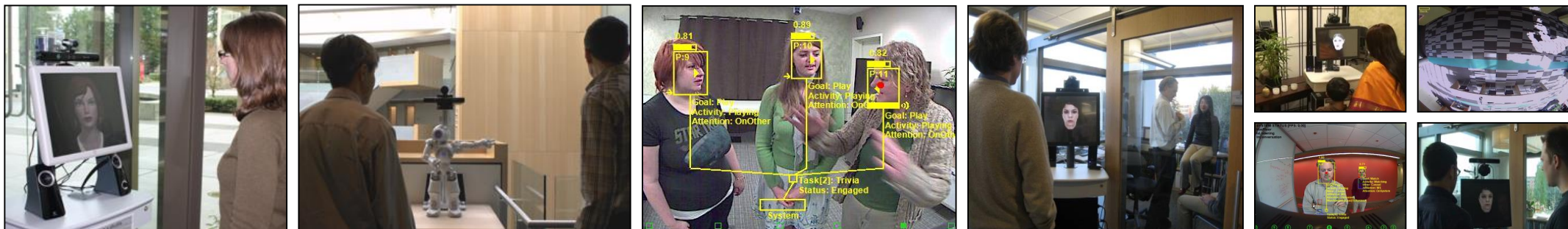


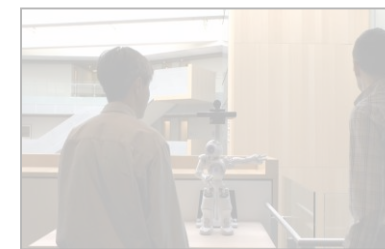
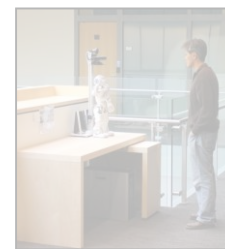
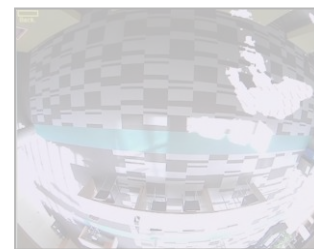
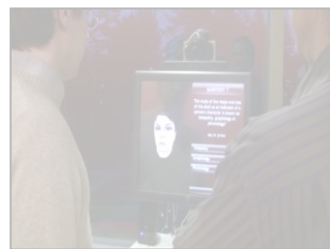
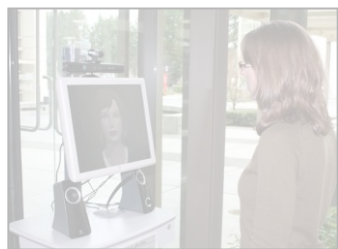
Timing and Coordination in Physically Situated Language Interaction



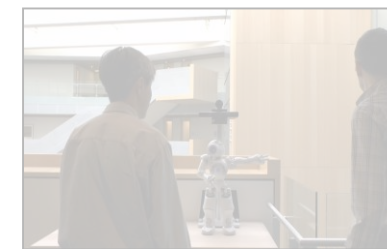
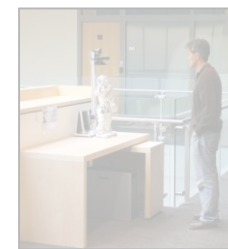
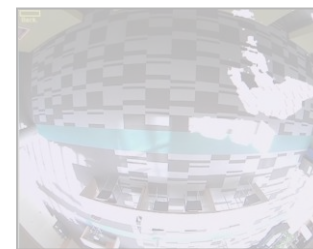
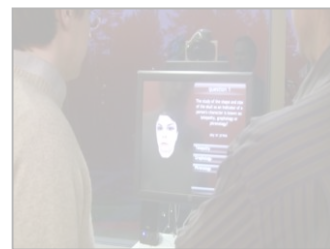
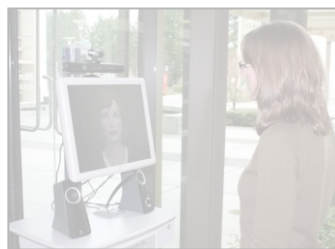
Dan Bohus
Microsoft Research

collaborators: Sean Andrist, Qin Cai, Rich Caruana, Pali Choudhuri, Ashley Feniello, Don Gillett, Richard Hughes, Ece Kamar, Mihai Jalobeanu, Zicheng Liu, Tomislav Pejsa, Nick Saw, Richard Roberts, Stephanie Rosenthal, Anne Loomis Thompson, Oriol Vinyals, Cha Zhang, Zhou Yu

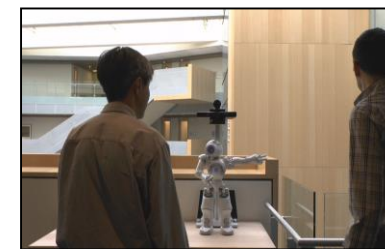
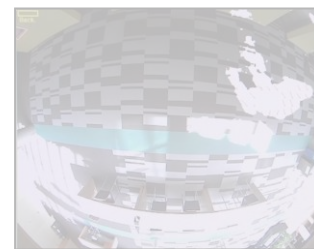
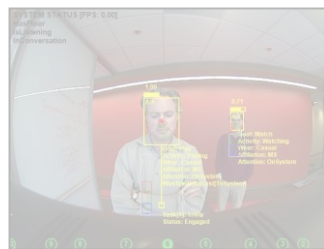
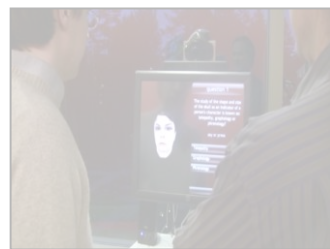
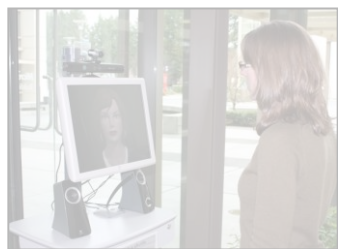
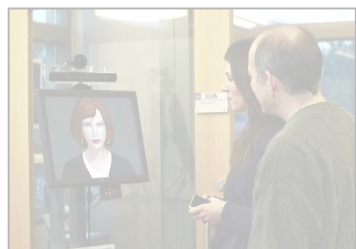
Physically situated language interaction



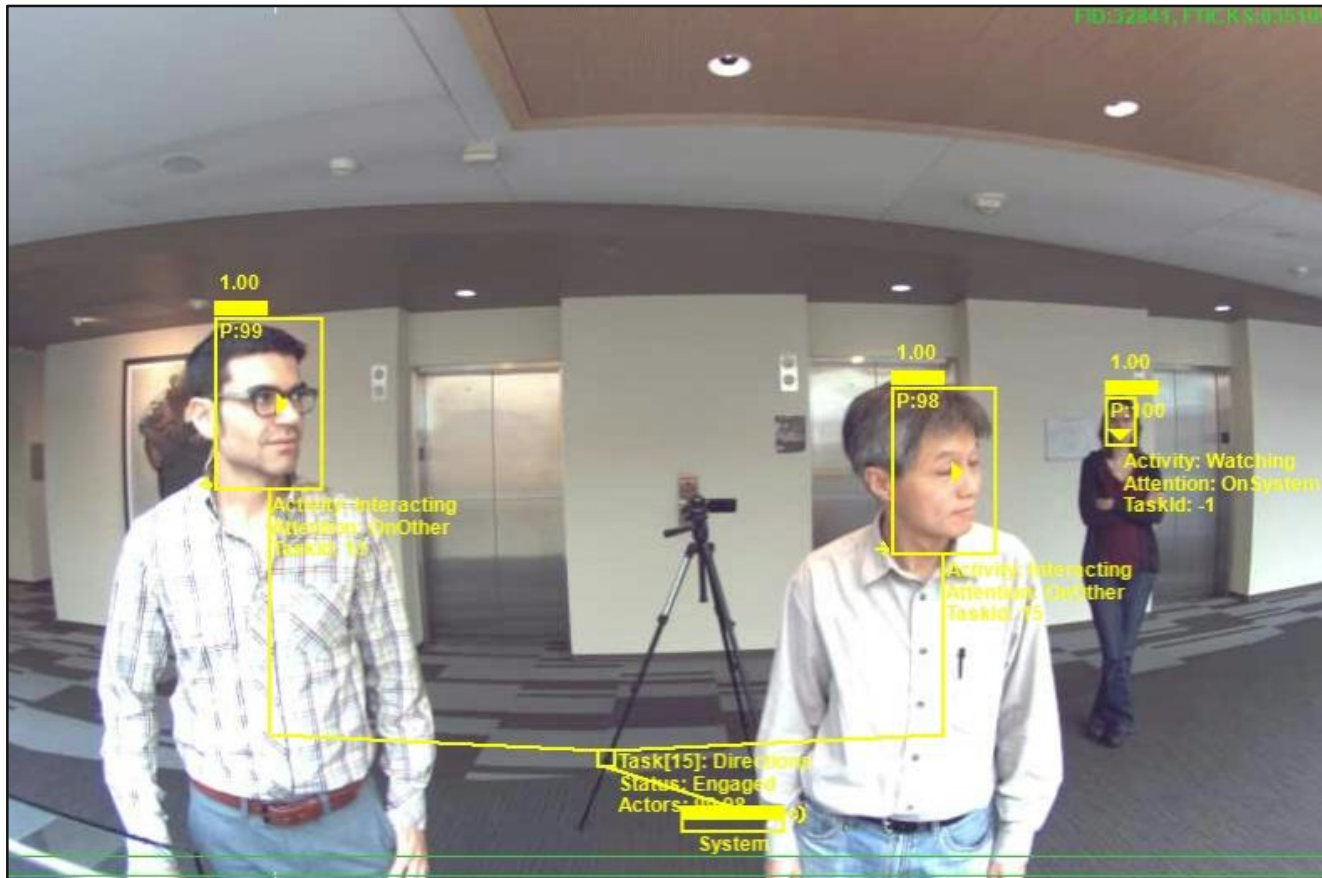
Physically situated language interaction



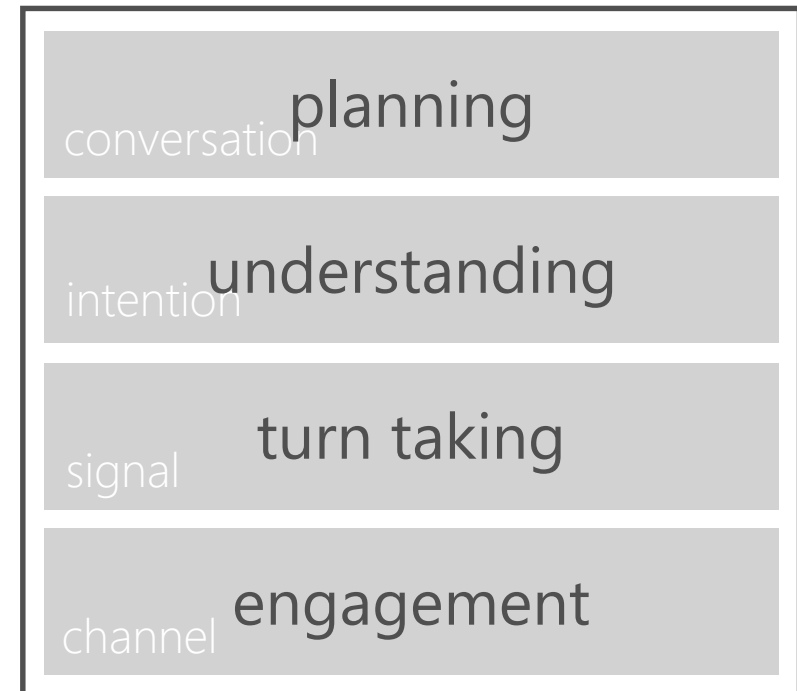
Physically situated language interaction



Physically situated language interaction



interaction skills



Physically situated language interaction



situational context

why: goals and intentions

sense and reason about beliefs, intentions, goals and long-term plans

what: situation and activity

sense and reason about relevant events and activities of self and others

who: physical awareness

identify, track, and characterize relevant actors, objects, states and relationships

interaction skills

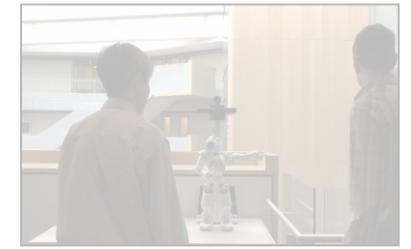
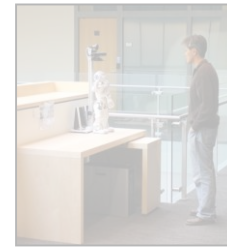
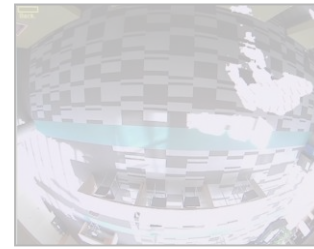
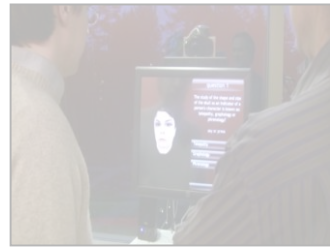
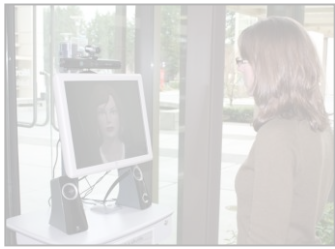
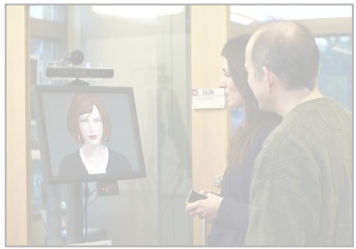
conversation **planning**

intention **understanding**

signal **turn taking**

channel **engagement**

Physically situated language interaction



situational context

interaction skills

why: goals and intentions

sense and reason about beliefs, intentions, goals and long-term plans

what: situation and activity

sense and reason about relevant events and activities of self and others

who: physical awareness

identify, track, and characterize relevant actors, objects, states and relationships

planning

conversation

understanding

intention

turn taking

signal

engagement

channel

Engagement

Visualization

SYSTEM STATUS [FPS: 0.00]
HasFloor
IsListening
InConversation

0.88
A:1
Dan Bohus
0.87

Goal: Shuttle
Activity: Interacting
Wear: Casual
Affiliation: MS
Attention: OnOther
Task[2]: NotifyWait
Status: Engaged

Goal: Shuttle
Activity: Interacting
Wear: Casual
Affiliation: MS
Attention: OnOther
Was SpeakingLast[ToSystem]
Task[1]: Receptionist
Status: Suspended

Actors Interaction Tasks Reactive state Display configuration

Actor ID: 1 [face id 1]

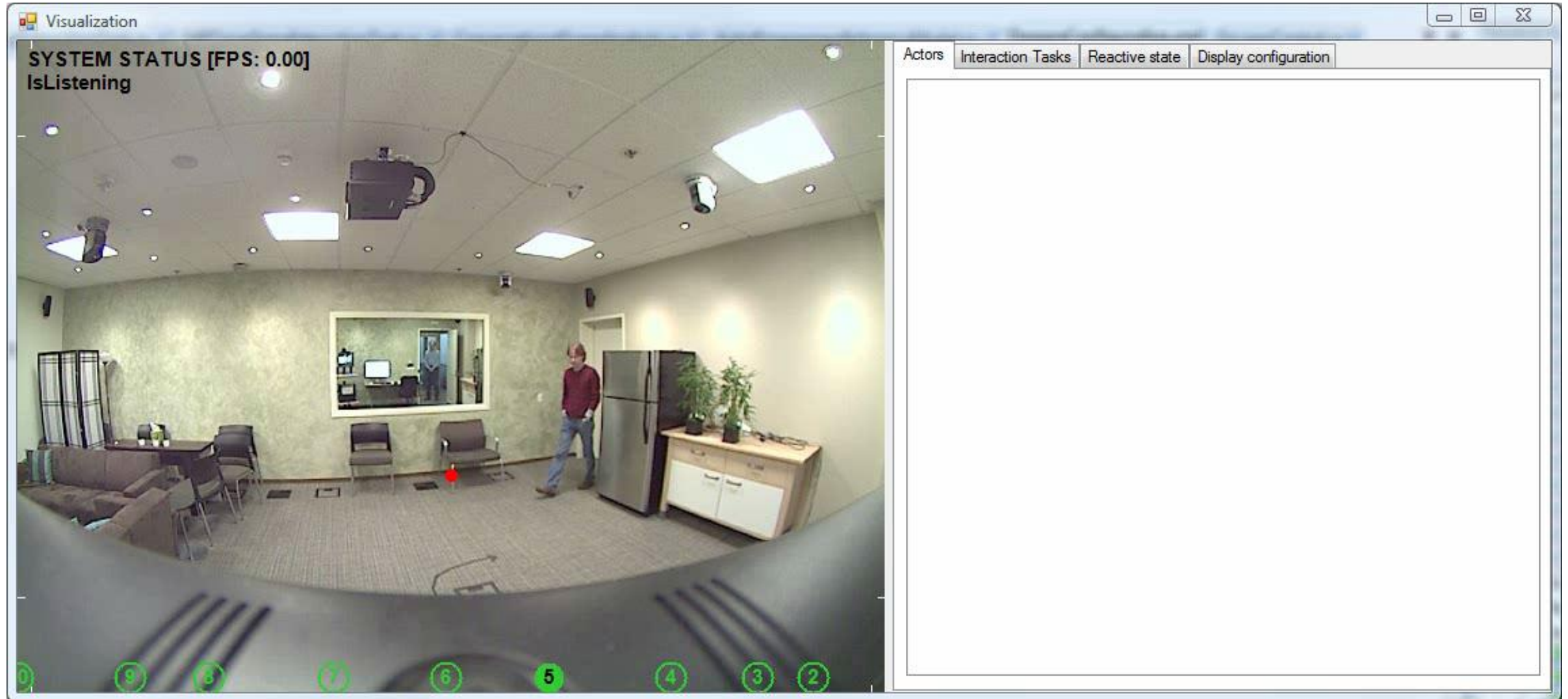
Engagement state	Floor info & actions	Clothing
Engaged <input type="checkbox"/>	HasFloor False	Formal <input type="checkbox"/>
NotEngaged <input checked="" type="checkbox"/>	None <input checked="" type="checkbox"/>	Casual <input checked="" type="checkbox"/>
Engagement action	Hold <input type="checkbox"/>	Affiliation
NoAction <input checked="" type="checkbox"/>	Take <input type="checkbox"/>	MS <input checked="" type="checkbox"/>
Engaging <input type="checkbox"/>	ReleaseToS <input type="checkbox"/>	External <input type="checkbox"/>
Maintaining <input type="checkbox"/>	ReleaseToO <input type="checkbox"/>	Goal
Disengaging <input type="checkbox"/>	Speech <input type="checkbox"/>	Shuttle <input checked="" type="checkbox"/>
Engagement goal	IsSpeaking False	Register <input type="checkbox"/>
Engaged <input type="checkbox"/>	SpeakingLast True	Other <input type="checkbox"/>
NotEngaged <input checked="" type="checkbox"/>	Speech target	Activity
Attention state	System <input checked="" type="checkbox"/>	Interacting <input checked="" type="checkbox"/>
System <input type="checkbox"/>	Other <input type="checkbox"/>	Waiting(Recep) <input type="checkbox"/>
Other <input checked="" type="checkbox"/>		Waiting(Shuttle) <input type="checkbox"/>
		PassingBy <input type="checkbox"/>

Actor ID: 2 [face id 2]

Engagement state	Floor info & actions	Clothing
Engaged <input checked="" type="checkbox"/>	HasFloor False	Formal <input type="checkbox"/>
NotEngaged <input type="checkbox"/>	None <input checked="" type="checkbox"/>	Casual <input checked="" type="checkbox"/>
Engagement action	Hold <input type="checkbox"/>	Affiliation
NoAction <input type="checkbox"/>	Take <input type="checkbox"/>	MS <input checked="" type="checkbox"/>
Engaging <input type="checkbox"/>	ReleaseToS <input type="checkbox"/>	External <input type="checkbox"/>
Maintaining <input checked="" type="checkbox"/>	ReleaseToO <input type="checkbox"/>	Goal
Disengaging <input type="checkbox"/>	Speech <input type="checkbox"/>	Shuttle <input checked="" type="checkbox"/>
Engagement goal	IsSpeaking False	Register <input type="checkbox"/>
Engaged <input checked="" type="checkbox"/>	SpeakingLast False	Other <input type="checkbox"/>
NotEngaged <input type="checkbox"/>	Speech target	Activity
Attention state	System <input checked="" type="checkbox"/>	Interacting <input checked="" type="checkbox"/>
System <input type="checkbox"/>	Other <input type="checkbox"/>	Waiting(Recep) <input type="checkbox"/>
Other <input checked="" type="checkbox"/>		Waiting(Shuttle) <input type="checkbox"/>
		PassingBy <input type="checkbox"/>

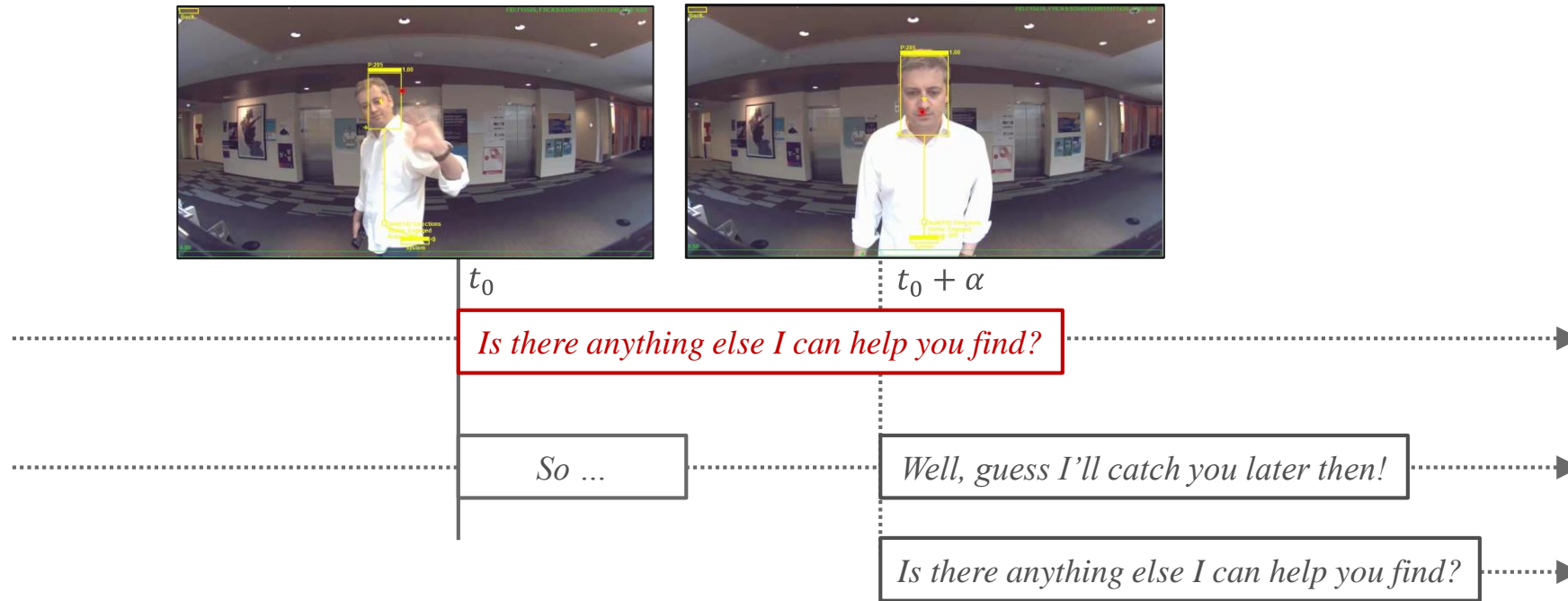
0 9 8 7 6 5 4 3 2

Engagement



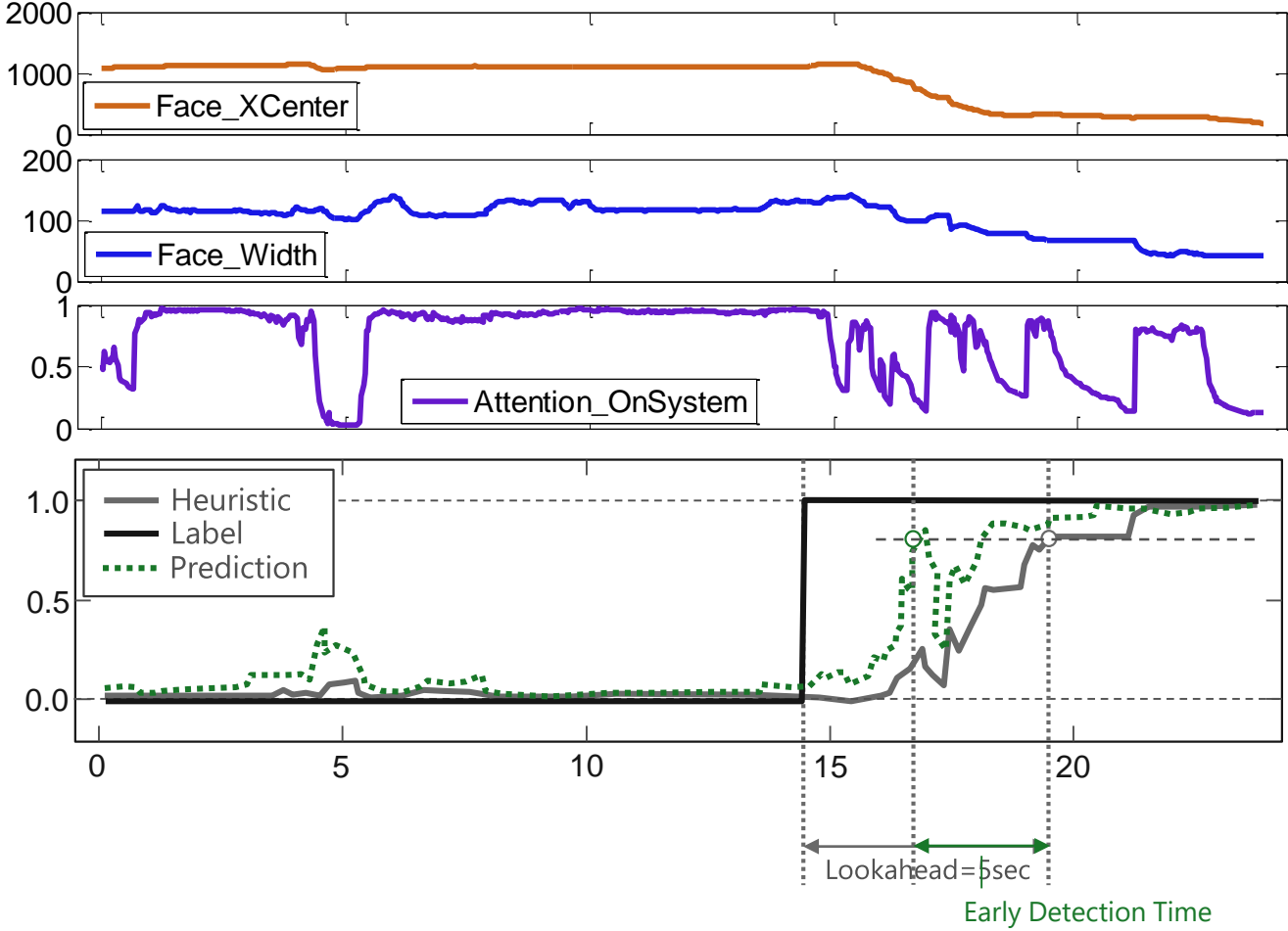
Engagement models: in practice

Forecasting disengagement

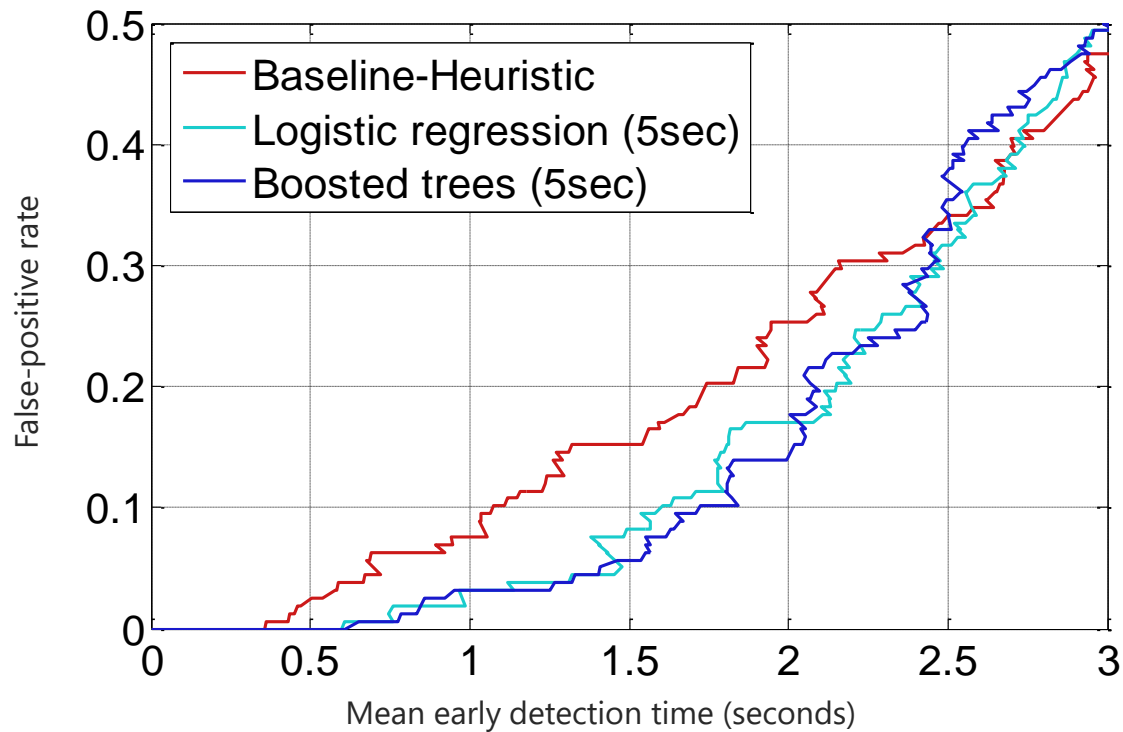


1. Forecast disengagement in a future time window
2. Use hesitations to mitigate high uncertainty

Forecasting disengagement: approach

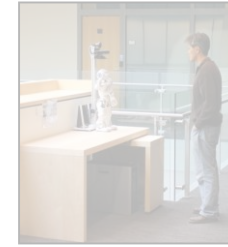
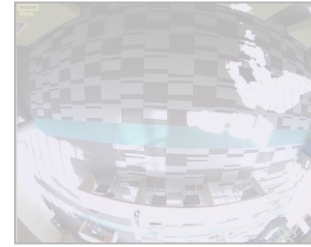
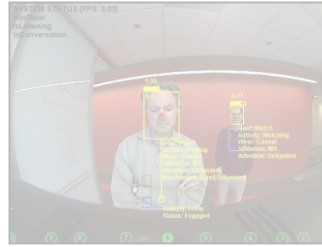
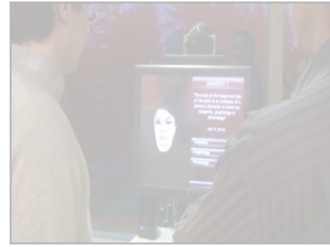
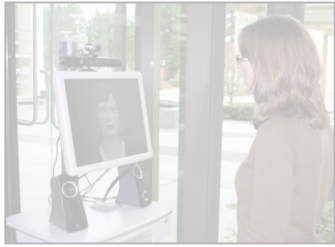
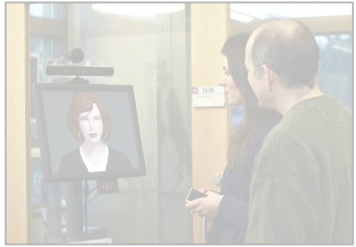


Forecasting disengagement: results



Model	Classification error	Mean squared error
Majority Baseline	24.9%	0.1880
Heuristic based model (H)	13.9%	0.1134
FocusOfAttention (A)	17.0%	0.1249
FaceLocation (L)	16.3%	0.1265
TrackingConfidence (C)	18.8%	0.1468
Interaction/Dialog (D)	21.3%	0.1526
A+L	13.5%	0.1057
A+L+C	11.7%	0.0955
A+L+C+D	11.5%	0.0924
Full model (A+L+C+H+H)	10.9%	0.0863

Physically situated language interaction



situational context

interaction skills

why: goals and intentions

sense and reason about beliefs, intentions, goals and long-term plans

what: situation and activity

sense and reason about relevant events and activities of self and others

who: physical awareness

identify, track, and characterize relevant actors, objects, states and relationships

planning

conversation

understanding

intention

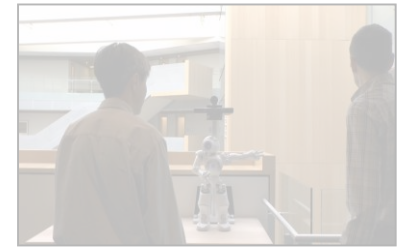
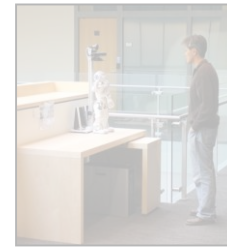
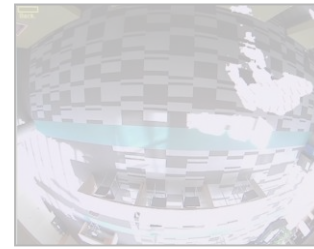
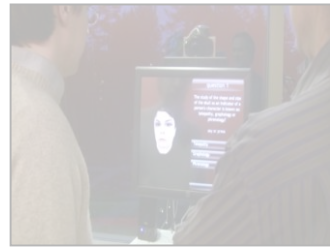
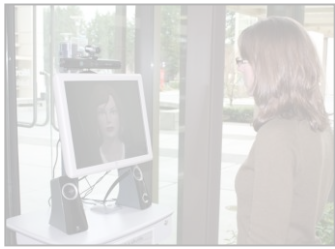
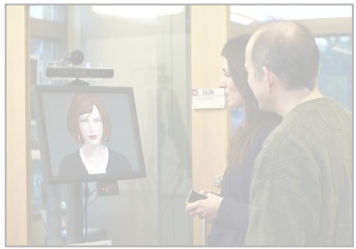
turn taking

signal

engagement

channel

Physically situated language interaction



situational context

interaction skills

why: goals and intentions

sense and reason about beliefs, intentions, goals and long-term plans

what: situation and activity

sense and reason about relevant events and activities of self and others

who: physical awareness

identify, track, and characterize relevant actors, objects, states and relationships

planning

conversation

understanding

intention

turn taking

signal

engagement

channel

Coordination of attention and speech production

Charles Goodwin: disfluencies and attention

from *Conversational Organization: Interaction Between Speakers and Hearers*

Speaker: Anyway, (0.2) Uh:, (0.2) We went t- I went ta bed

Listener: 

Speaker: Brian you're gonna hav- You kids'll *have* to go

Listener: 

Speaker: I come int- I no sooner sit down on the couch

Listener: 

Our system ...

Model

Attentional demands

Defined at phrase level
Specified at onset and production
Define expected targets



Attentional supply

Infers attention on various targets
Relies on ML model for geometric visual attention
Leverage features from visual subsystems



Coordinative policy



... .. *Excuse me!* *To get* *To get to 3800* *go to [...]*

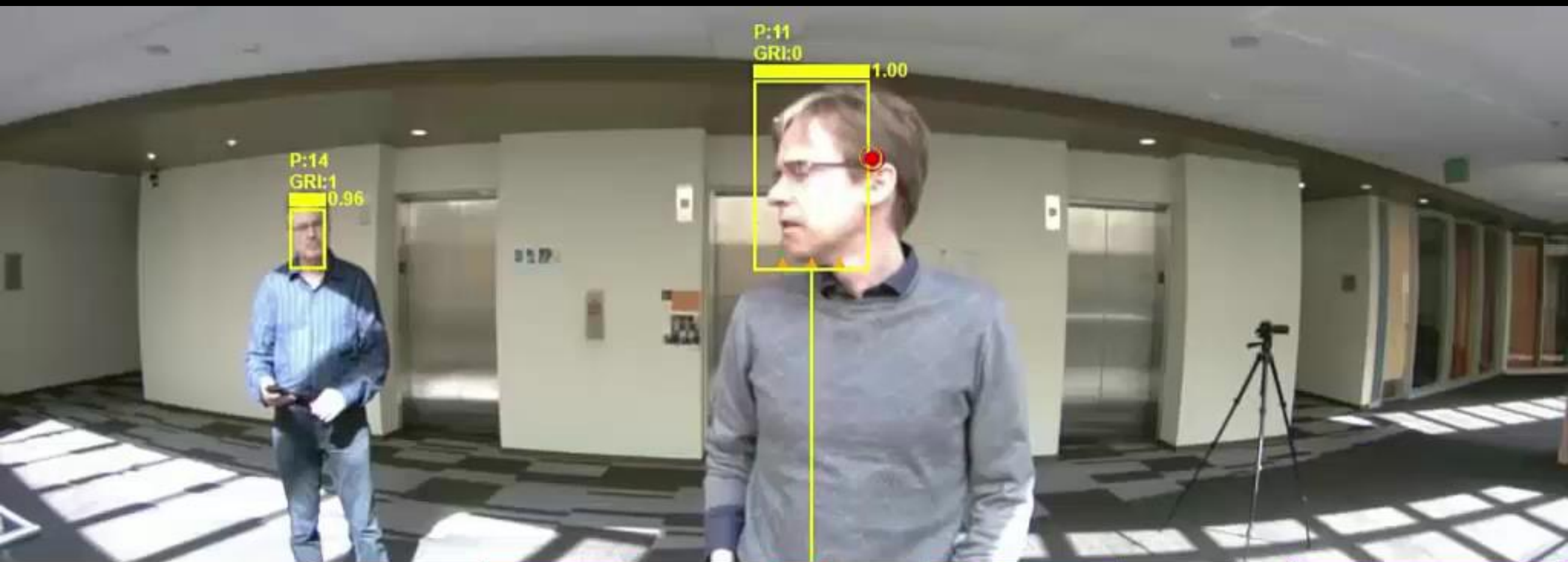
Phrase 1

Phrase 2

a demonstration video ...



the robot's view ...
sensing and computation details

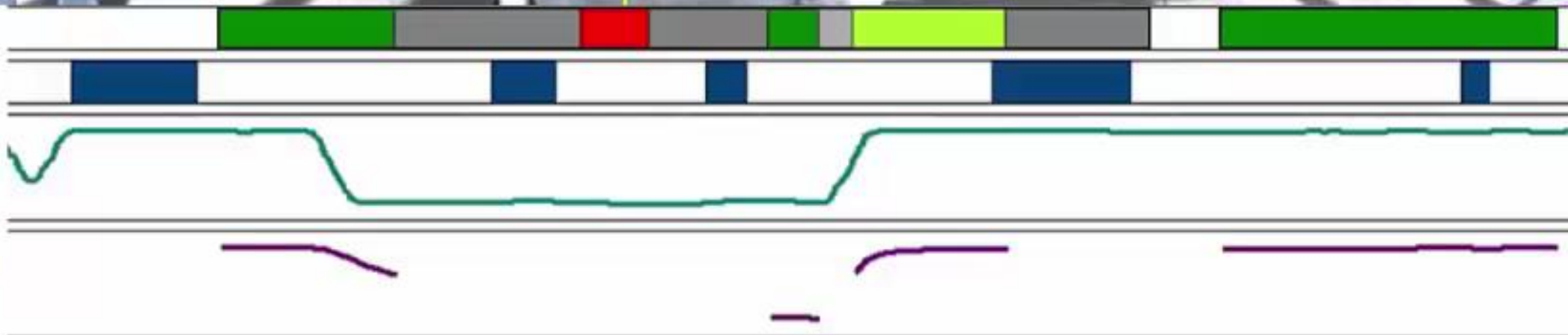


Robot

Participant

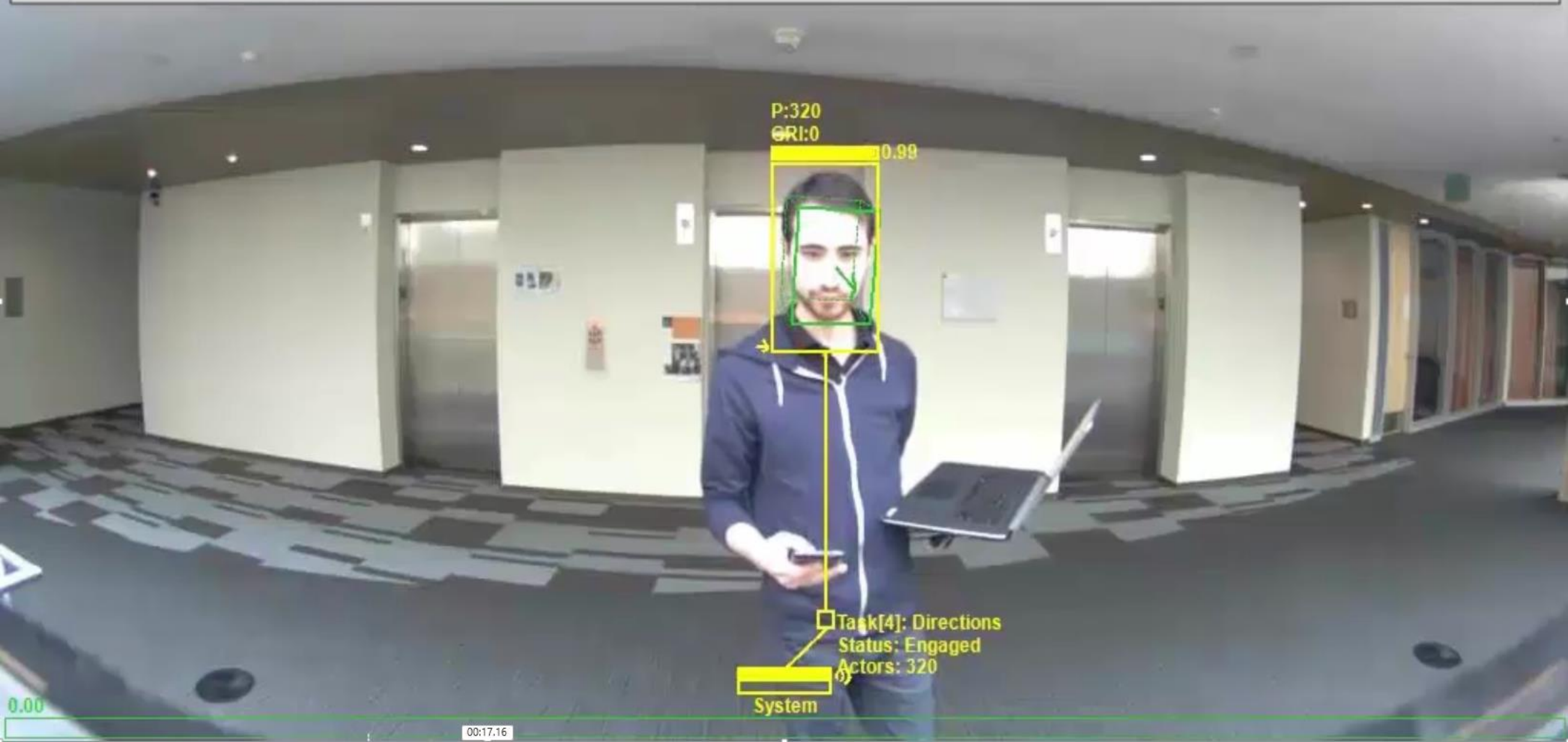
Participant's
attention

Phrase production
attention supply

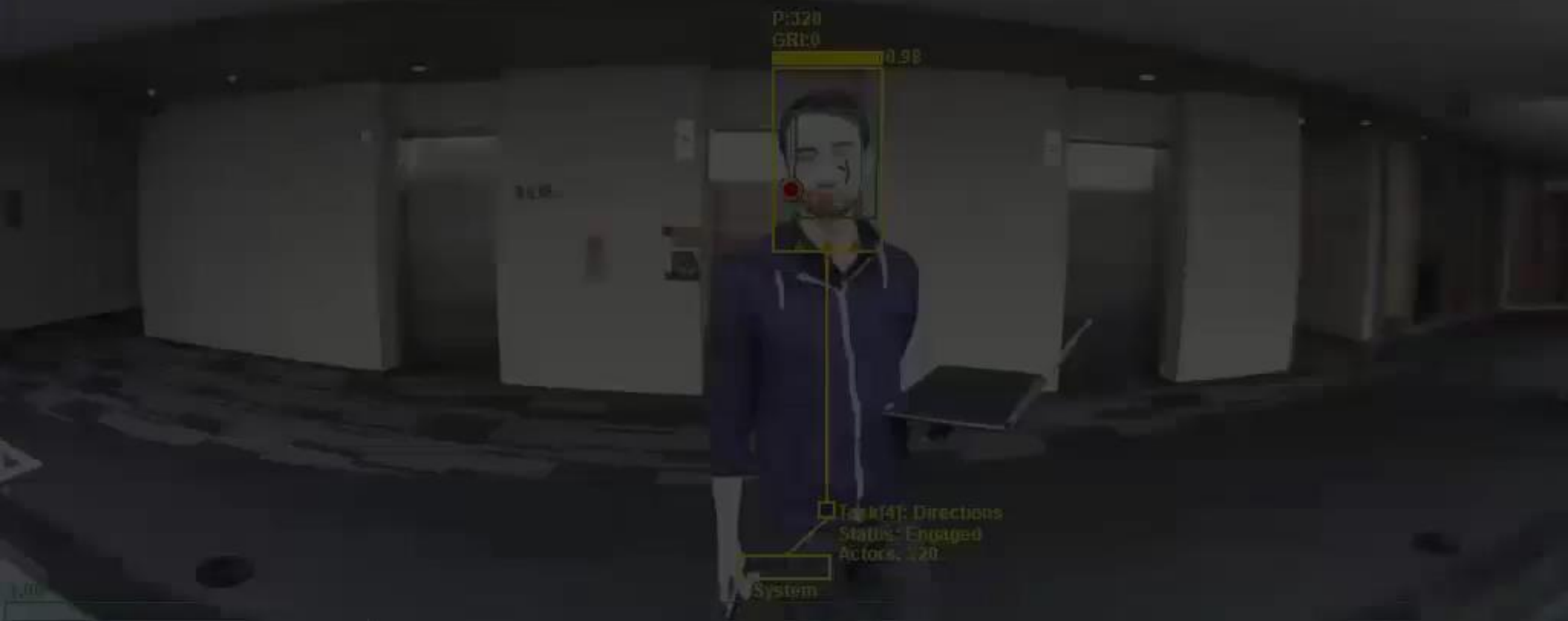


sample natural interactions ...

Dialog act: To get to 2800 | take the elevator down to the 2nd floor | turn left as you walk out of the elevator and continue on to the end of that hallway | ... | Excuse me | ... | 2800 will be on that side of the building.



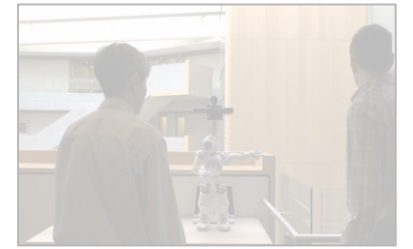
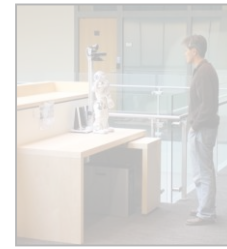
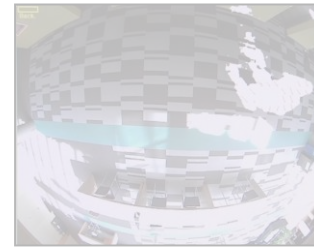
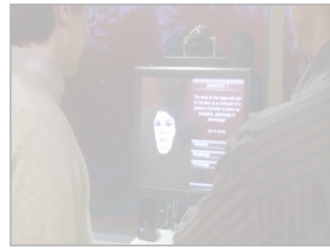
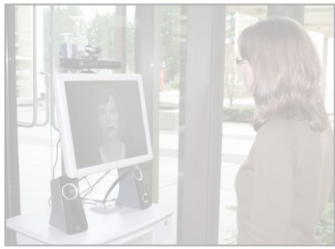
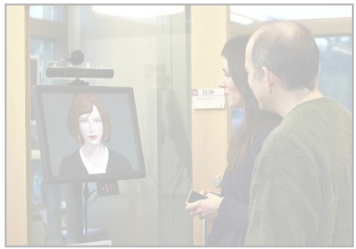
Dialog act:



... but still, far from solved

another interesting failure mode ...

Physically situated language interaction



situational context

interaction skills

why: goals and intentions

sense and reason about beliefs, intentions, goals and long-term plans

what: situation and activity

sense and reason about relevant events and activities of self and others

who: physical awareness

identify, track, and characterize relevant actors, objects, states and relationships

planning

conversation

understanding

intention

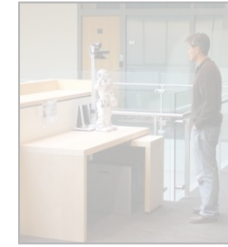
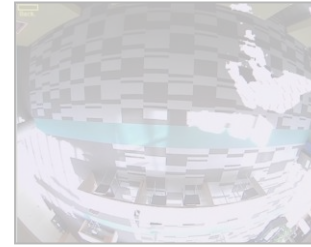
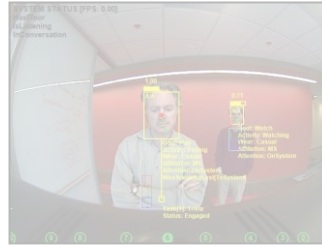
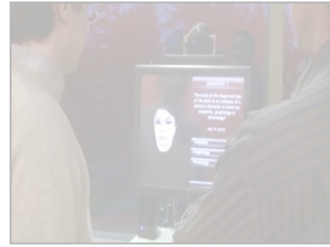
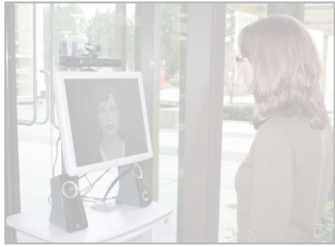
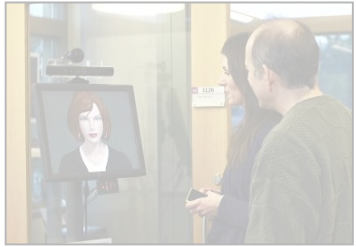
turn taking

signal

engagement

channel

Physically situated language interaction



situational context

interaction skills

why: goals and intentions

sense and reason about beliefs, intentions, goals and long-term plans

what: situation and activity

sense and reason about relevant events and activities of self and others

who: physical awareness

identify, track, and characterize relevant actors, objects, states and relationships

planning

conversation

understanding

intention

turn taking

signal

engagement

channel