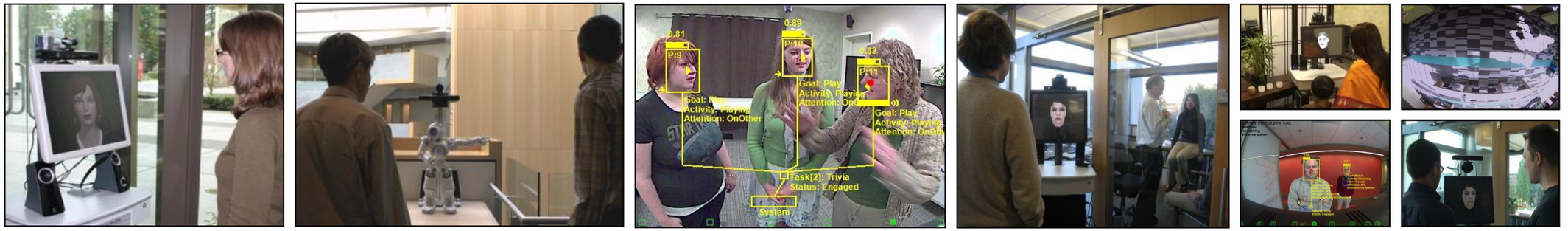
Microsoft

Physically Situated Language Interaction: an integrative-AI challenge

Dan Bohus

Microsoft Research
Faculty Summit
2015

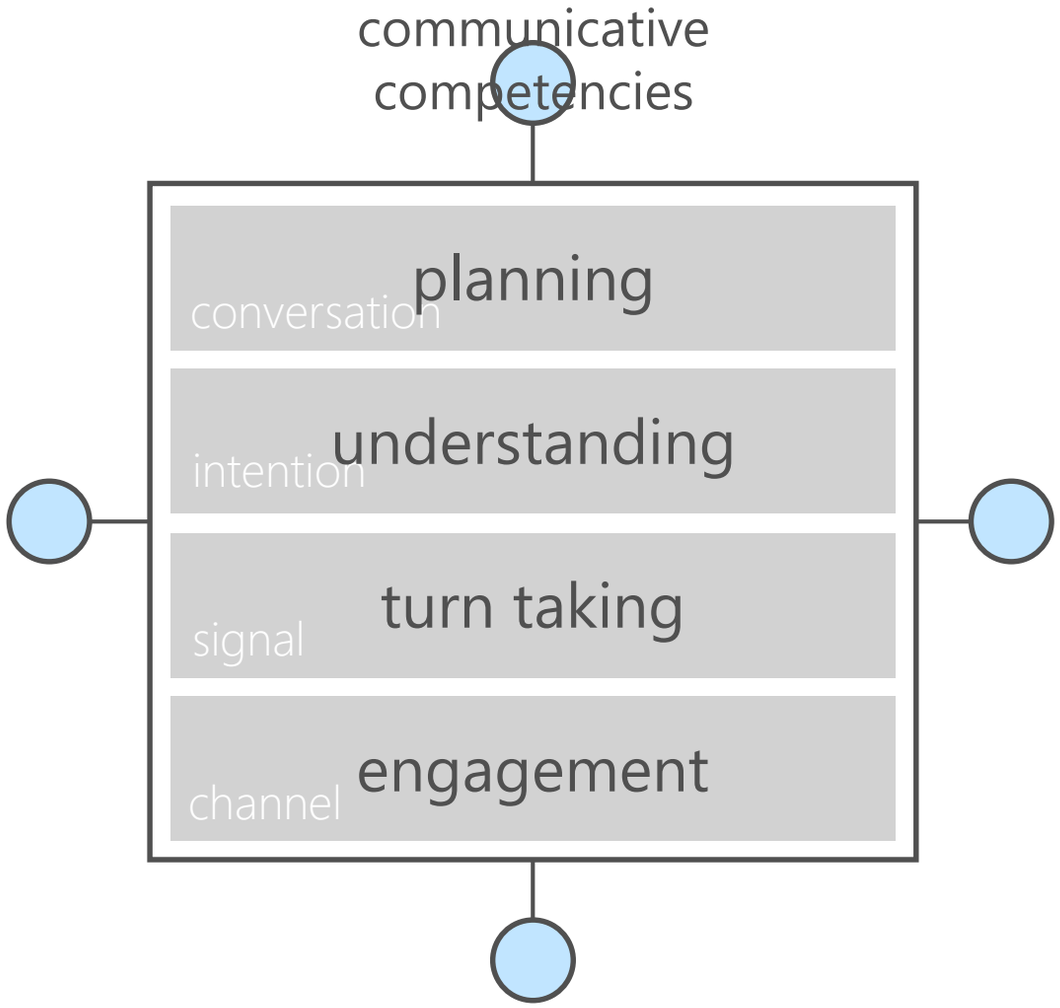
Physically Situated Language Interaction: an integrative-AI challenge



Dan Bohus

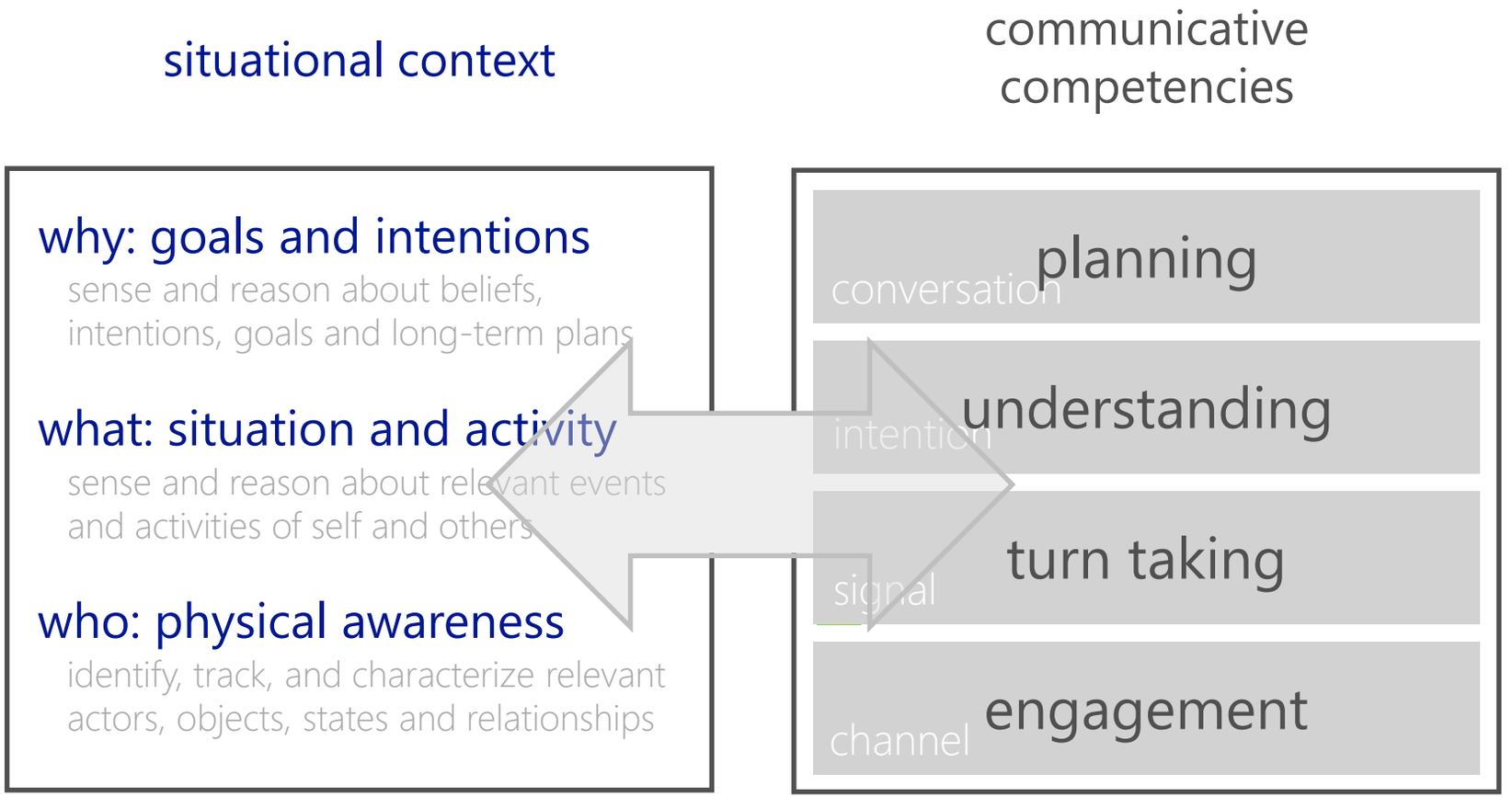
dbohus@microsoft.com
Senior Researcher
Adaptive Systems and Interaction
Microsoft Research

Physically situated language interaction

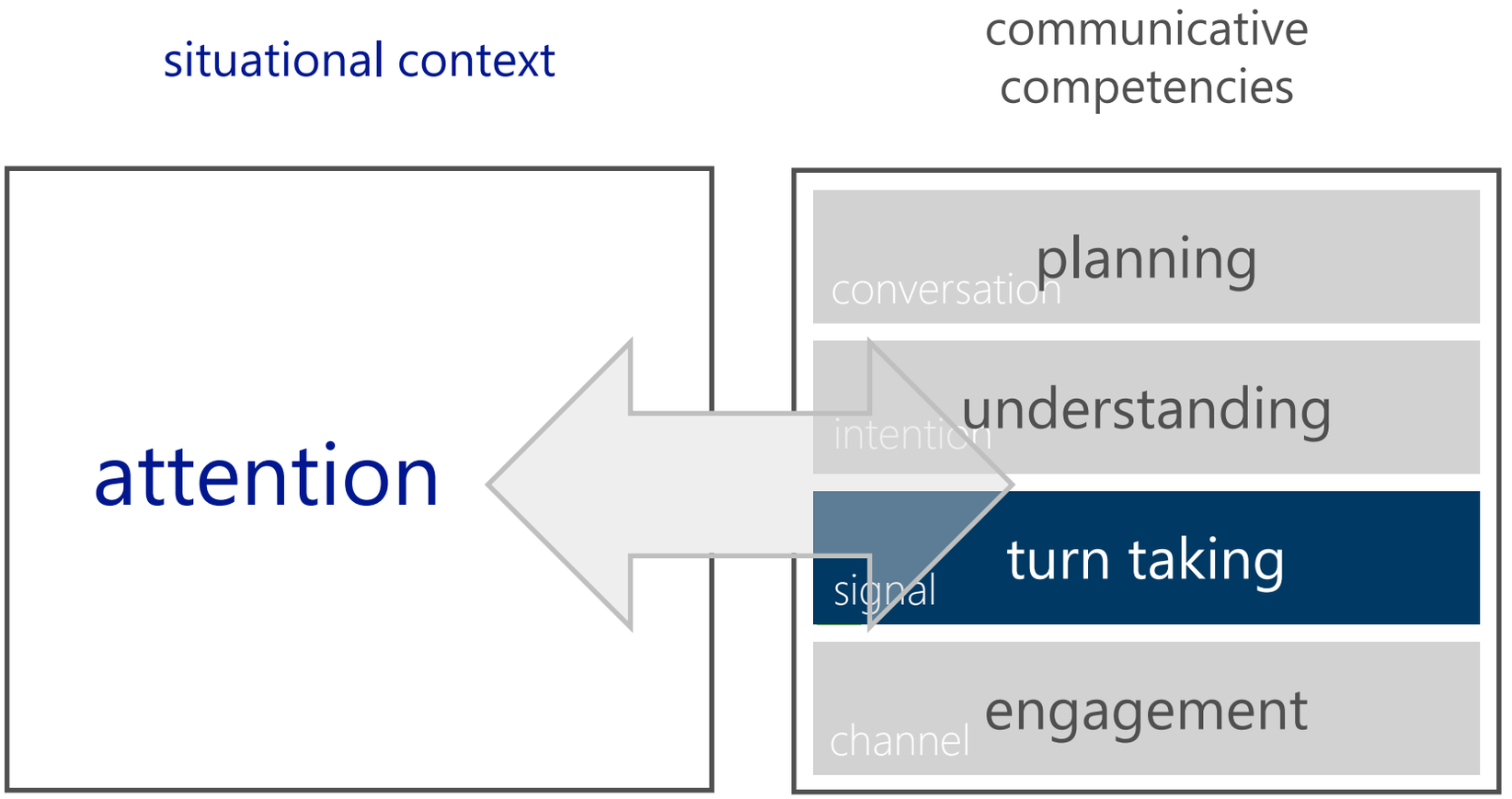


Physically situated language interaction

Physically situated language interaction



Coordination of attention and speech production



Coordination of attention and speech production

[Yu, Bohus and Horvitz, 2015 – *Incremental Coordination: Attention-Centric Speech Production in a Physically Situated Conversational Agent*, to appear in SIGdial'2015]

Charles Goodwin: disfluencies and attention

from *Conducting Interaction: Achieving Mutual Orientation at Turn Beginning*

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

Listener:

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

Listener:

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

Listener:

Coordination of attention and speech production

[Yu, Bohus and Horvitz, 2015 – *Incremental Coordination: Attention-Centric Speech Production in a Physically Situated Conversational Agent*, to appear in SIGdial'2015]

Charles Goodwin: disfluencies and attention

from *Conducting Interaction: Achieving Mutual Orientation at Turn Beginning*

Speaker: She– she's reaching the p– She's at the point I'm

Listener:

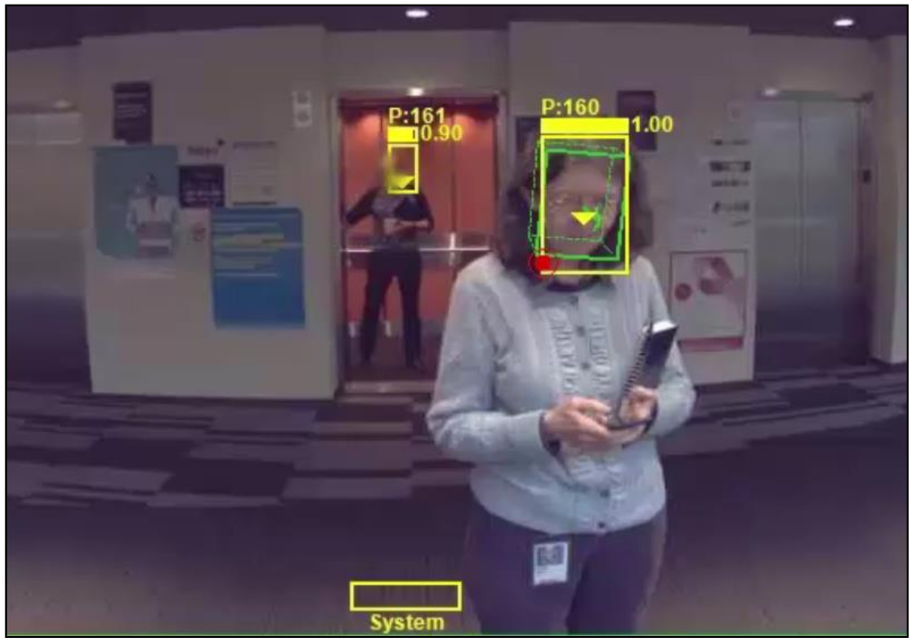
Speaker: I ask him, (0.1) I ask him if he– (0.4) could– If *you* could call 'im when you go *in*.

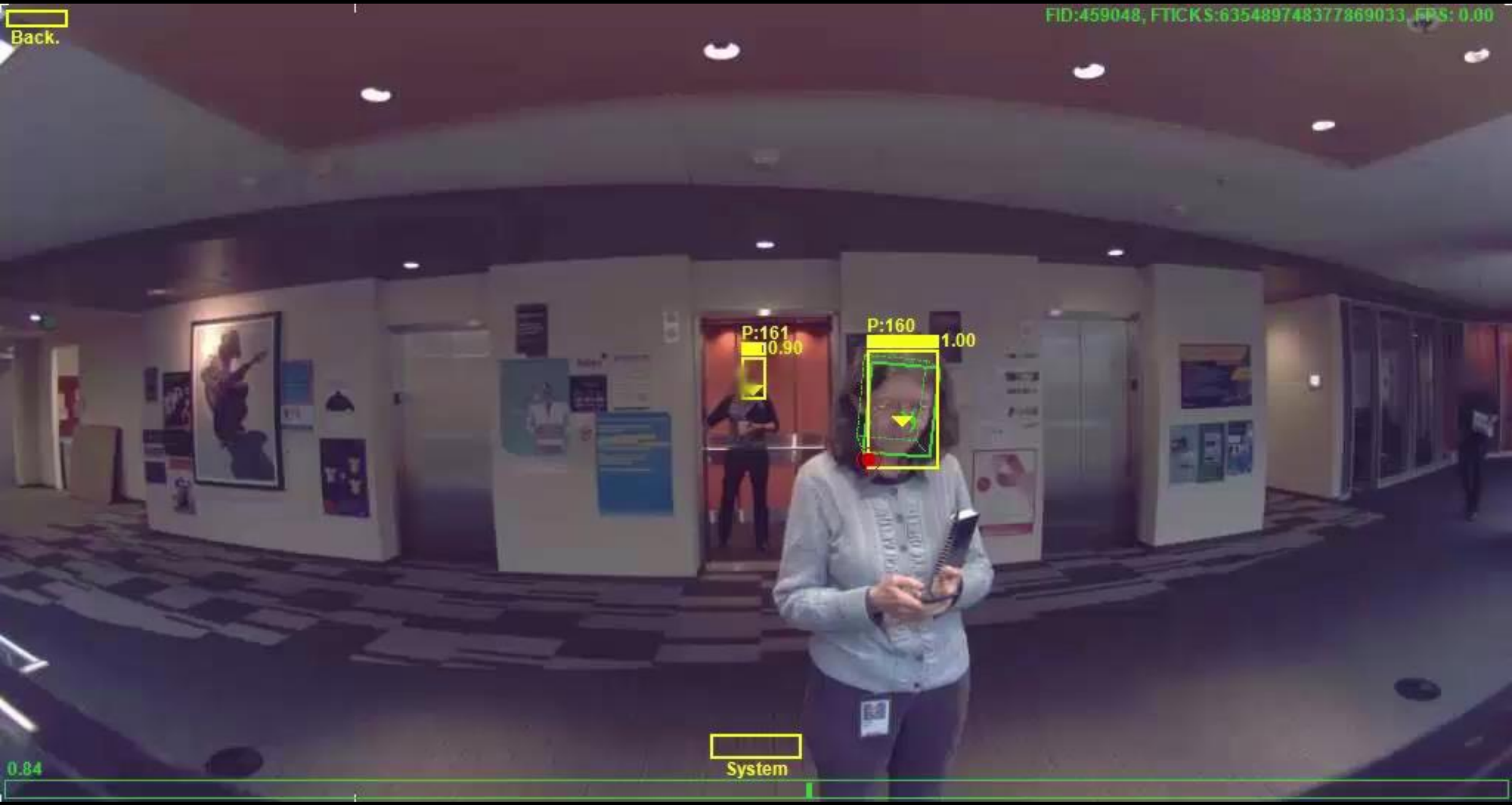
Listener:

Turn taking models dialog systems

Push to talk

You-speak-then-I-speak




Back.
FID:459048, FTICKS:635489748377869033, FPS: 0.00
P:161
0.90
P:160
1.00
System
0.84

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

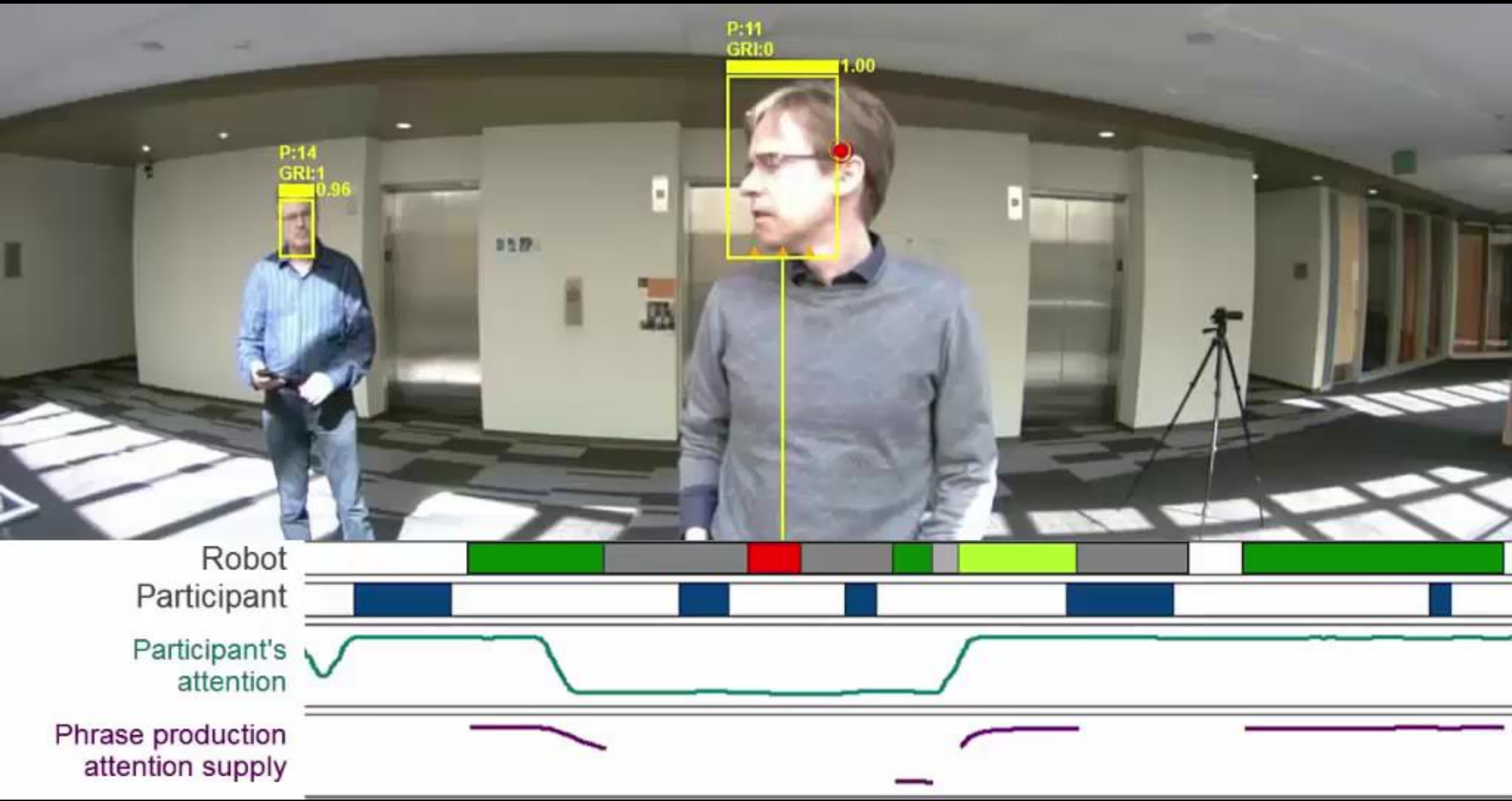
Coordinative policy

… … … Excuse me! … … … To get … … … To get to 3800 go to […]

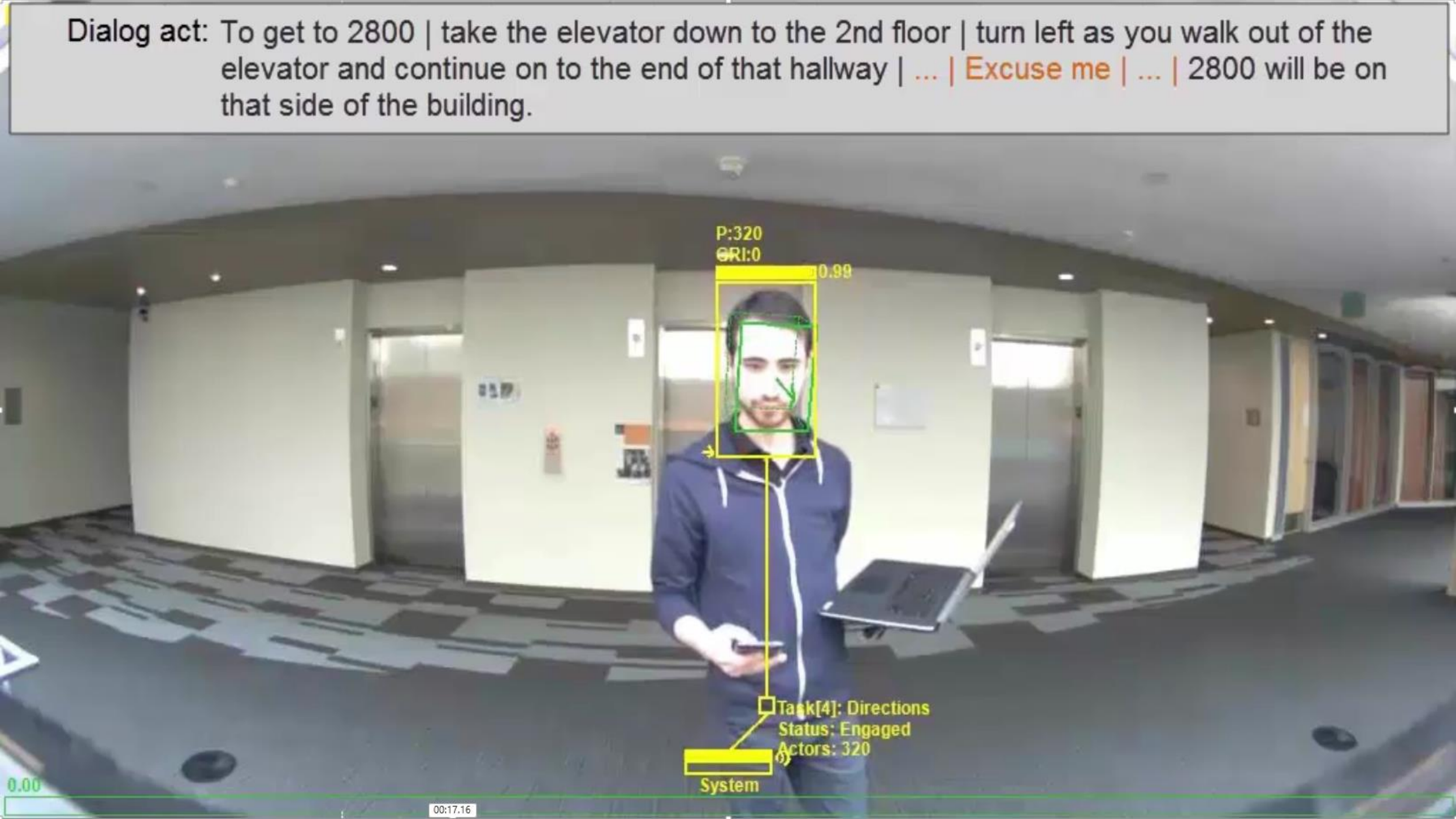
Phrase 1 | Phrase 2

a demonstration video …

the robot's view …
sensing and computation details


P:11
GRI:0
1.00
P:14
GRI:1
0.96
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.
P:320
GRI:0
0.99
Task[4]: Directions
Status: Engaged
Actors: 320
System
0.00
00:17.16

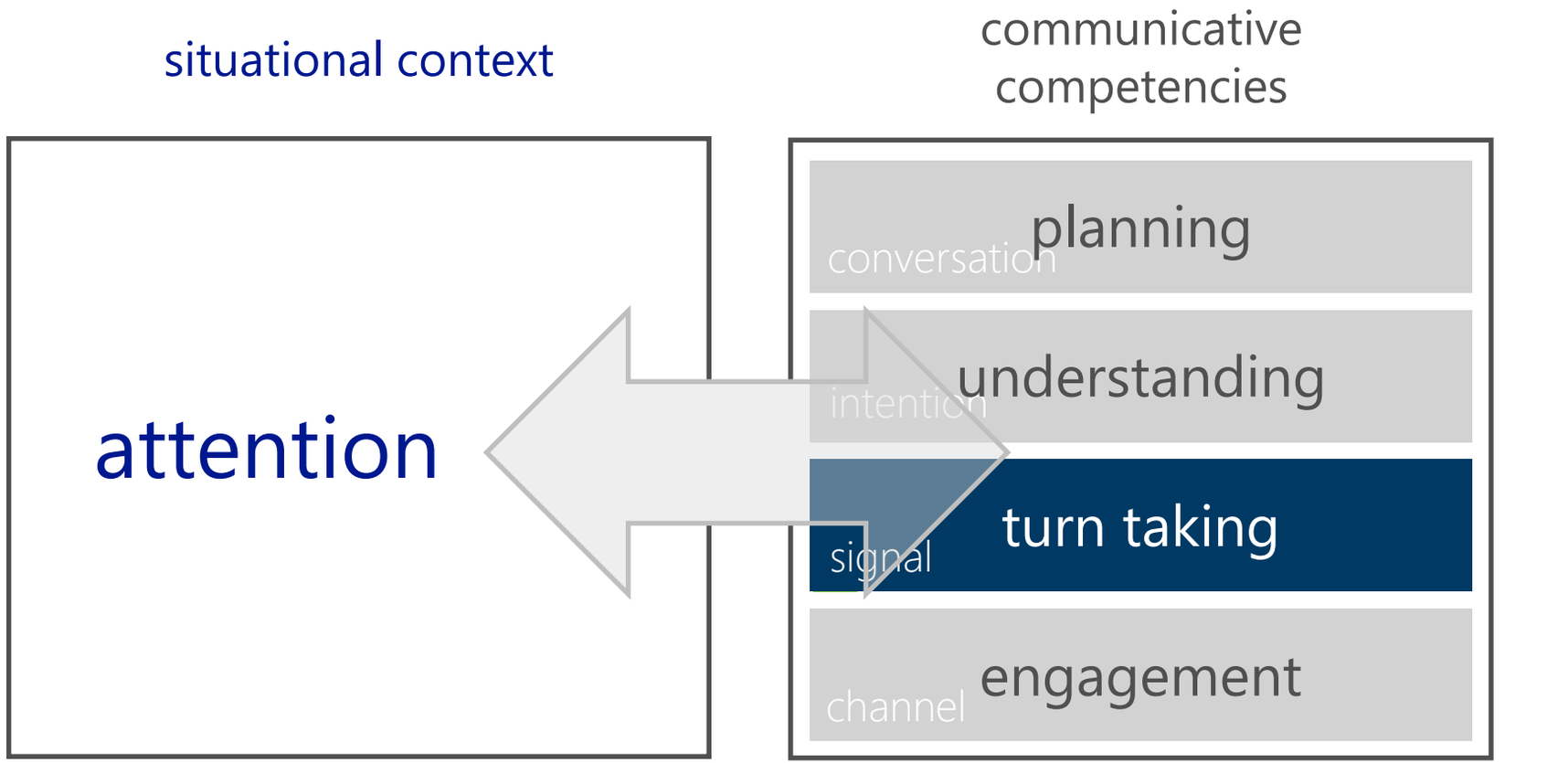
Dialog act:


P:320
GR:0
0.98
Task[4]: Directions
Status: Engaged
Actors: 20
System

failure cases …

system incorrectly infers user is not attending and inappropriately triggers pauses, interjections and restarts

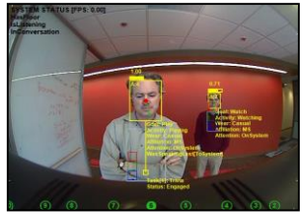
Physically situated language interaction



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

communicative competencies

- planning
- understanding
- turn taking
- engagement

conversation

intention

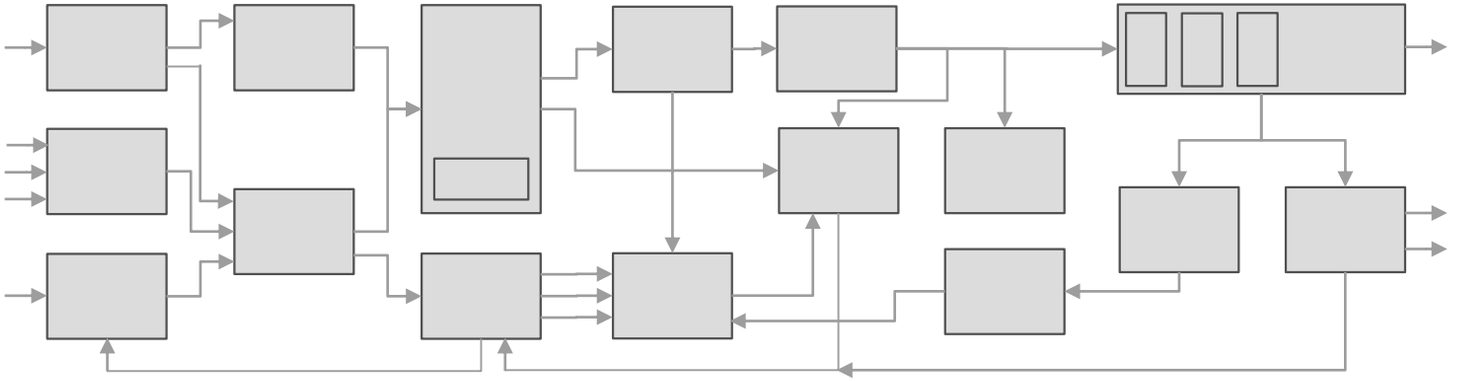
signal

channel

Challenges with integrative-AI systems

Challenges with integrative-AI systems

complex (many components)



Microphone array capture
Sound source localization
Speech recognition
Language understanding
Infrared proximity sensors
Badge sensors
Face detection and tracking
Head-pose tracking

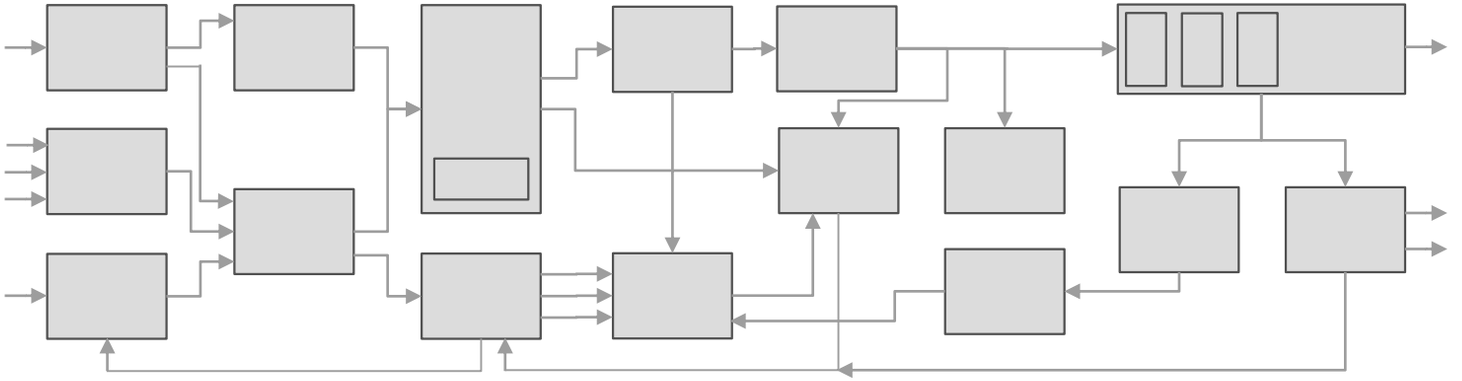
Facial feature tracking
Face identity recognition
Gender detection
Attention models
Engagement models
Turn-taking models
Behavioral control

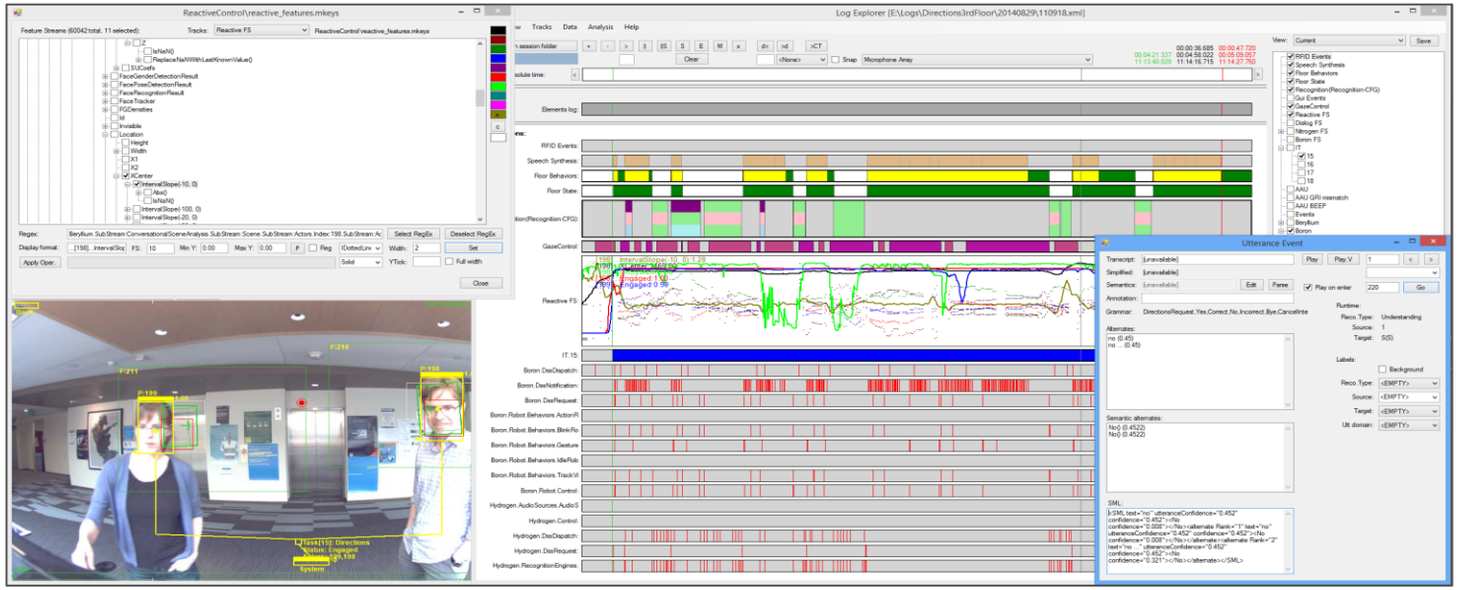
Dialog management
Natural language generation
Speech synthesis
Avatar synthesis
Robot motion control
Floor-plan models
User models

Challenges with integrative-AI systems

complex (many components)

programming models for coordinated computation; tools





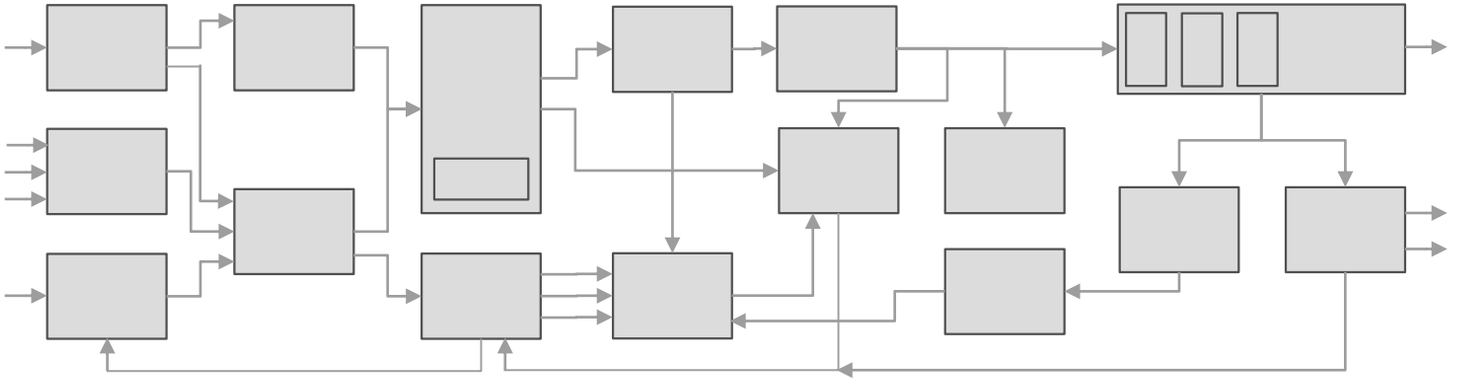
Challenges with integrative-AI systems

complex (many components)

programming models for coordinated computation; tools

act in real-time, under uncertainty

evolve programming languages? e.g. time & uncertainty

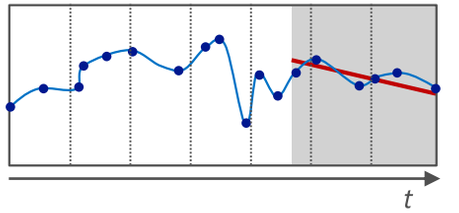


Time and streams as 1st order citizens

double f; → stream double f;

f=3; f=x*f-y;

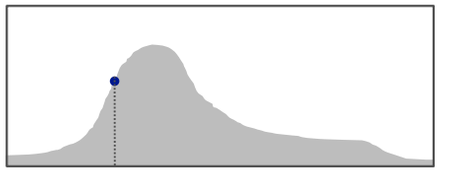
Automatic persistence, historical access, sampling, transforms



Uncertainty as 1st order citizen

double f; → uncertain double f;

Representation, sampling, inference, belief updates



Challenges with integrative-AI systems

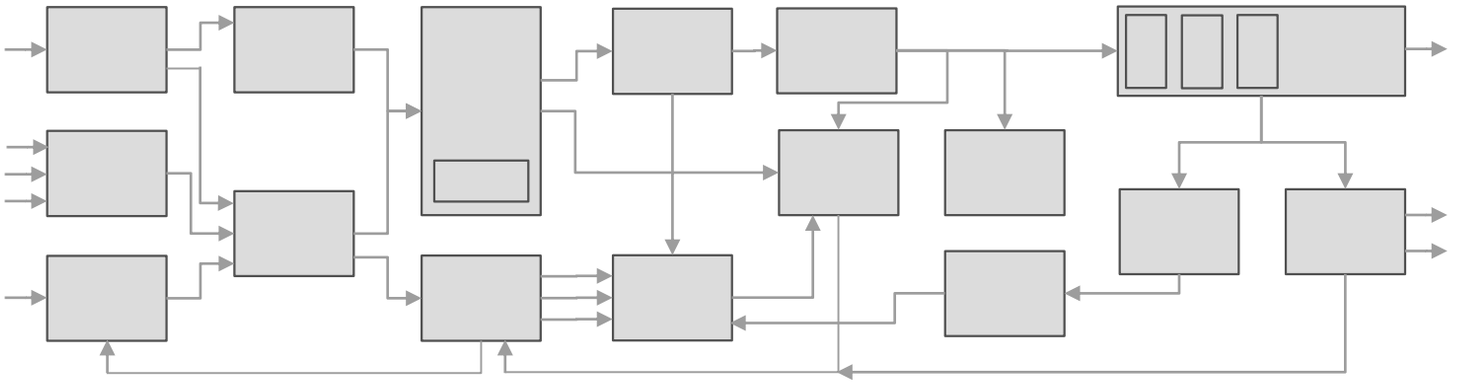
complex (many components)

programming models for coordinated computation; tools

act in real-time, under uncertainty

evolve programming languages? e.g. time & uncertainty

integration of human- and machine-authored components



Microphone array capture
Sound source localization
Speech recognition
Language understanding
Infrared proximity sensors
Badge sensors
Face detection and tracking
Head-pose tracking

Facial feature tracking
Face identity recognition
Gender detection
Attention models
Engagement models
Turn-taking models
Behavioral control

Dialog management
Natural language generation
Speech synthesis
Avatar synthesis
Robot motion control
Floor-plan models
User models

Challenges with integrative-AI systems

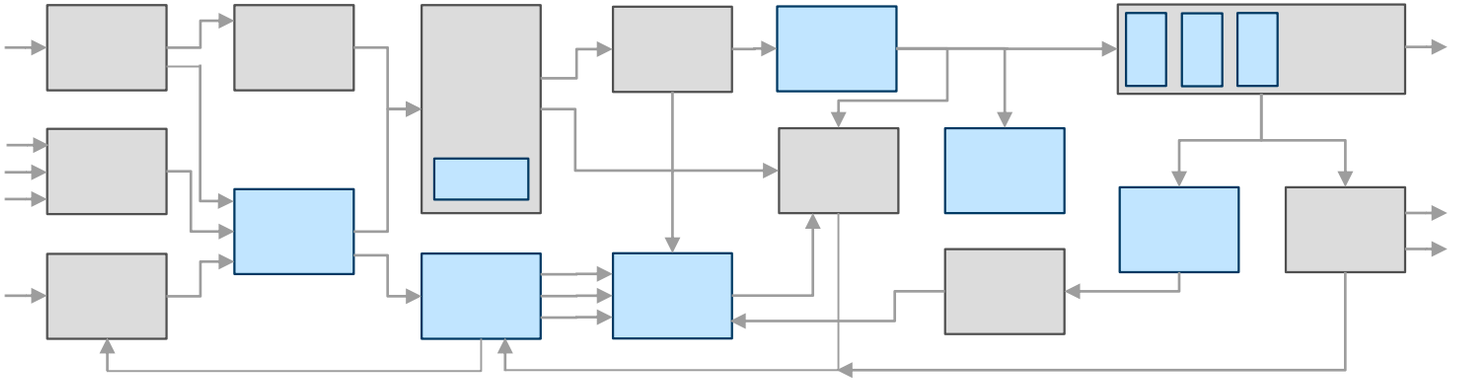
complex (many components)

programming models for coordinated computation; tools

act in real-time, under uncertainty

evolve programming languages? e.g. time & uncertainty

integration of human- and machine-authored components



Microphone array capture
Sound source localization
Speech recognition
Language understanding
Infrared proximity sensors
Badge sensors
Face detection and tracking
Head-pose tracking

Facial feature tracking
Face identity recognition
Gender detection
Attention models
Engagement models
Turn-taking models
Behavioral control

Dialog management
Natural language generation
Speech synthesis
Avatar synthesis
Robot motion control
Floor-plan models
User models

Challenges with integrative-AI systems

complex (many components)

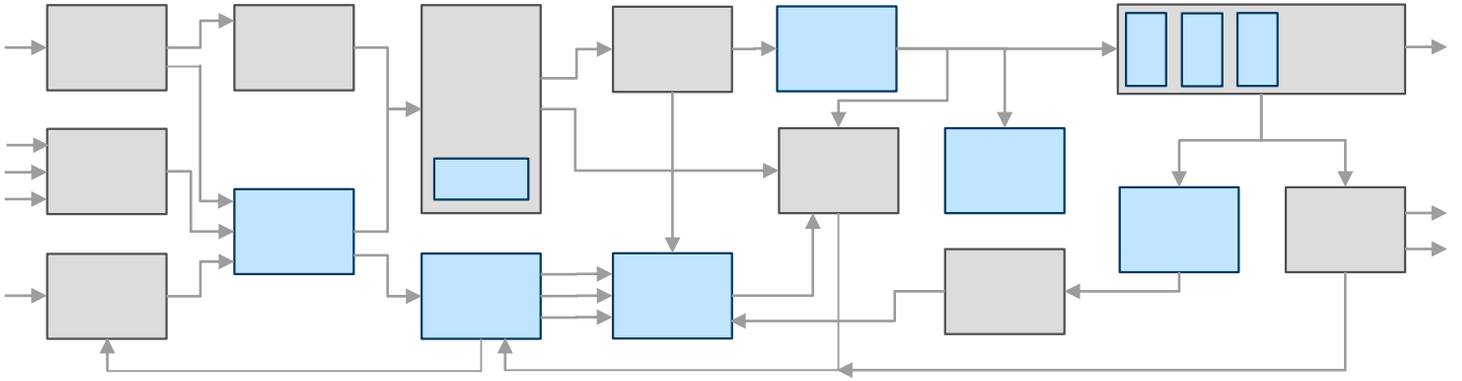
programming models for coordinated computation; tools

act in real-time, under uncertainty

evolve programming languages? e.g. time & uncertainty

integration of human- and machine-authored components

engineering of integrated learning systems



Engineering of integrated learning systems

learning in connected systems (new frontiers for ML & software engineering?)

learning in interactive settings: online, lifelong vs. batch

Challenges with integrative-AI systems

complex (many components)

programming models for coordinated computation; tools

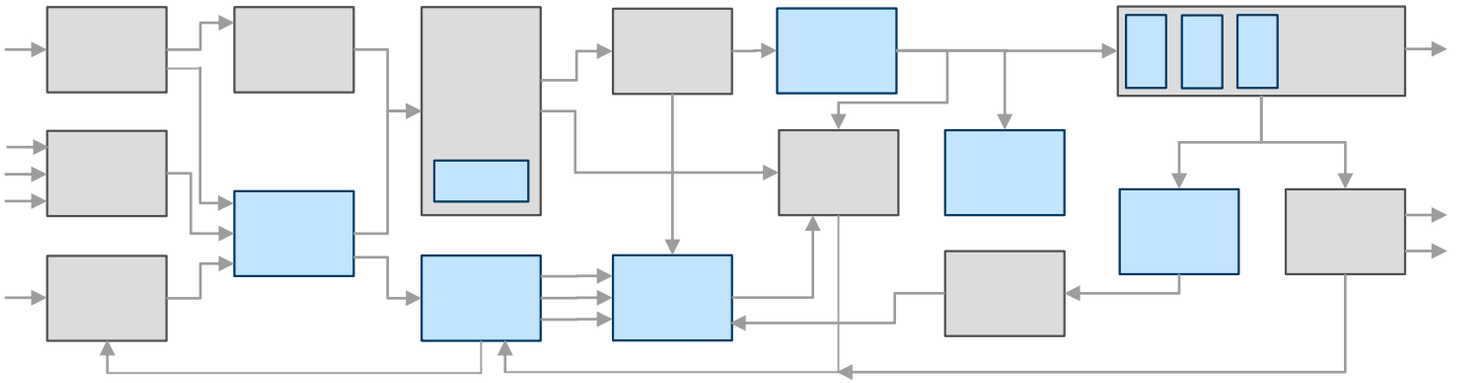
act in real-time, under uncertainty

evolve programming languages? e.g. time & uncertainty

integration of human- and machine-authored components

engineering of integrated learning systems

meta-reasoning & system-level (self)-optimization



Engineering of integrated learning systems

learning in connected systems (new frontiers for ML & software engineering?)

learning in interactive settings: online, lifelong vs. batch

Meta-reasoning and system-level (self)-optimization

self-monitoring and diagnosis / blame assignment

self-optimization

Challenges with integrative-AI systems

complex (many components)

programming models for coordinated computation; tools

act in real-time, under uncertainty

evolve programming languages? e.g. time & uncertainty

integration of human- and machine-authored components

engineering of integrated learning systems

meta-reasoning & system-level (self)-optimization