Situated Interaction in the Open World:
New Systems and Challenges

Sean Andrist
with Dan Bohus, Stuart Dent, Ashley Feniello, Eric Horvitz, Mihai Jalobeanu, Ece Kamar, Nick Saw, Pat Sweeney, Andy Wilson, Zhou Yu, and more

Microsoft Research AI
Physically Situated AI Systems

Verbal + nonverbal
Embodiment
Physical context
Social context
History
....
Today

Some existing systems
Research challenges
Platform for situated intelligence
Smart Elevator
Virtual Assistant
Directions Robots
Mobile Guide Robot ("PsiBot")
Challenge: Engagement
Engagement

The process by which participant initiate, maintain, and break their perceived connection
Approach: Forecasting and Hesitations

Strategically use hesitations to mitigate high uncertainty

Is there anything else I can help you find?

So ...

Well, guess I’ll catch you later then!

Is there anything else I can help you find?
Challenge: Turn-Taking
Turn-Taking

Process by which participants synchronize their verbal exchanges
Approach: Coordinate Speech and Attention
Challenge: Spatial Configurations

Too close

Too far

Close bystander

Large group
**F-Formations**

Spatial and orientational relationship between two or more people in a focused interaction (Kendon, 1980)

**Proxemics**

Organization of spatial relationships and interpersonal distances in everyday life (Hall, 1963)
Approach: Scene Shaping
Common Themes

In-the-Wild

Problems not apparent or reproducible in lab settings

Multimodal Reasoning and Fusion

Forecasting

Online and self-supervised
Building/Maintaining Situated AI Systems

- 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
Platform for Situated Intelligence (\psi)

an open, extensible framework for developing and studying situated, integrative-AI systems
Overall Architecture of \psi

Enable easy development while retaining high-performance

RUNTIME
- time-aware streams
- isolation
- scheduling
- persistence
Runtime: Coordinated Pipelines

```javascript
// instantiate camera and microphone
var camera = new Camera();
var microphone = new MicrophoneArray();

// instantiate speech source detector
var ssi = new SpeechSourceIdentification();

// track faces from camera and connect to
// speech source detector
camera.Out
  .TrackFaces()
  .ConnectTo(ssi.FacesInput);

// do localization and connect to speech
// source detector
microphone.Out
  .Localize()
  .ConnectTo(ssi.SourceAngleInput);
```
Overall Architecture of \psi

Enable a fast debugging + visualization + ML cycle

Enable easy development while retaining high-performance

TOOLS
- visualization & debugging
- data processing
- machine learning

RUNTIME
- time-aware streams
- isolation
- scheduling
- persistence
Tools: Visualization
Overall Architecture of ψi

- Ecosystem of pluggable AI components
- Enable a fast debugging + visualization + ML cycle
- Enable easy development while retaining high-performance

**COMPONENTS**
- sensors | imaging | vision | audio | speech | interaction | ...

**TOOLS**
- visualization & debugging | data processing | machine learning

**RUNTIME**
- time-aware streams | isolation | scheduling | persistence
Thank You!

Perception and Interaction Group

Situated Interaction Project
https://www.microsoft.com/en-us/research/project/situated-interaction/

Platform for Situated Intelligence (psi)