

Platform for Situated Intelligence (or \psi)

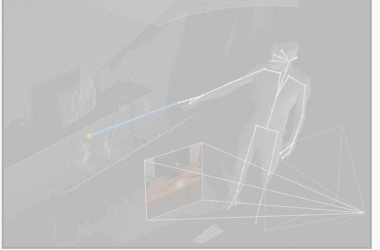
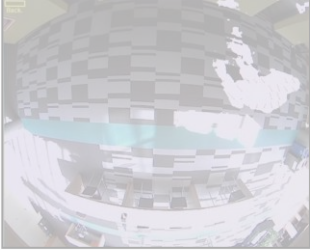
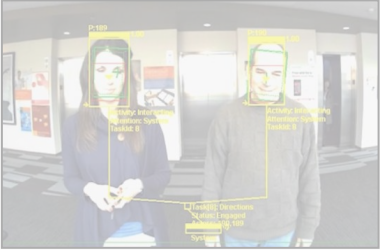
an open-source framework for multimodal, integrative-AI

Dan Bohus, Sean Andrist

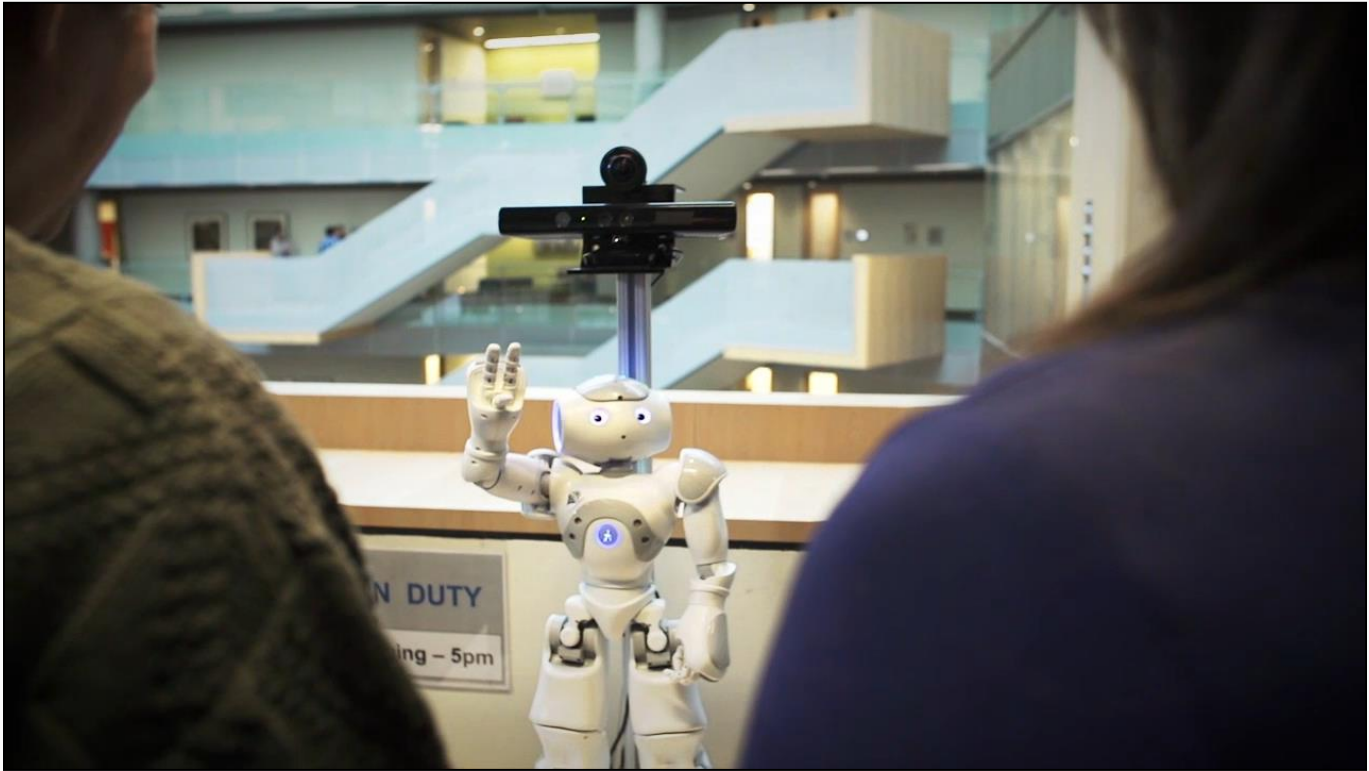
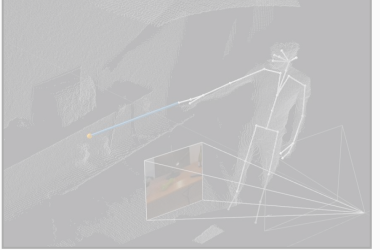
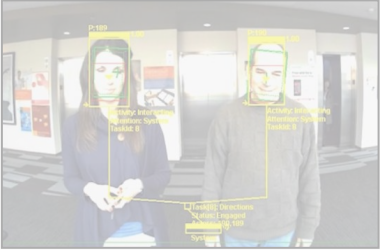
Contributors and collaborators: Mike Barnett, John Elliott, Ashley Feniello, Don Gillett, Eric Horvitz, Mihai Jalobeanu, Daniel McDuff, Lev Nachmanson, Ann Paradiso, Kael Rowan, Nick Saw, Patrick Sweeney, Anne Loomis Thompson

Microsoft Research
Redmond, WA, USA

SITUATED INTERACTION RESEARCH

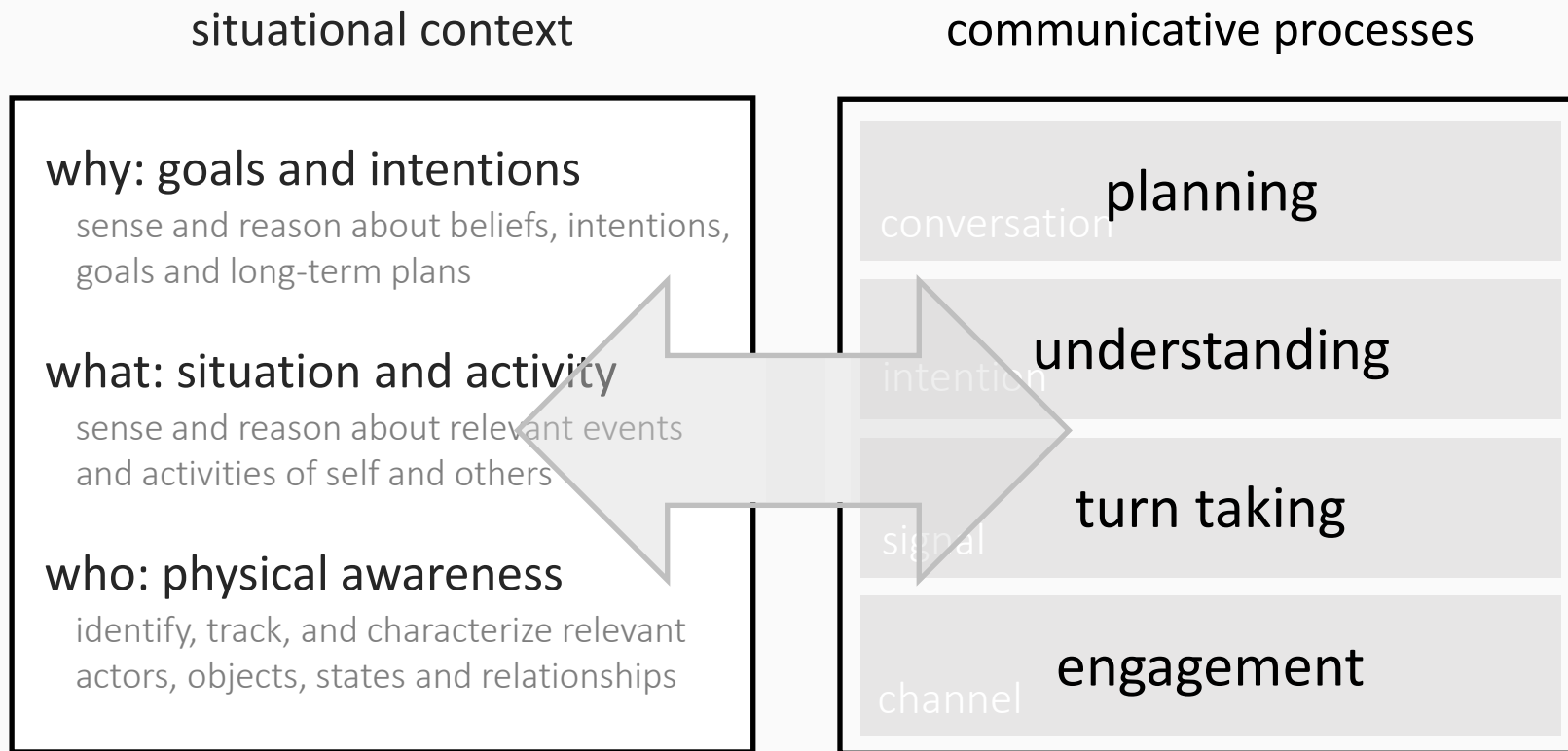


SITUATED INTERACTION RESEARCH



SITUATED INTERACTION RESEARCH

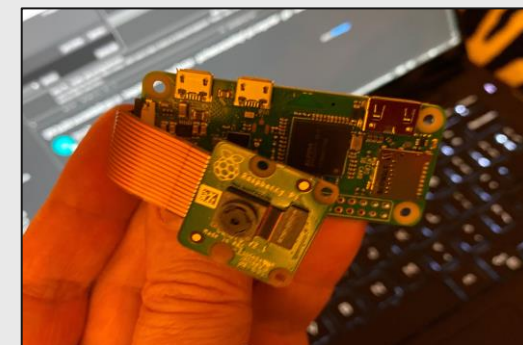
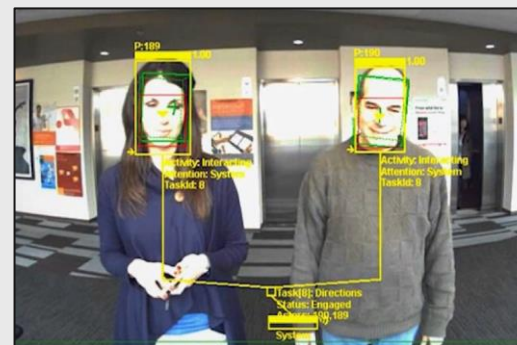
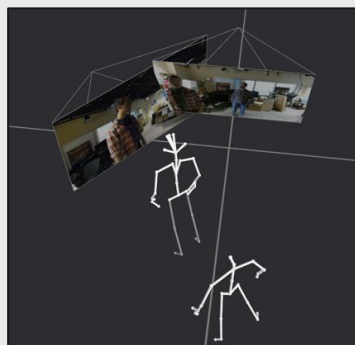
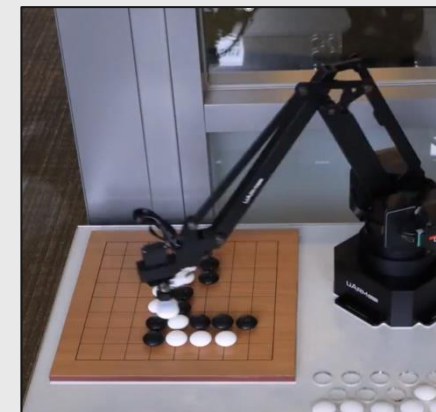
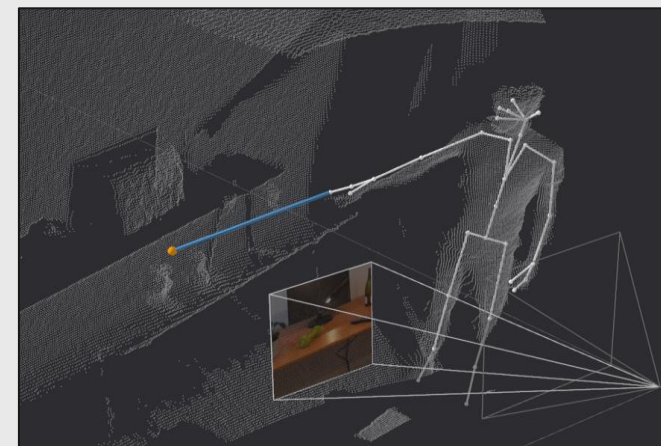
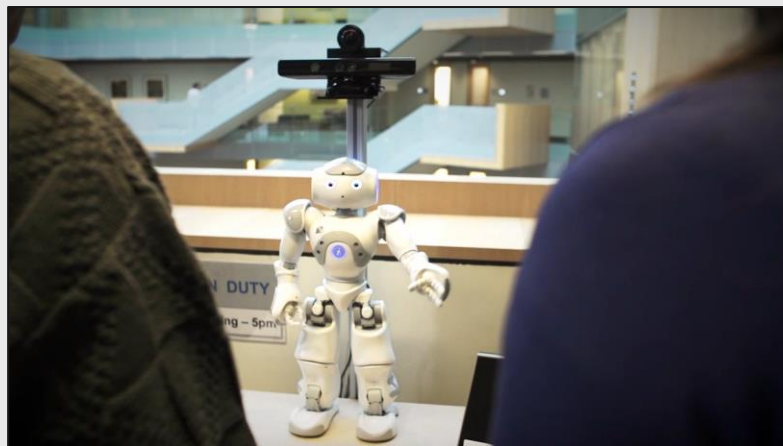
enable computers to understand and participate in the social physics of human interactions



Multimodal
Streaming Data

Integrate Many
Technologies

Operate Under
Latency Constraints



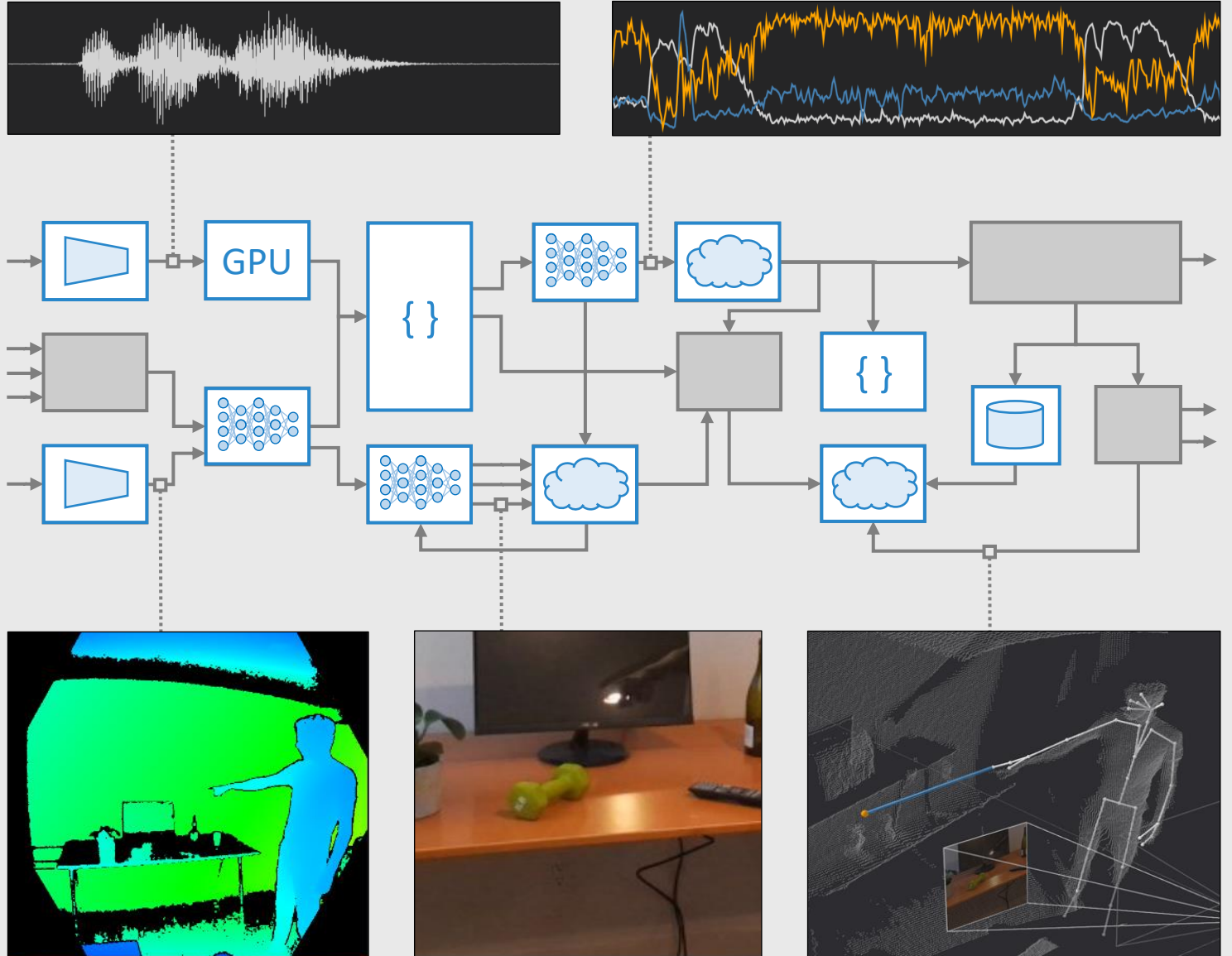
CHALLENGES

Multimodal

Integrative

Primitives

Dev Tools



Platform for Situated Intelligence

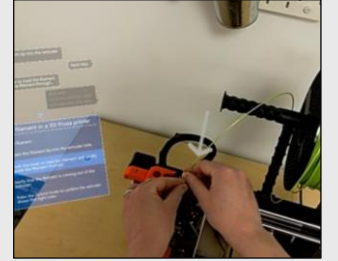
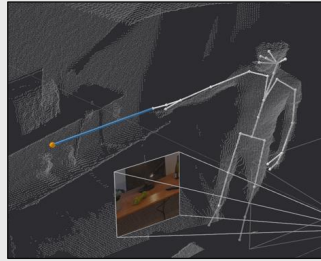
Lower the engineering costs

Foster innovation and research

Open-source: www.github.com/microsoft/psi

.NET standard, cross-platform

Platform for Situated Intelligence



COMPONENTS

sensors | imaging | audio | vision | speech | language | ...

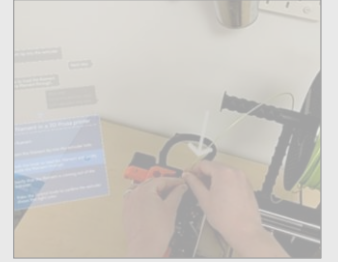
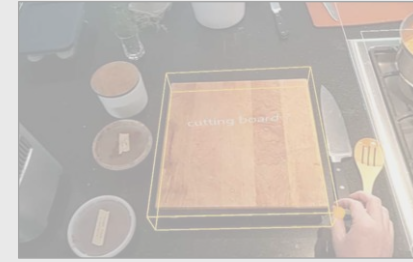
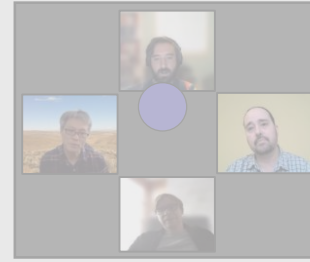
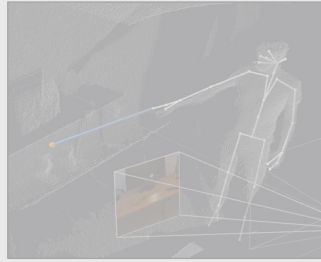
TOOLS

data visualization | debugging | annotation | processing

RUNTIME

streaming | logging | parallel coordinated computation

Platform for Situated Intelligence



COMPONENTS

sensors | imaging | audio | vision | speech | language | ...

TOOLS

data visualization | debugging | annotation | processing

RUNTIME

streaming | logging | parallel coordinated computation

APPLICATIONS

```
// create pipeline
var p = Pipeline.Create();

// instantiate components
var microphone = new AudioCapture(p);
var camera = new MediaSource(p);

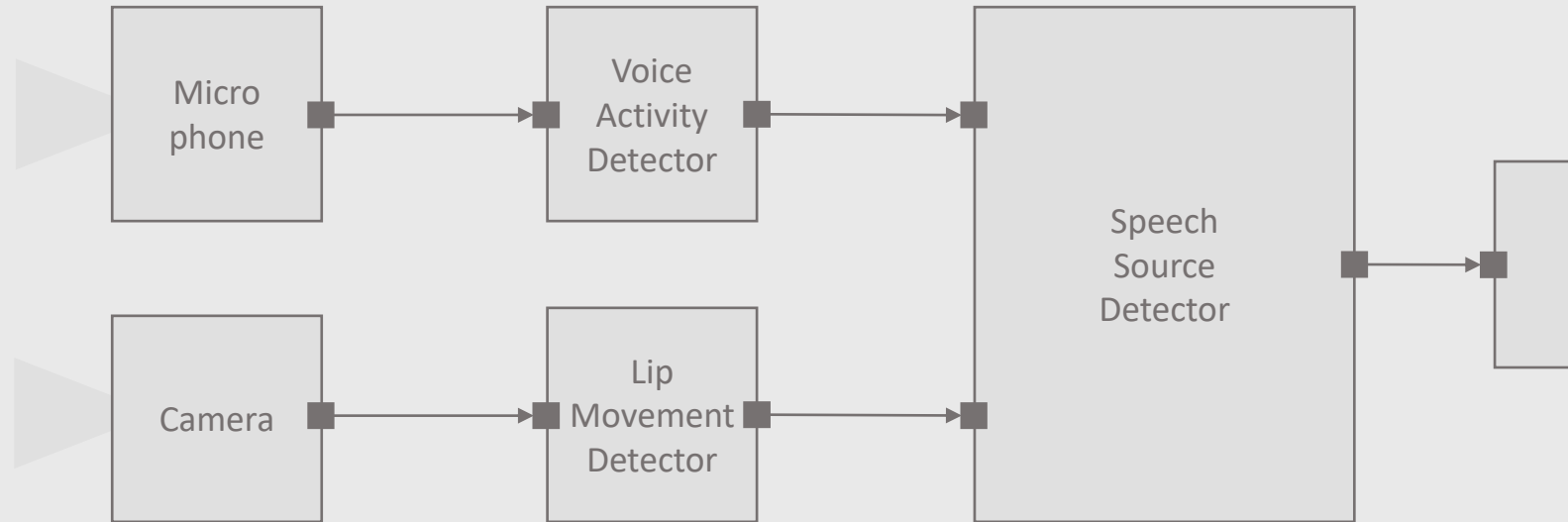
...

// connect components
microphone.PipeTo(voiceActivityDetector);
camera.PipeTo(lipMovementDetector);

...

// run the pipeline
p.Run();
```

RUNTIME



EXECUTION

```
// create pipeline
var p = Pipeline.Create();

// instantiate components
var microphone = new AudioCapture(p);
var camera = new MediaSource(p);

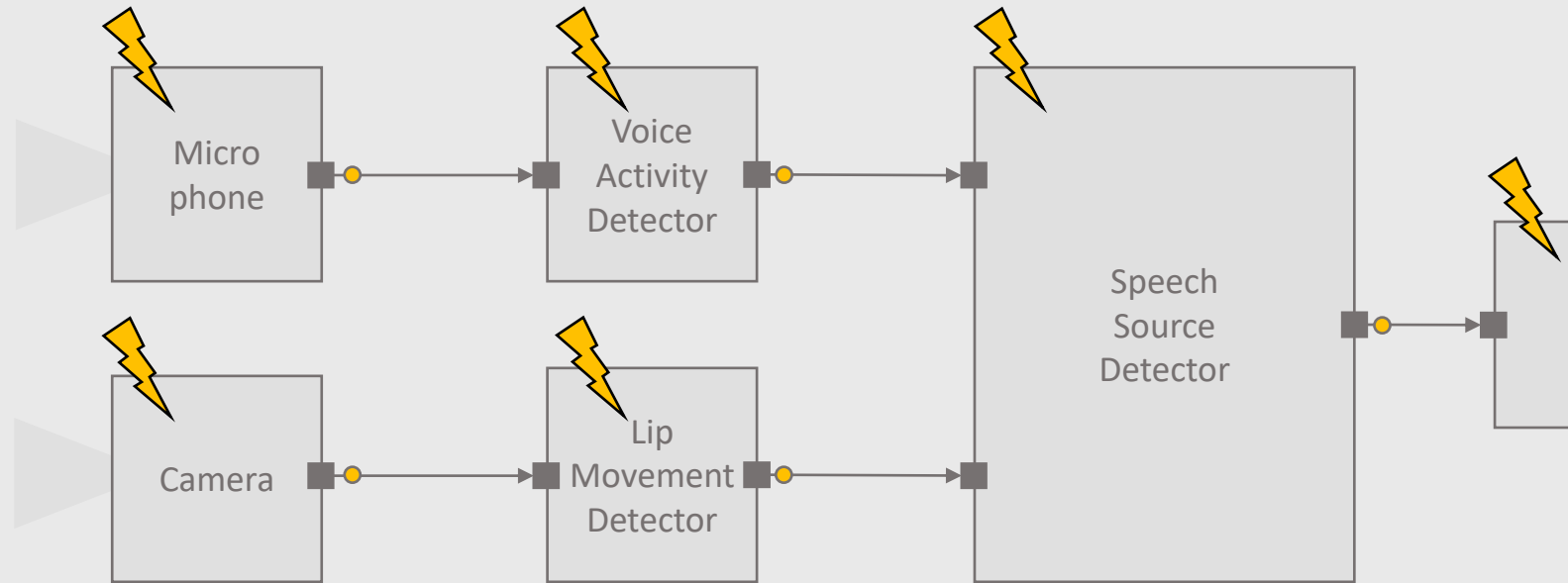
...

// connect components
microphone.PipeTo(voiceActivityDetector);
camera.PipeTo(lipMovementDetector);

...

// run the pipeline
p.Run();
```

RUNTIME

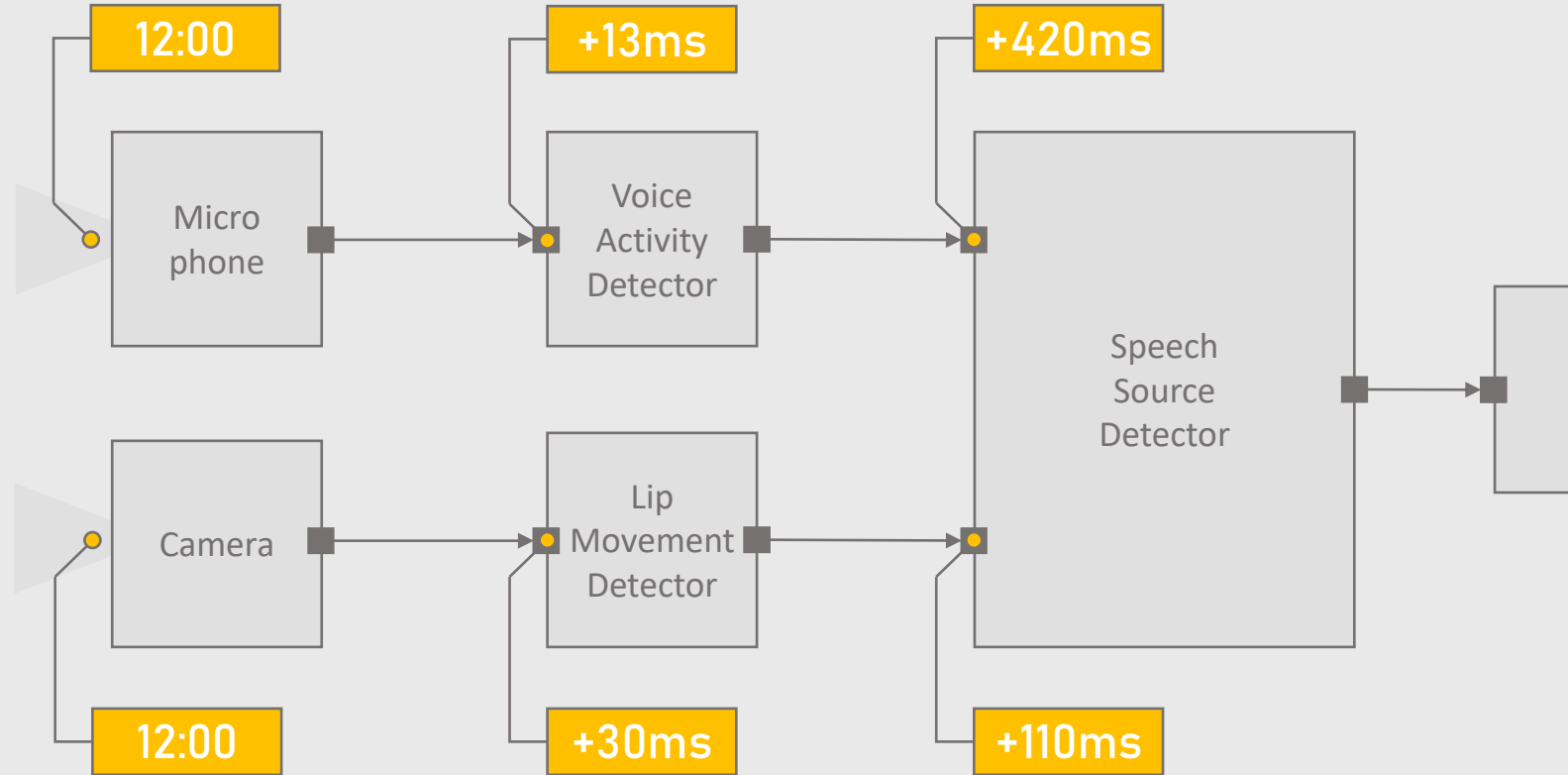


STREAMS

Originating Times

Synchronization

RUNTIME



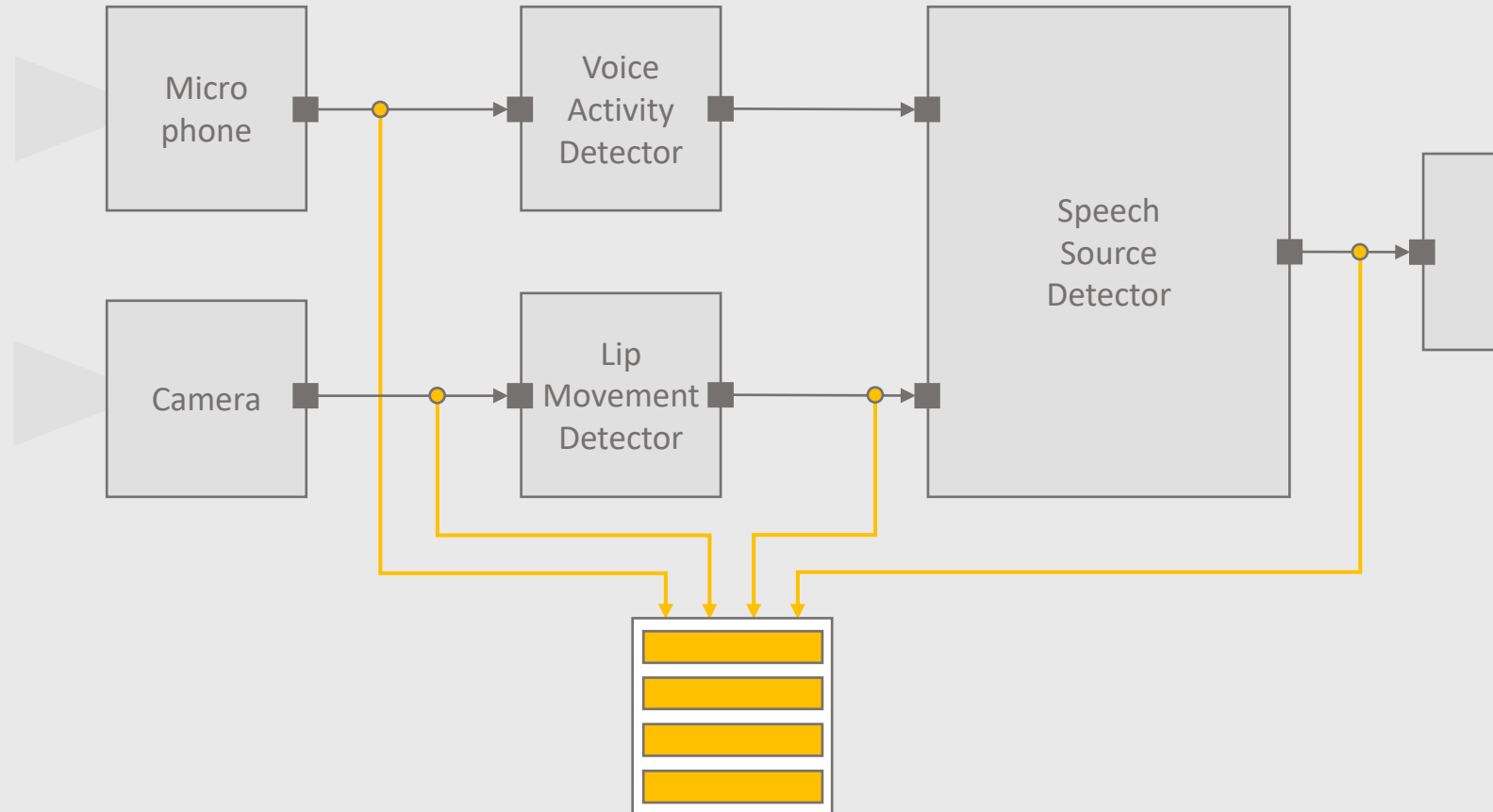
STREAMS

Originating Times

Synchronization

Persistence

RUNTIME



STREAMS

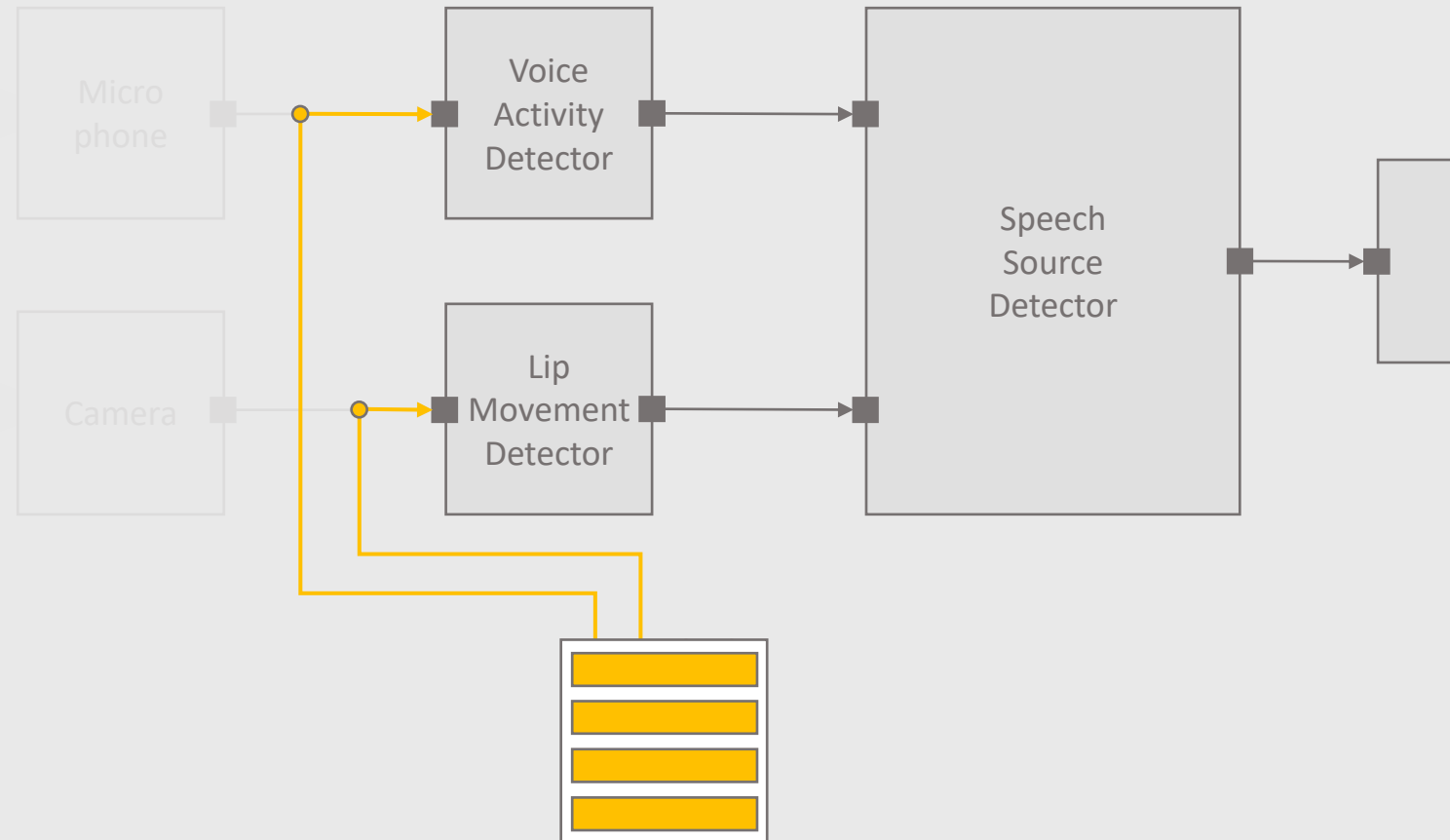
Originating Times

Synchronization

Persistence

Reproducible Replay

RUNTIME



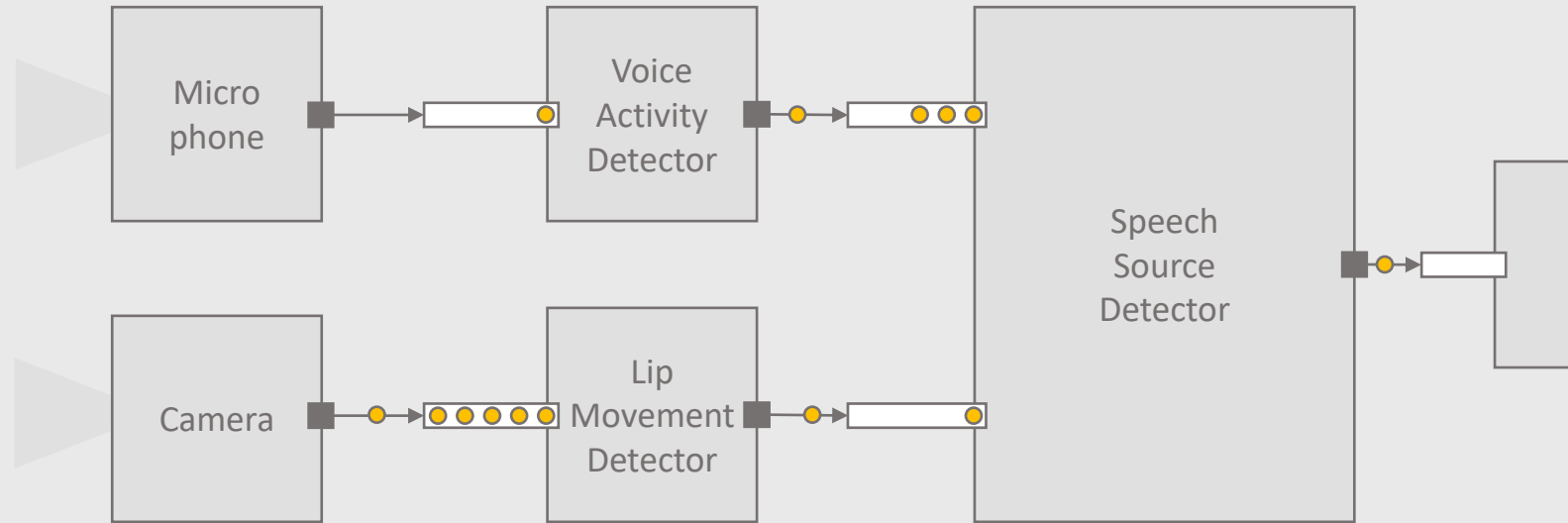
EFFICIENCIES

Pipeline Parallelism

Time-aware
Scheduling

Fine-grained Control

RUNTIME



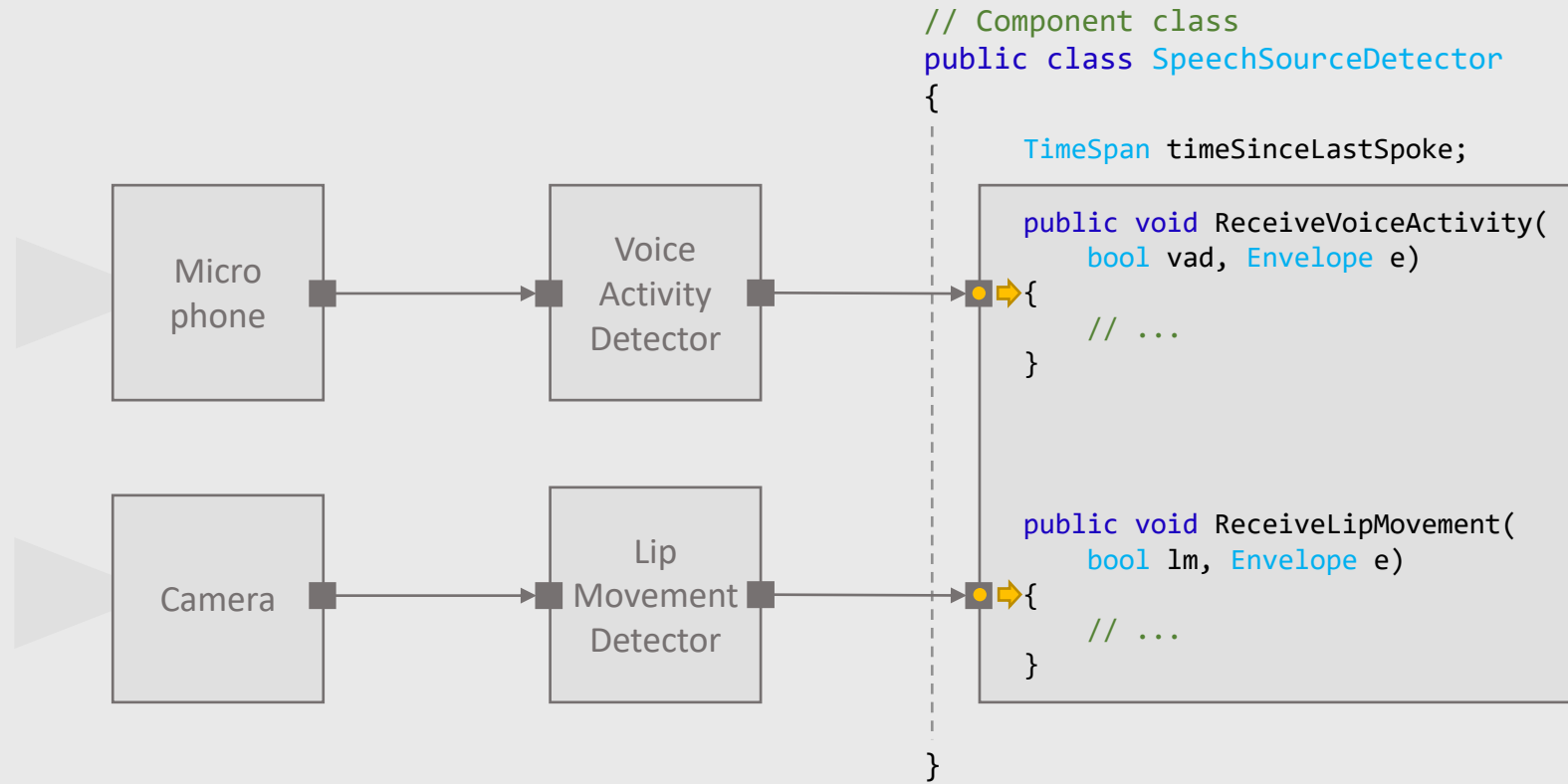
WRITING COMPONENTS

Easy to Write

State Protection

Isolated Execution

RUNTIME



WRITING COMPONENTS

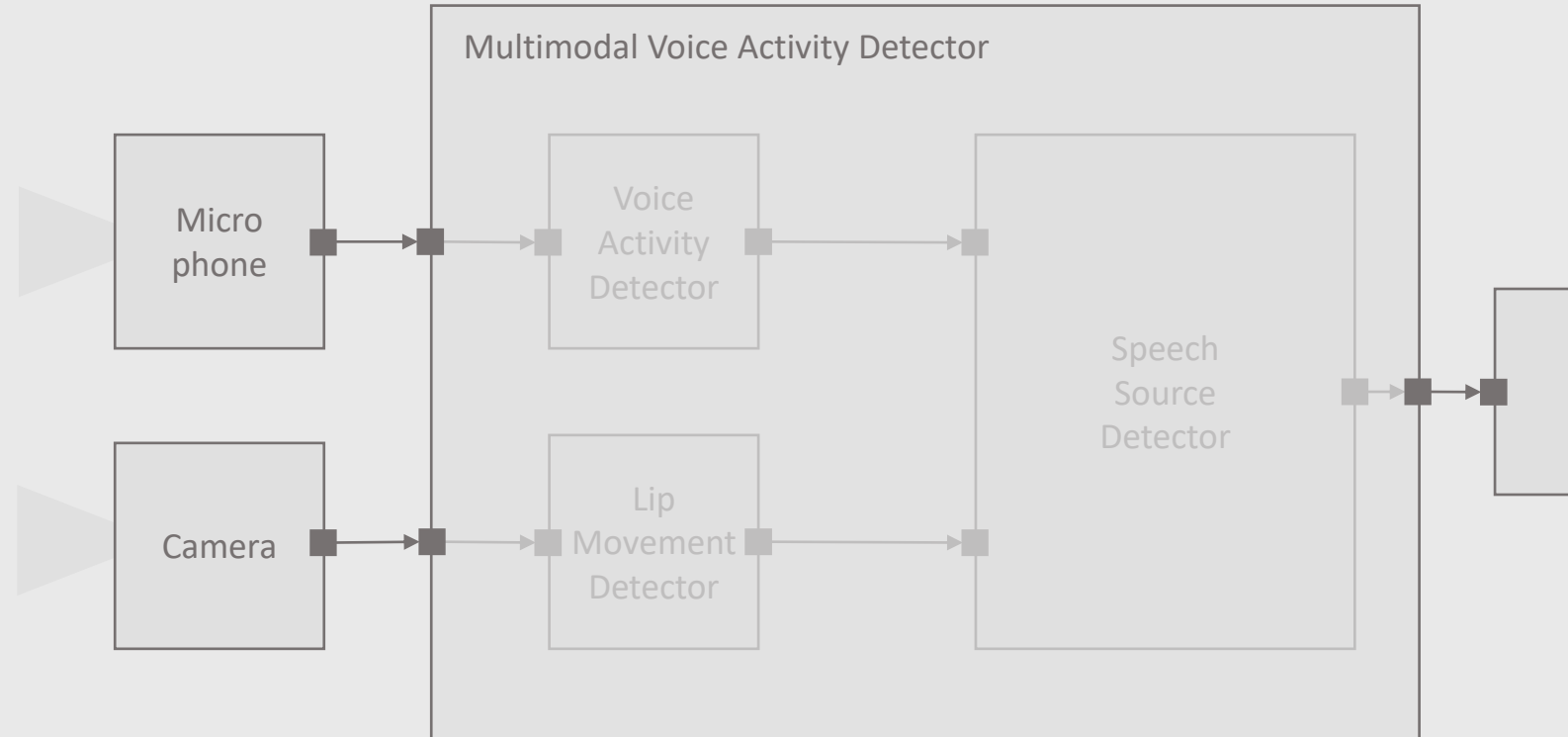
Easy to Write

State Protection

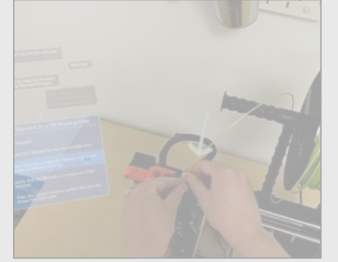
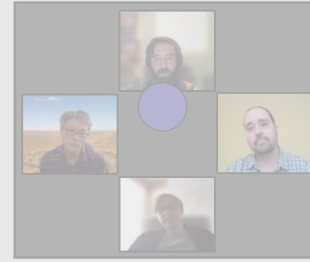
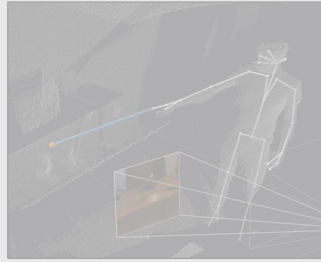
Isolated Execution

Hierarchical
Composition

RUNTIME



Platform for Situated Intelligence



COMPONENTS

sensors | imaging | audio | vision | speech | language | ...

TOOLS

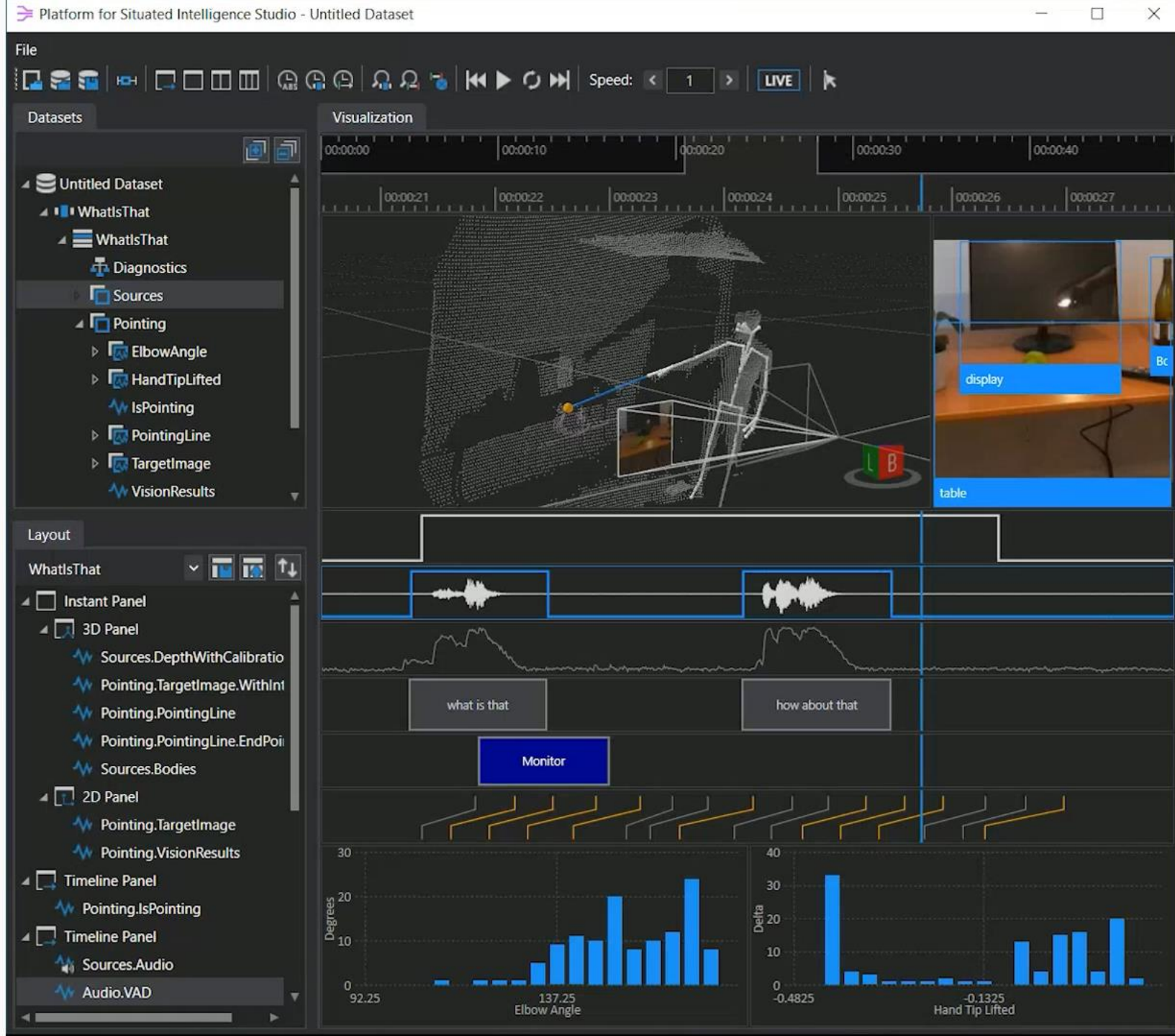
data visualization | debugging | annotation | processing

RUNTIME

streaming | logging | parallel coordinated computation

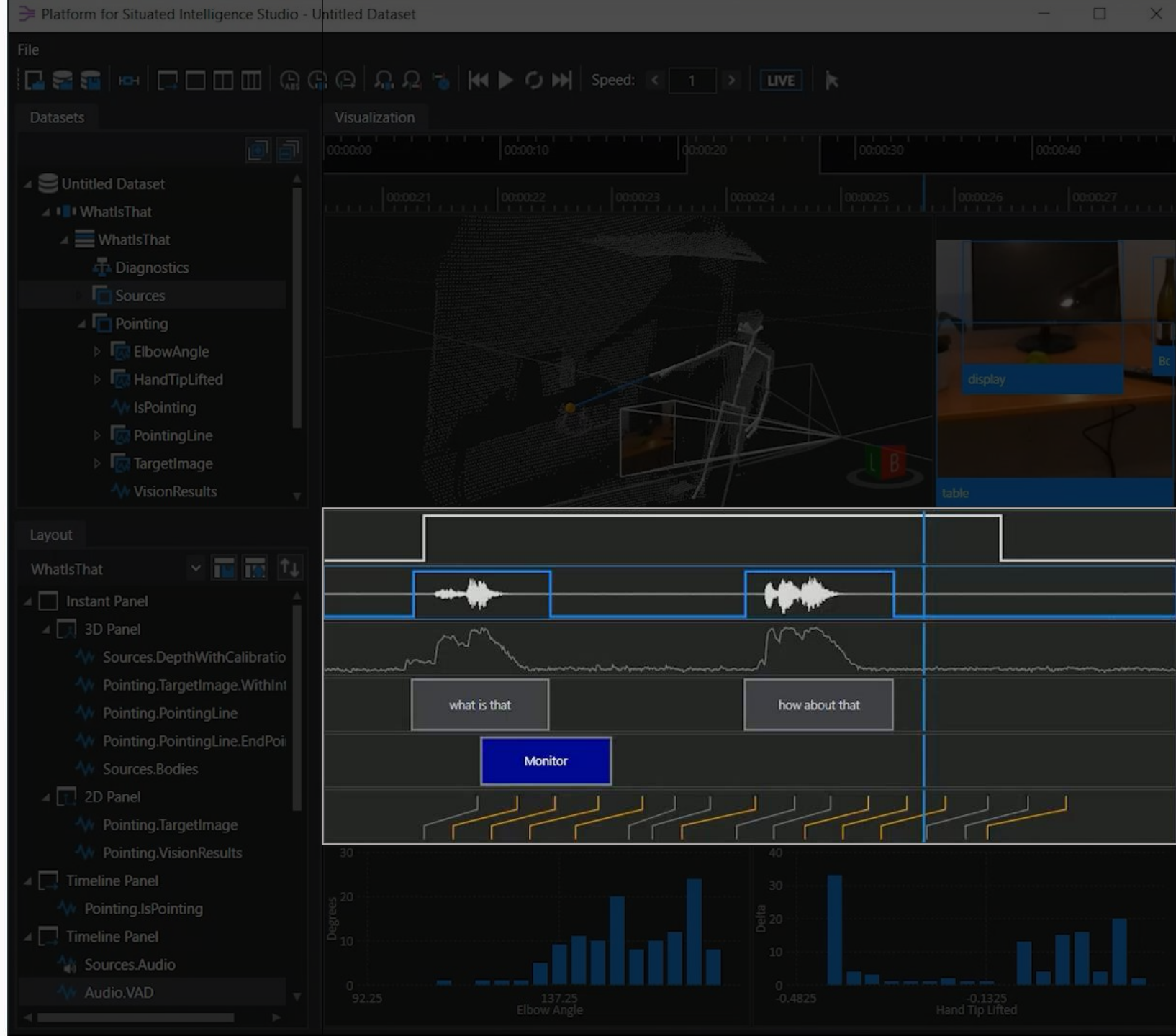
TOOLS

Data Visualization



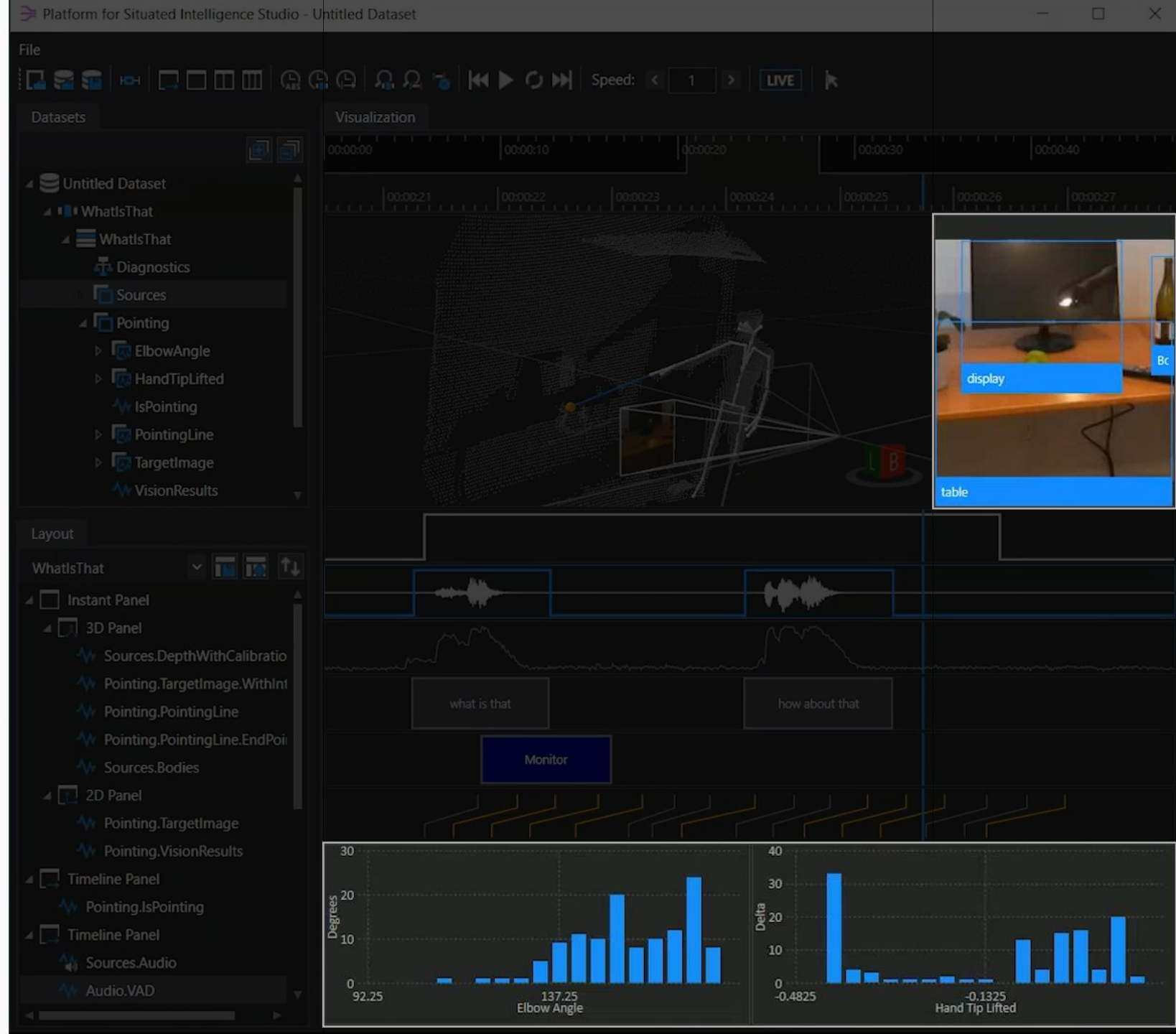
TOOLS

Data Visualization



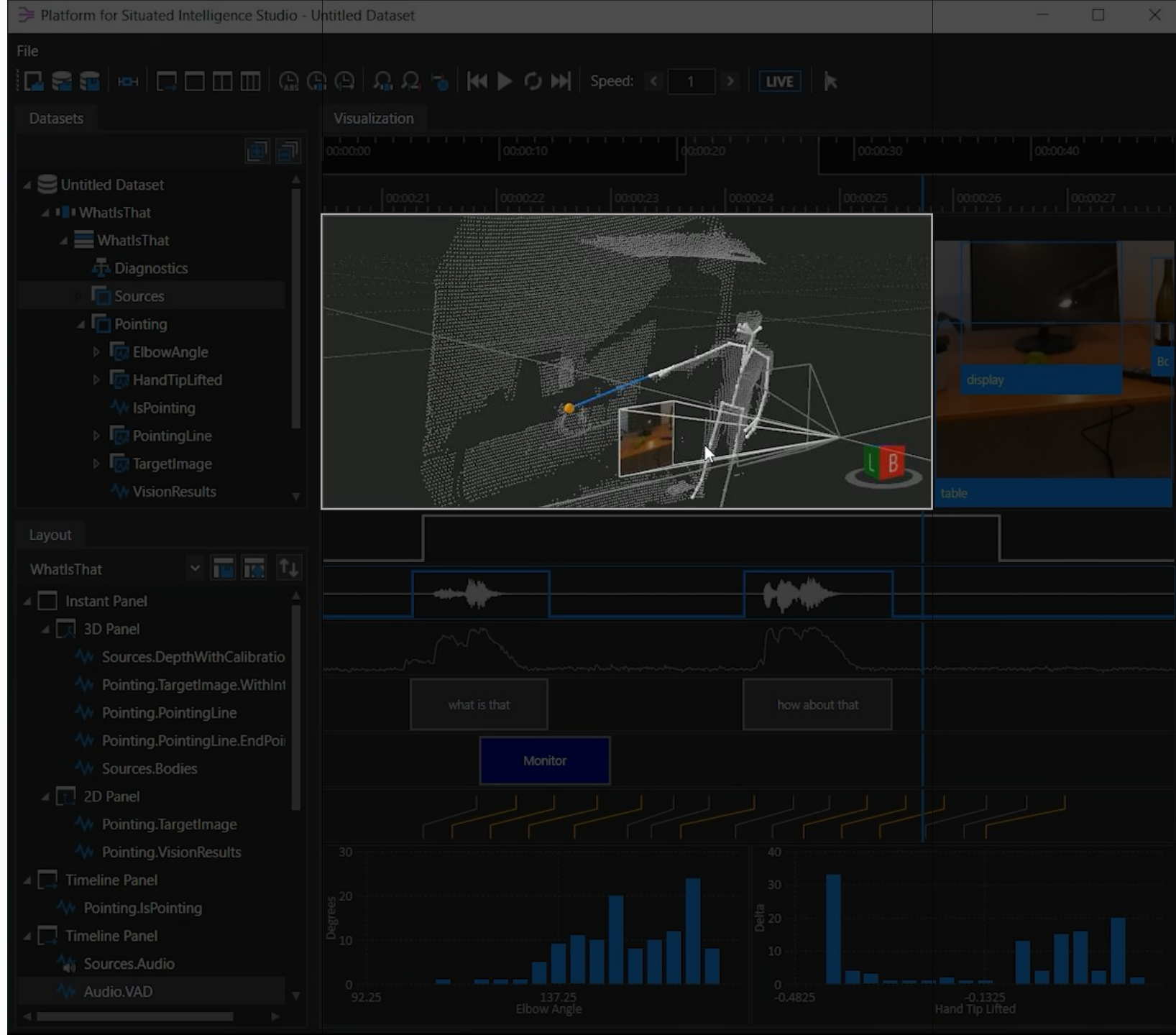
TOOLS

Data Visualization



TOOLS

Data Visualization



TOOLS

Data Visualization

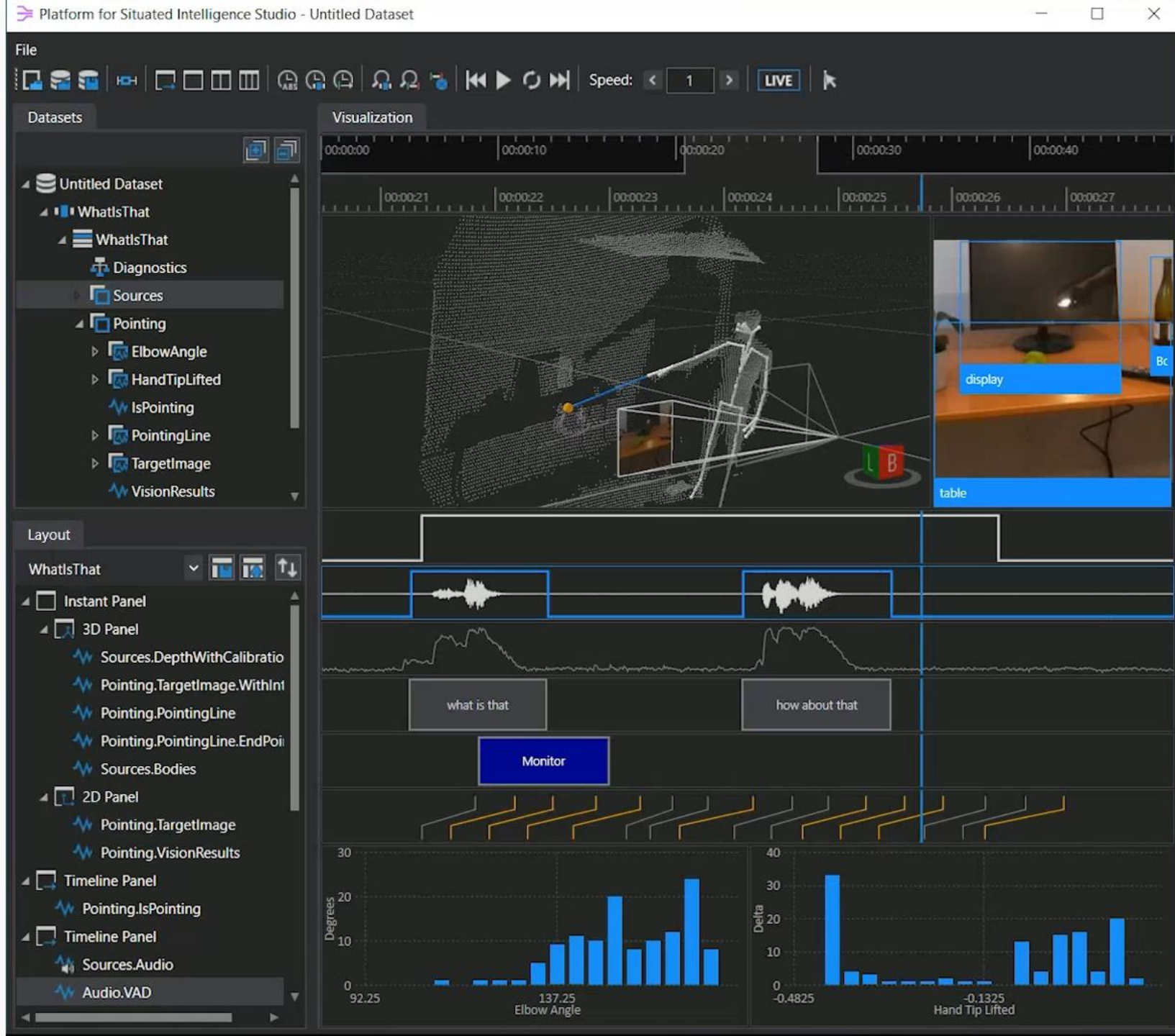
Navigation

Live

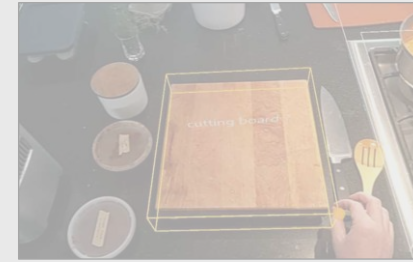
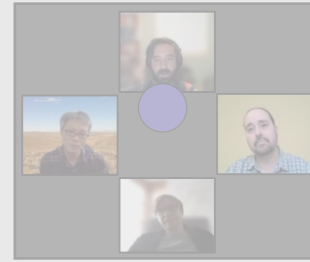
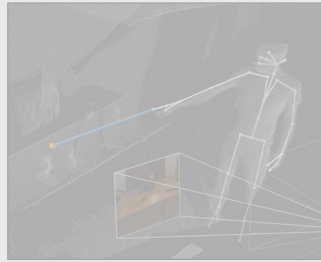
Annotations

Batch Processing

Etc.



Platform for Situated Intelligence



COMPONENTS

sensors | imaging | audio | vision | speech | language | ...

TOOLS

data visualization | debugging | annotation | processing

RUNTIME

streaming | logging | parallel coordinated computation

COMPONENTS

Sensors

Audio, Imaging

Speech, Vision

Language

ML / Onnx

Cloud Services

...

a growing ecosystem

INTEROP

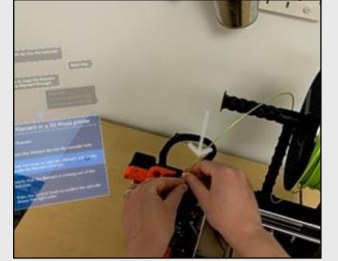
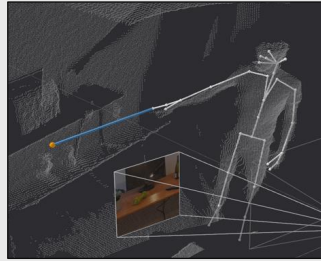
ROS

Python / Pytorch

Javascript

an open framework

Platform for Situated Intelligence



COMPONENTS

sensors | imaging | audio | vision | speech | language | ...

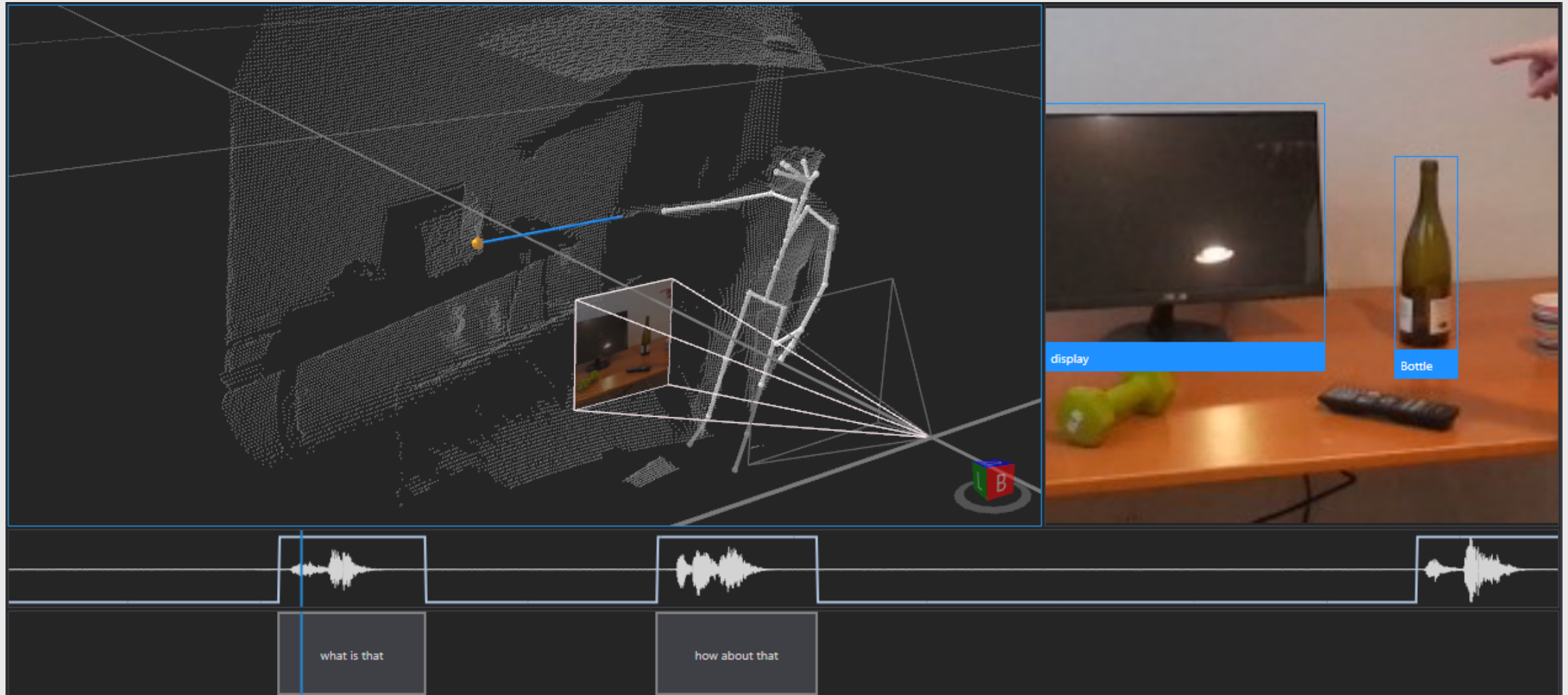
TOOLS

data visualization | debugging | annotation | processing

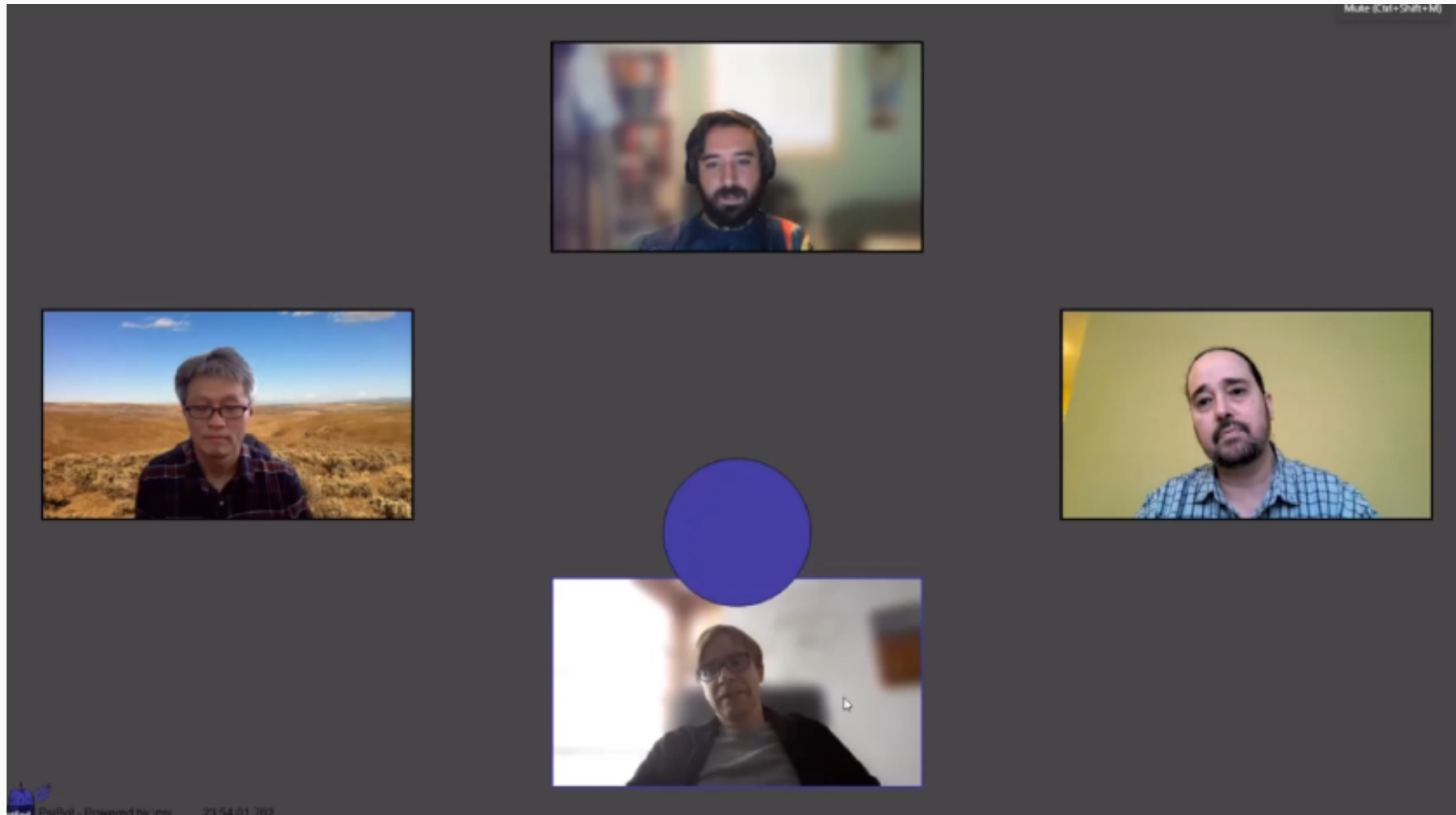
RUNTIME

streaming | logging | parallel coordinated computation

WHAT IS THAT? (SAMPLE APP)



BOT FOR MICROSOFT TEAMS



HOLOLENS 2

4 visible light cameras

2 infrared cameras

time-of-flight depth sensor

RGB camera

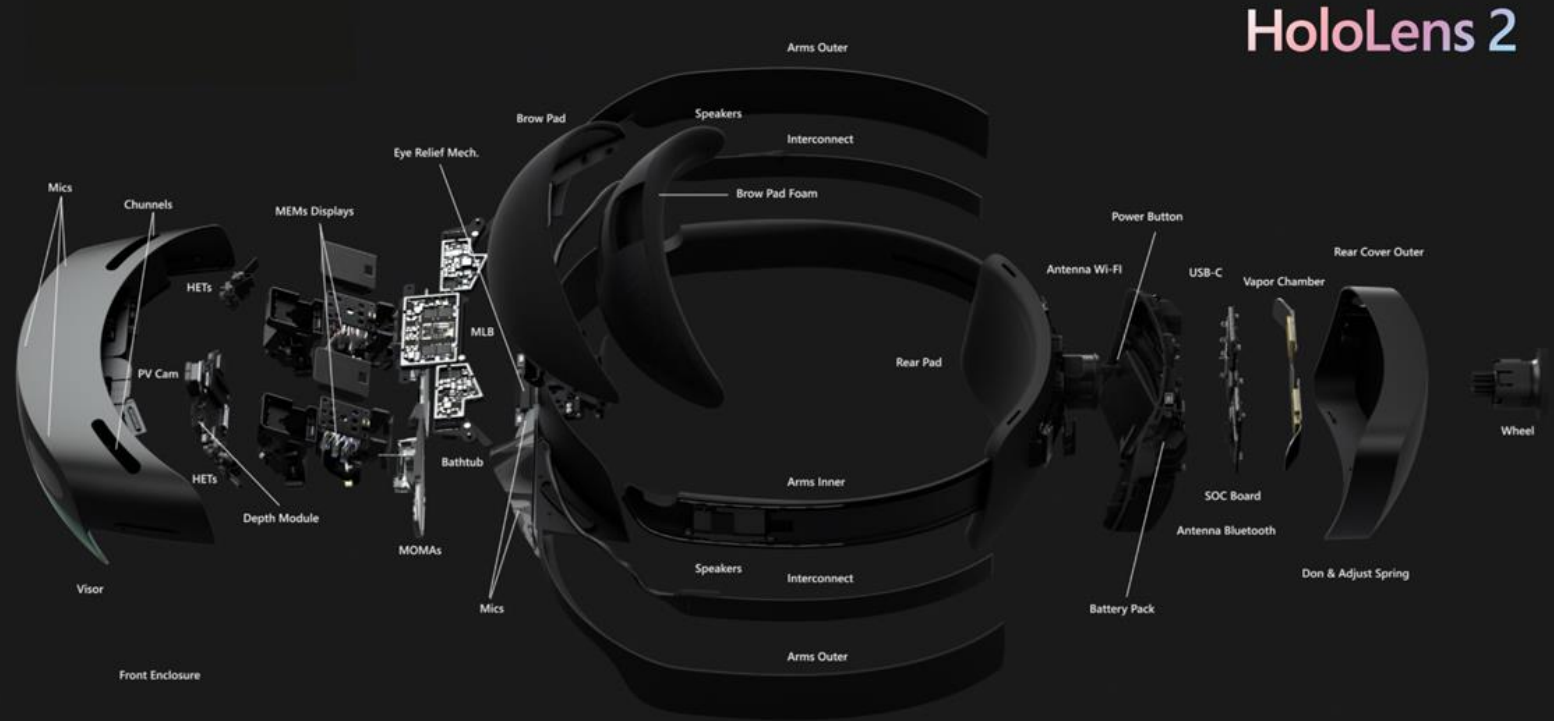
accelerometer

gyroscope

magnetometer

5-channel microphone array

spatial sound

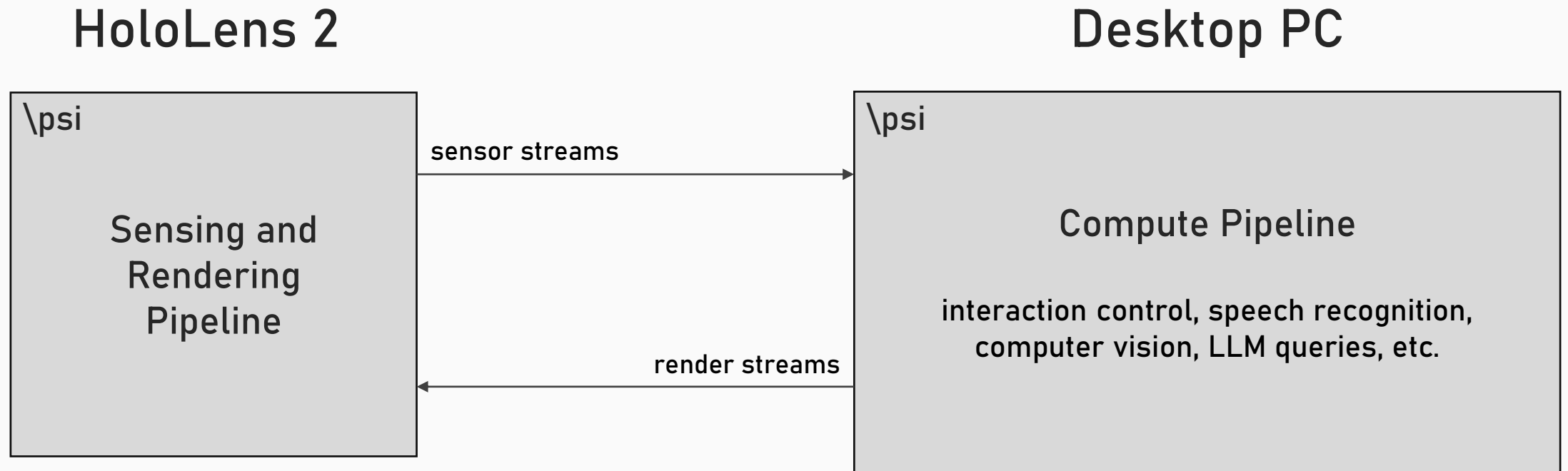


HoloLens 2

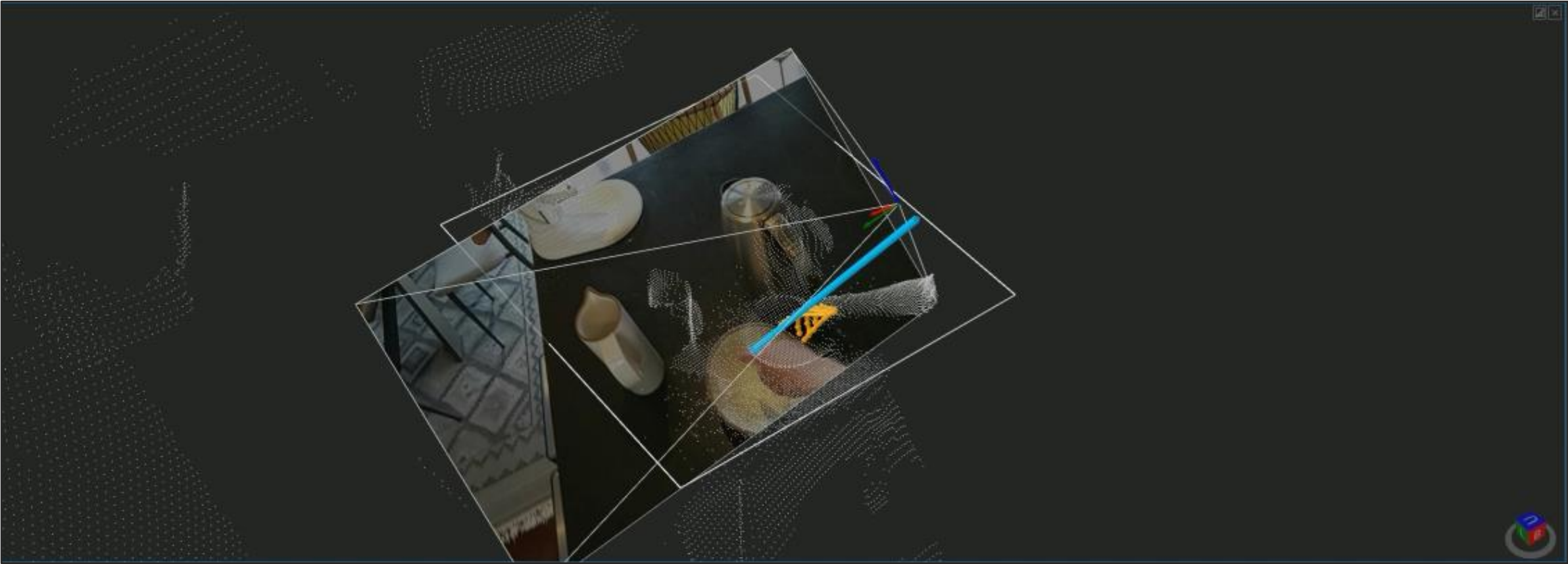


PSI HOLOLENS INTEGRATION

Client-Server architecture: HoloLens 2 with off-loaded compute



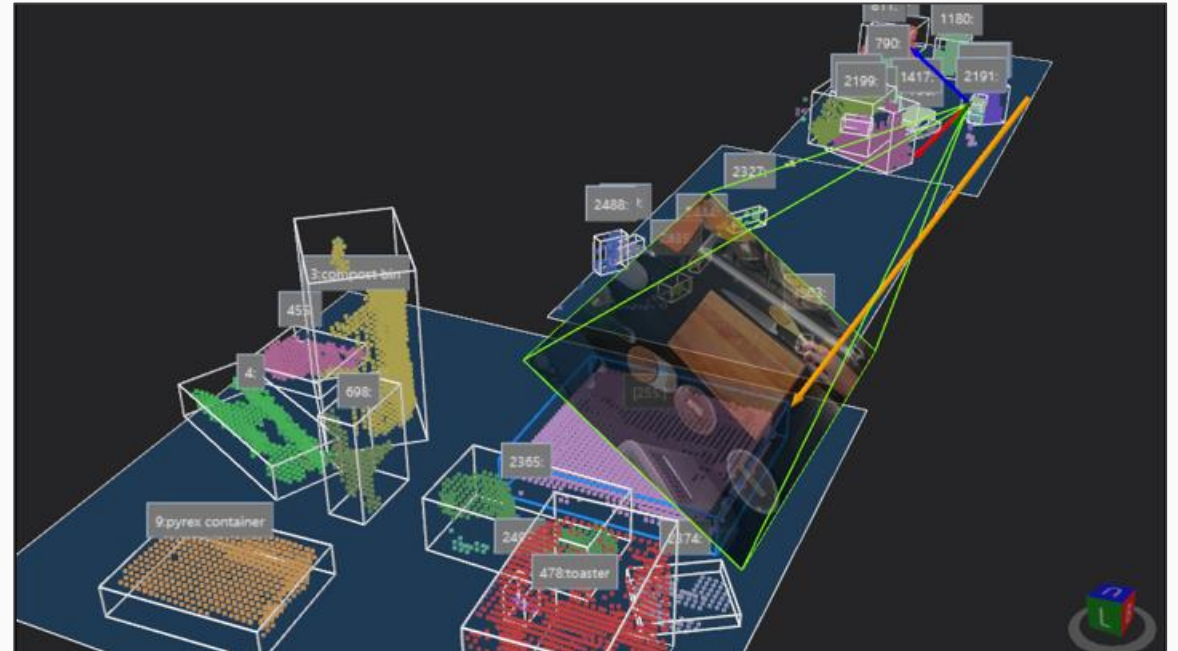
PSI HOLOLENS INTEGRATION



CONTINUAL LEARNING ABOUT OBJECTS

Bohus, Andrist, Feniello, Saw, and Horvitz. "Continual Learning about Objects in the Wild: an Interactive Approach", in Proc. of ICMI 2022

A multimodal, interactive approach for learning to recognize an open-ended set of objects online and onsite, in mixed reality



CONTINUAL LEARNING ABOUT OBJECTS

Bohus, Andrist, Feniello, Saw, and Horvitz. *"Continual Learning about Objects in the Wild: an Interactive Approach"*, in Proc. of ICMI 2022

Edit segmentations in stream



TASK ASSISTANCE IN THE PHYSICAL WORLD

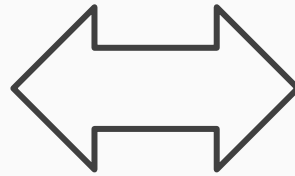
HOLOASSIST DATASET

Wang, Kwon, Rad, Pan, Chakraborty, Andrist, Bohus, Feniello, Tekin, Frujeri, Joshi, Pollefeys. *"HoloAssist: an Egocentric Human Interaction Dataset for Interactive AI Assistants in the Real World"*, in Proc. of ICCV 2023

An egocentric human interaction (task assistance) dataset



Worker

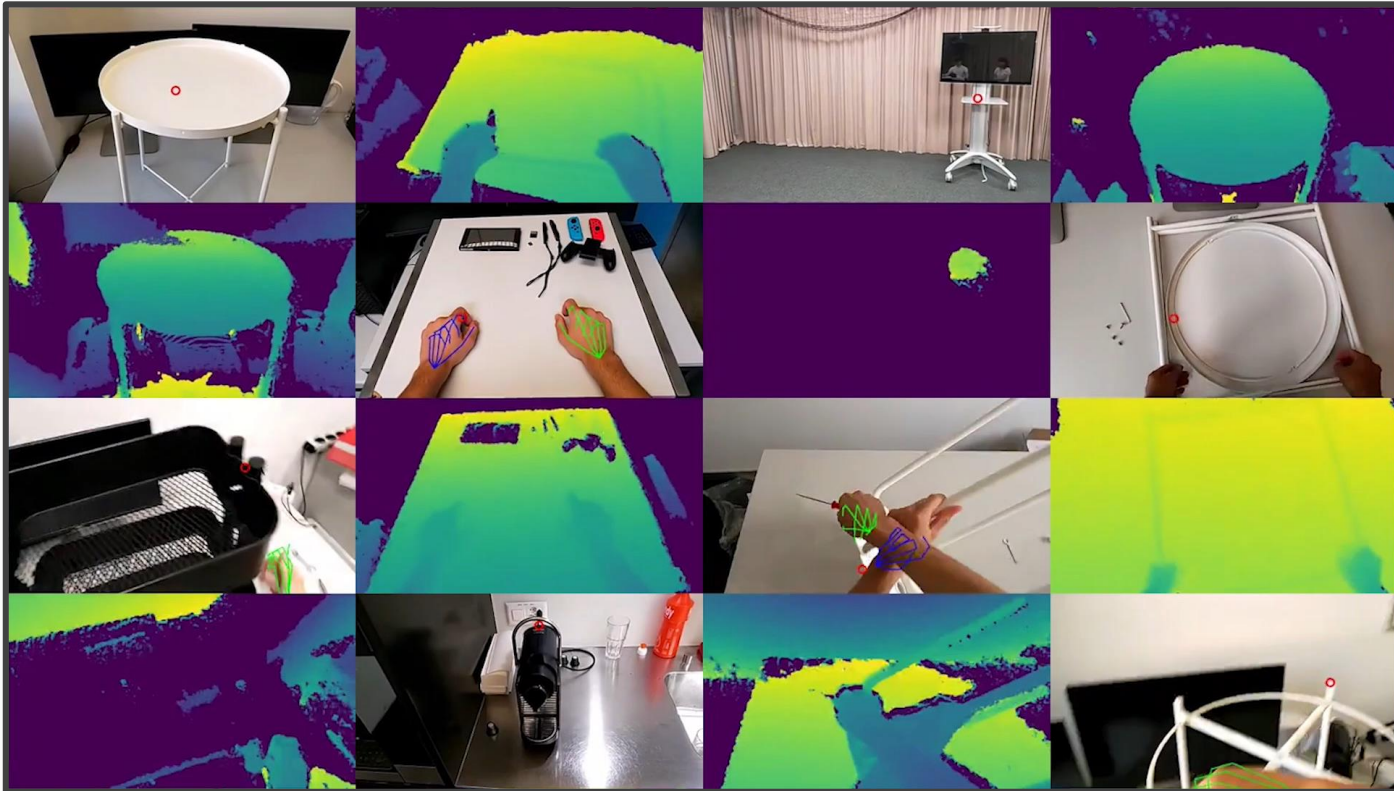


Instructor

HOLOASSIST DATASET

Wang, Kwon, Rad, Pan, Chakraborty, Andrist, Bohus, Feniello, Tekin, Frujeri, Joshi, Pollefeys. *"HoloAssist: an Egocentric Human Interaction Dataset for Interactive AI Assistants in the Real World"*, in Proc. of ICCV 2023

an egocentric human interaction (task assistance) dataset



166 hours

222 participants

350 instructor-worker pairs

20 object-centric tasks

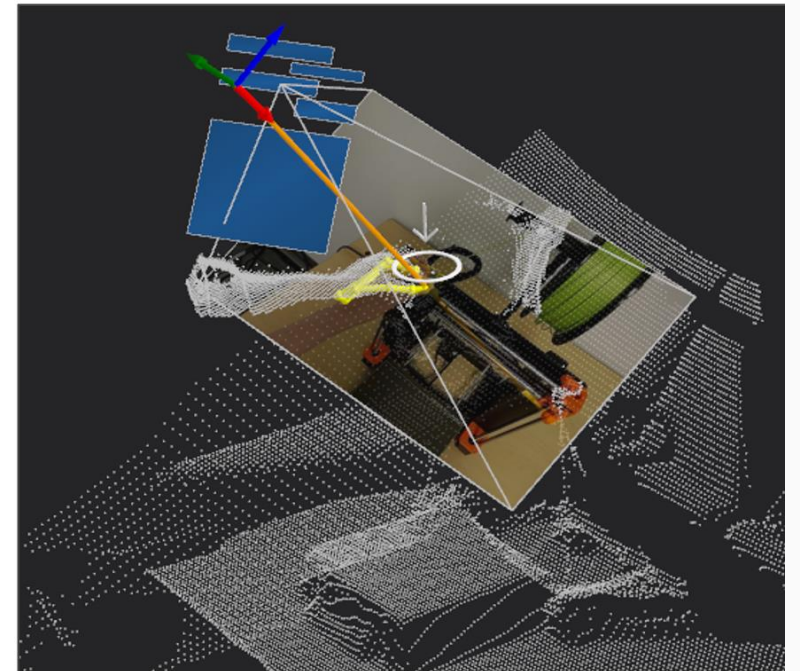
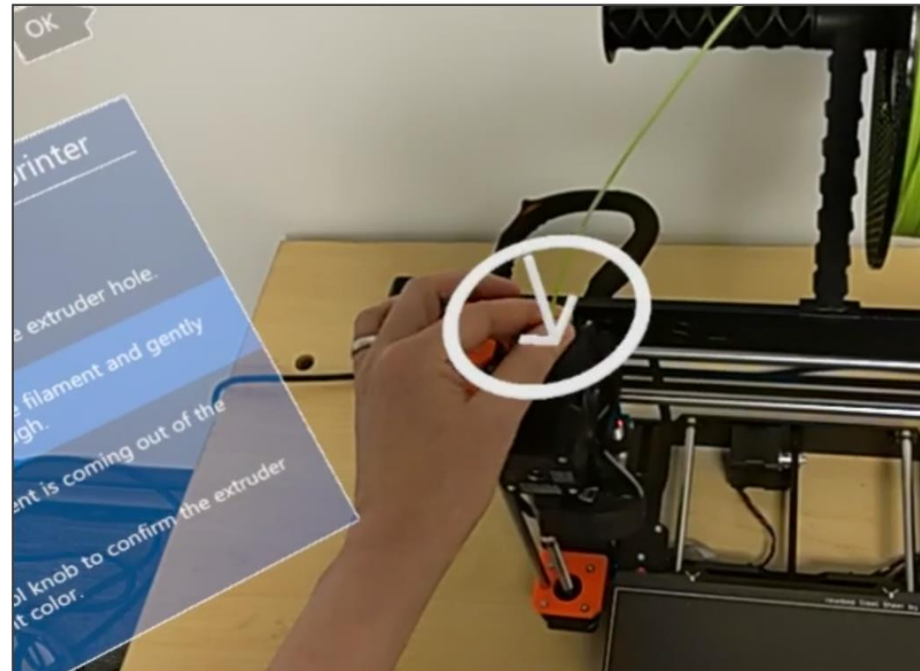
multiple modalities

rich temporal annotations

SIGMA

Bohus, Andrist, Paradiso, Saw, Rad, Chakraborty. "SIGMA: An Open-Source Interactive System for Mixed-Reality Task Assistance Research", in Proc. of IEEE VR VRW 2024

A research testbed for mixed-reality task assistance research
Build on \psi, open, extensible
Large language and vision models



Now that you have everything we need, let's move on to the next step.

The next step is to start-up the machine

First, power-on the machine by pressing the power button on the back right side.

I'm done.
Let's abandon this task.

Load the filament in a 3D Prusa printer

1 Objects:

- scissors
- filament spool

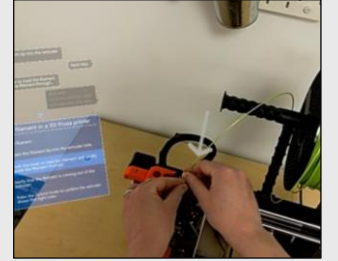
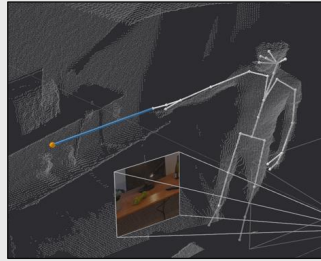
2 Start-up the machine

- 1 Power-on the machine by pressing the power button on the back right side.

3 Prepare the new filament

The next step is to start up the machine.

Platform for Situated Intelligence



COMPONENTS

sensors | imaging | audio | vision | speech | language | ...

TOOLS

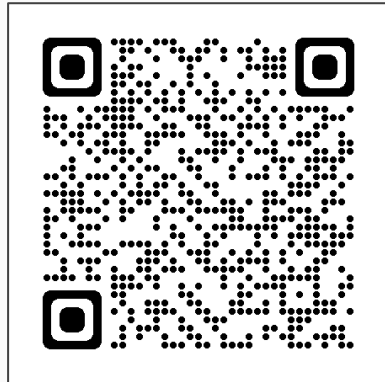
data visualization | debugging | annotation | processing

RUNTIME

streaming | logging | parallel coordinated computation

GET INVOLVED

Use it
github.com/Microsoft/psi



Report issues
Contribute
Tell others

A screenshot of the GitHub repository page for Microsoft/psi. The browser address bar shows the URL https://github.com/microsoft/psi/. The repository name is 'microsoft / psi' and it is public. The page shows navigation tabs for Code, Issues (47), Pull requests, Discussions, Actions, Wiki, Security, Insights, and Settings. Below the repository name, there are statistics: Edit Pins, Unwatch (37), Fork (90), and Starred (504). The main content area displays the README for 'Platform for Situated Intelligence'. It includes a description of the framework, a list of features provided, and a diagram of the infrastructure for logging. The right sidebar contains an 'About' section with a list of repository details and a 'Releases' section with 19 tags and a 'Packages' section with no published packages. The bottom of the sidebar shows 19 contributors.