr:deeper	⟹	Hd:up→ Hd:dwn
			Hd:dwn	⟸	Gz:dwn, Er:drop
	ack				

Findings

- Ready and effective use of expressive feedback
  - To acknowledge understanding
  - modulate intensity of their response
  - modulate intensity of robot's response to them
- Themed variations
- Empathic reactions
- Affective mirroring
  - Synchrony

Study 2: Regulation of vocal turn taking

Vocal Turn-Taking

- Cornerstone of human-style communication, learning, and instruction
- Four phases of turn cycle
  - Acquire floor
  - Hold floor/ speak
  - Relinquish floor
  - Listen to speaker
- Paralinguistic envelope displays regulate transitions
  - Raising brows
  - Establish eye contact
  - Break eye contact
  - Posture, gesture

Evaluation with subjects

- Naive subjects (n=5)
  - 2M, 3F
  - 25 to 28 years of age
  - All young professionals.
  - No prior experience with Kismet
  - Video recorded

Examples of turn-taking

Turn Taking

Two People

One Person

Annotation of observable measures

Annotations for Proto-dialog Experiment		
Type	*Option*	*Annotation*
Listener, Speaker	Human	H
	Robot	R
Turn Phase	Acquire Floor	Aq
	Start Speech	St
	Stop Speech	Sp
	Hold Floor	Hd
	Relinquish Floor	Rq
Cue	avert gaze	
	eye contact	
	elevate brows	
	lean forward	
	lean back	
	blink	
	"*utterance*"	
Turns	clean turn	#
	Interrupt	I
	Missed	M
	Pause	P

Annotated interaction

Envelope Displays During a Proto-Dialog...continued

Time Code	Speaker S	Ph	Cue	Listener L	Cue	Turns
07:13:05	H	Aq	eye contact	R	eye contact	11
		St	"Did you ask me how I am? I'm fine. How are you?"			
07:14:25		Sp:Rq				
07:17:09	R	Aq	avert gaze	H		12
07:17:10		St	*babble*			
07:18:03		Sp	eye contact			
07:20:05		Hd	avert gaze			
07:21:24		Rq	eye contact raise brows			
07:22:23	H	Aq		R	eye contact	13
		St	"Are you speaking another language, Kismet?"	*babble*		I
07:24:23		Sp:Rq				14
07:24:06	R	Aq:St	*babble*	H		15
07:25:04		Sp	blink			
		Rq	elev brows			
07:25:14	H	Aq:St	"Sounds like you're speaking Chinese."	R	eye contact	16
07 27:10		St:Rq				
07:27:20	R	Aq	lean forward	H		17
07:27:45		St	*babble*			
07:28:03		Sp	eye contact			
07:28:25		Rq	elev brows			
07:30:08	H	Aq:St	"Hey!"	R	avert gaze	18
07:30:15		Sp:Rq	lean forward		eye contact	
07:31:08	R	Aq:St	*babble*	H	eye contact	19
07:33:01		Sp	blink eye contact			
07:33:30		Rq	elev brows			
07:34:01	H	Aq:St	"What are you saying?"	R	eye contact	20
07:34:26		Sp:Rq				
07:36:04	R	Aq:St	*babble*	H	eye contact	21
07:37:00		Sp	blink			
07:38:19		Rq	lean forward, elev brows, eye contact		lean forward nod head	
07:40:00		Aq	lean back, avert gaze			

Envelope Displays During a Proto-Dialog...continued

Time Code	Speaker S	Ph	Cue	Listener L	Cue	Turns
07:41:13		St	*babble*			
07:42:11		Sp:Rq	eye contact			
07:45:05	H	Aq		R	eye contact	22
		St	"Did you know that you look like a gremlin?"			
07:47:05		Sp:Rq				
07:47:26	R	Aq	avert gaze	H	eye contact	23
07:49:12		St	*babble*			
07:50:25		Sp:Rq	eye contact			
07:52:22	H	Aq:St	"All right..."	R	eye contact,	24
07:53:05		Sp			eye contact	
07:54:18		St	"What are you going to do the rest of the day?"			
07:55:29		Sp:Rq				
07:56:14	R	Aq:St	*babble*	H	eye contact	25
07:57:29		Sp:Rq	blink eye contact		avert gaze	
08:03:01	H	Aq:St	"My name is Carol. you have to remember that I'm Carol. "	R	eye contact *babble*	26 I
08:05:25		Sp:Rq	(pause)			P
08:06:31		St	"If you see me again, I'm Carol."		eye contact	27
08:07:17		Sp:Rq	(pause)			P
08:08:26		St	"Hello!"			28
08:09:21		Sp			blink	
		Rq	lean forward			
08:10:13	R	Aq	avert gaze	H	lean back (laugh)	29
08:10:40		St	*babble*			
08:11:17		Sp	eye contact, blink			
08:11:45		Rq	lean forward			
08:12:19	H	Aq:St	"Hello!"	R		30
08:12:54		Sp:Rq				
08:13:23	R	Aq:St	*babble*	H		31
08:14:25		St:Rq	eye contact, elev brows			
08:15:05	H	Aq:St	"Hello!"	R		32
08:15:35		St:Rq				

Turn taking performance

- Turn-taking performance
  - 82.5% “clean” turn transitions
  - 10.9% interruptions
  - 6.3% delays followed by prompting
- Significant flow disturbances
  - Tend to occur in clusters
  - 6% of the time, but rate diminishes over time

	Sub 1		Sub 2		Sub 3		Sub 4		Avg
	Data	%	Data	%	Data	%	Data	%	%
Clean Turns	35	83	45	85	38	84	83	78	82.5
Interrupts	4	10	4	7.5	5	11	16	15	10.9
Pauses	3	7	4	7.5	2	4	7	7	6.3
Significant Flow Distrb.	3	7	3	5.7	2	4	7	7	6
Total Speaking Turns	42		53		45		106		

Evidence of entrainment

		Time Stamp (min:sec)	*Clean Turns 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 6:47	6:47–6:54	3
		6:55–7:21	7
		7:22–7:57	11
	end 8:44	8:03–8:44	16
subject 4	start 4:52	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	9:20–10:40	80+

- Evidence for entrainment
  - Shorter phrases
  - Wait longer for response
  - Read turn-taking cues
  - 0.5—1.5 seconds between turns

Findings

- Ready use of envelope displays to regulate interaction
  - Benefits interaction
- Captured dynamics of interaction
  - It's a Dance!
  - Tempo & synchrony
  - Entrainment

Kismet: Summary

- Socially engaging on many levels
  - Readable social cues
  - Responsive to social cues
  - Fine grained dynamics & synchrony
- Strong social presence
- Socially pro-active
- Mutually beneficial interactions
- Computational models supports aspects of attributed social model
  - Ethological models of emotions, drives, attention, behavior, etc.

Study 3: Social Presence

Robot versus Animation

Social presence: A comparison

(Cory Kidd, MAS MS student)

- Social presence: how closely a mediated experience is to an actual, “live” experience
- Naïve subjects interact with
  - A robot
  - An animated character
  - A human
- Simple visual task

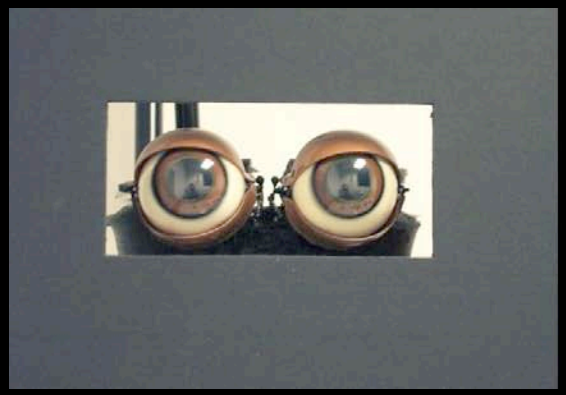
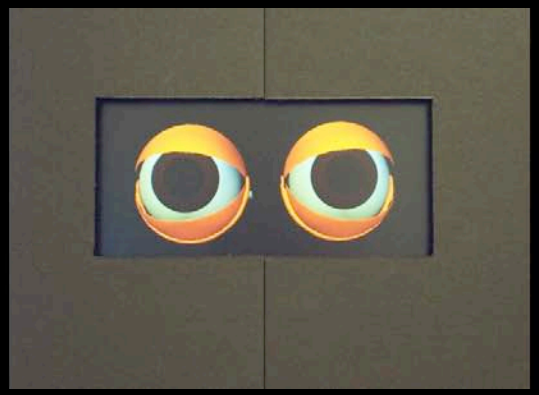
Measures

- social presence measures
  - Questionnaire
  - Video analysis (3 cameras)
    - Reaction time
    - Proximity, personal space
- Arousal measures
  - Galvanic skin response

The Questionnaire

- Robot as a media
- Based on Lombard & Ditton scale for social presence (7 point scale)
  - Social richness
  - Realism
  - Shared space
  - Immersion (psychological & perceptual)
  - Social actor within medium
  - Medium as a social actor
- Set list of adjectives (7 point scale)
- Set of open ended questions

The Protocol


- (n=32) naïve subjects
  - 18-47 years (M=27, SD=9)
  - 50% M, 50% F
- Only see eyes to minimize appearance effects
- Wizard of Oz
  - Pre-recorded female human voice, same for all characters
  - Preset order of interaction with each character (all 6 used)
  - Each character has own fixed ordering of its requests
  - Fixed timing of interactions

The interaction

Commands spoken while looking at a particular block:

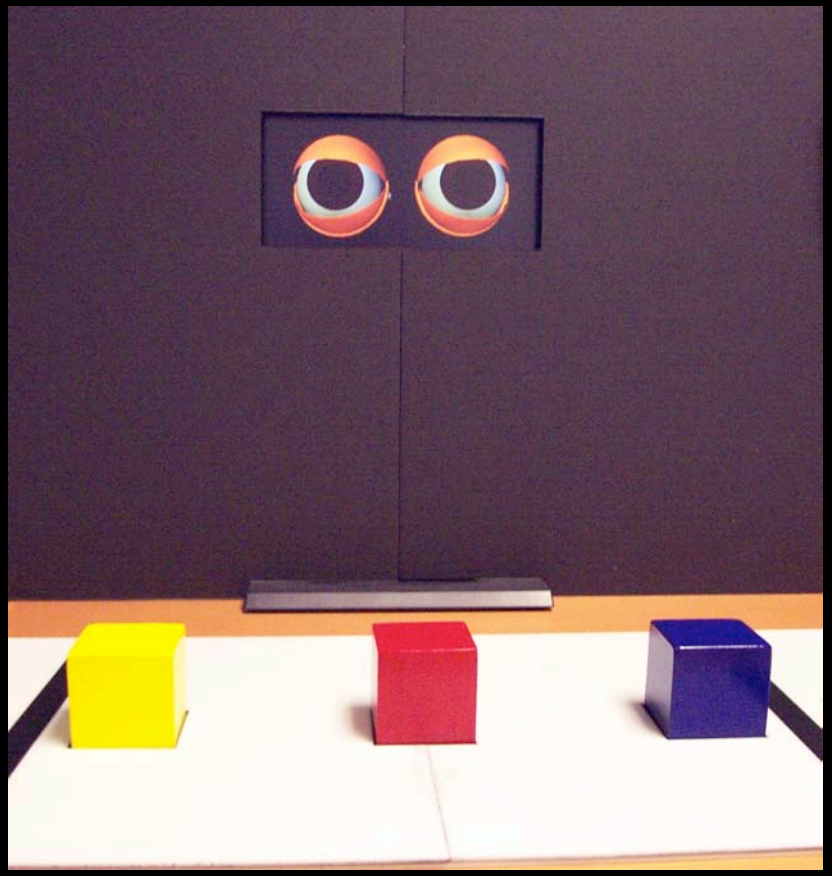
- **Move this block towards me.**
- **Move that block off the table.**
- **Hold that block up so I can see it.**

Commands spoken while looking at a point on the table:

- **Move the blue block there.**
- **Put the yellow block here.**

Commands spoken while looking at the subject:

- **Move the red block towards me.**
- **Put the blue block where I can't see it.**
- **Please move the yellow block to my left.**
- **Put the yellow block where I can't see it.**



Level of engagement

Question	F	P-value	Human	Robot	Screen
1. How often did you feel that the character was really alive and interacting with you? (higher response = more often)	df(2,93) = 20.33	<0.0001	5.88	3.97	3.31
2. How completely were your senses engaged? (higher response = very much)	df(2,93) = 10.64	0.0001	5.59	4.75	3.97
3. To what extent did you experience a sensation of reality? (higher response = very much)	df(2,93) = 9.83	0.0001	5.69	4.41	3.97
4. How well were you able to view the character from different angles? (higher response = very well)	df(2,92) = 8.03	0.0006	5.74	5.69	4.22
5. How engaging was the interaction? (higher response = very much)	df(2,93) = 6.99	0.0015	5.53	4.72	4.09
6. The experience caused real feelings and emotions for me. (higher response = strongly agree)	df(2,93) = 5.26	0.0068	5.16	4.16	3.63
7. How much attention did you pay to the display devices/equipment rather than to the interaction? (higher response = very much)	df(2,93) = 2.66	0.0754	3.97	4.97	4.47
8. How relaxing or exciting was the experience? (higher response = very exciting)	df(2,93) = 2.60	0.0800	4.59	4.44	3.78

Subject reaction to character

Question	F	P-value	human	robot	screen
1. How often did you have the sensation that the character could also see/hear you? (higher response = more often)	df(2,93) = 19.07	0.00001	5.94	3.91	3.19
2. How often did you want to or did you make eye contact with the character? (higher response = more often)	df(2,93) = 6.00	0.0035	4.97	6.25	5.78
3. How much control over the interaction with the character did you feel that you had? (higher response = more control)	df(2,93) = 5.23	0.0070	3.81	2.91	2.31
4. How often did you make a sound out loud in response to someone you saw or heard in the interaction? (higher response = more often)	df(2,93) = 5.47	0.0083	2.03	1.41	1.25

Involvement with characters

Question	F	P-value	human	robot	screen
1. He/she is a lot like me.	df(2,93) = 9.28	0.0002	4.59	3.09	2.69
2. If he/she were feeling bad, I'd try to cheer him/her up.	df(2,93) = 4.09	0.0199	5.44	4.91	4.09
3. He/she seemed to look at me often.	df(2,93) = 4.05	0.0207	5.97	5.44	4.78
4. I'd like to see/hear him/her again.	df(2,93) = 3.74	0.0273	4.13	5.41	4.56
5. If there were a story about him/her in a newspaper or magazine, I would read it.	df(2,90) = 3.38	0.0383	4.87	5.81	4.55
6. I would like to talk with him/her.	df(2,93) = 3.22	0.0444	4.97	5.00	3.97

Choice of adjectives

Adjective	P-value	Human	Robot	Screen
Convincing	0.0019	5.16	4.25	3.56
Varied	0.0196	4.13	3.45	2.90
Compelling	0.0307	4.97	4.56	3.84
Entertaining	0.0414	4.19	5.41	4.72
Enjoyable	0.0496	4.16	5.28	4.59
Credible	0.0820	4.97	4.38	3.94

- People rated the robot
  - More convincing
  - More compelling
  - More entertaining
    - … than the animated character

Summary

- People found the robot to be
  - Easier to read
  - More engaging of senses and emotions
  - More interested in them
    - …than the animated character.
- People often found the robot to be more like the human than the animated character