Demonstrating a Framework for Rapid Development of Physically Situated Interactive Systems

Sean Andrist
Microsoft Research
Redmond, WA, USA
sandrist@microsoft.com

Dan Bohus
Microsoft Research
Redmond, WA, USA
dbohus@microsoft.com

Ashley Feniello
Microsoft Research
Redmond, WA, USA
ashleyf@microsoft.com

***Abstract*—We demonstrate an open, extensible framework for enabling faster development and study of physically situated interactive systems. The framework provides a programming model for parallel coordinated computation centered on temporal streams of data, a set of tools for data visualization and processing, and an open ecosystem of components. The demonstration showcases an interaction toolkit of components for systems that interact with people via natural language in the open world.**

***Index Terms*—Situated Interaction; Tools; Multiparty Dialog; Multimodal Systems; Integrative AI**

Research in human-robot interaction is often hindered by the large engineering efforts required to develop end-to-end systems that can operate robustly in the open world. These systems often integrate a wide array of technologies, from sensors, such as cameras and microphones, to perceptual components such as speech recognition, language understanding, person detection and tracking, *etc.*, to higher level inference models that enable reasoning about attention, conversational engagement, turn taking, and more. Similarly, the rendering of output actions on various channels, such as speech and gesture, needs to be coordinated. Decisions must be typically made under uncertainty and latency constraints. The engineering challenges that arise make it none too surprising that a great deal of HRI research is conducted via Wizard-of-Oz studies, and that real-world, in-the-wild deployments are still very rare.

In this demonstration, we will showcase Platform for Situated Intelligence [1], [2] – an open-source, extensible framework that aims to alleviate these engineering challenges and enable the rapid development of physically situated interactive systems (see more info also at https://aka.ms/psi-b1). The platform subsumes a *runtime*, which provides a programming model for parallel coordinated computation centered on temporal streams of data, a set of *tools* for data visualization (see Figure 1) and processing, and an open ecosystem of *pluggable components*. The framework allows for easy integration with existing ecosystems such as ROS [3] and software implemented in other languages like Python.

The demo will focus on an interaction toolkit, *i.e.*, a set of components currently under development on this platform that simplifies the construction of physically situated interactive applications. Developers can specify the available sensors and their relative locations and a situated perception pipeline is configured. This pipeline provides a continuous understanding of what is happening in the interactive scene: 3D positions and poses of the participants and agents, who is engaged in the conversation, where is everyone's attention, who is talking to whom, what they are saying, etc. By managing perception and lower level engineering challenges such as asynchronous but coordinated execution, stream synchronization and fusion, *etc.*, the toolkit enables researchers and developers to focus on the higher level interaction control logic.

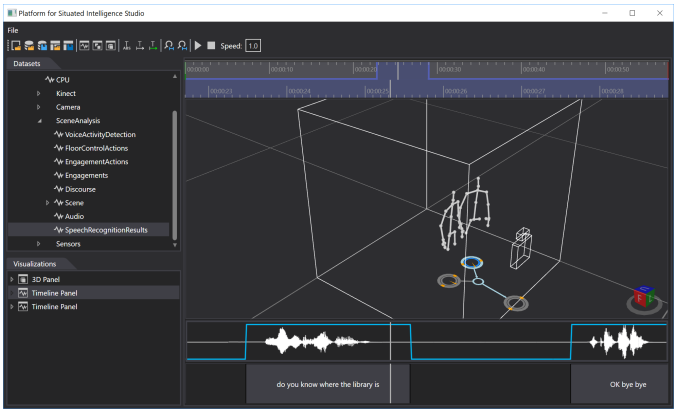

Fig. 1. The platform's visualization tool displaying persisted data and a 3D visualization of scene understanding in a multi-party human-agent interaction.

Our table demo will involve setting up and visualizing live (see Figure 1) a simple multi-party human-robot dialog application with coordinated speech and gaze, leveraging multiple sensors (*i.e.*, cameras, microphone). We will emphasize the small amount of code required to build such an application and further showcase platform principles, and the tools used for debugging and visualization. With this demonstration, we invite the HRI community to use, contribute to, and further develop this platform's ecosystem.

References

[1] D. Bohus, S. Andrist, and M. Jalobeanu, "Rapid development of multimodal interactive systems: A demonstration of platform for situated intelligence," in *Proc. of the 19th ACM Intl. Conf. on Multimodal Interaction (ICMI'17)*, Glasgow, UK, 2017.

[2] Microsoft. (2018) Platform for situated intelligence github page. [Online]. Available: https://github.com/Microsoft/psi

[3] M. Quigley, B. Gerkey, K. Conley, J. Faust, T. Foote, J. Leibs, E. Berger, R. Wheeler, and A. Ng, "Ros: an open-source robot operating system," in *Proc. of the IEEE Intl. Conf. on Robotics and Automation (ICRA'09) Workshop on Open Source Robotics*, Kobe, Japan, May 2009.