Can robots have social intelligence?

April 9 2014

Monica, a virtual assistant, greets Eric Horvitz outside his office.

Visiting Eric Horvitz at Microsoft Research headquarters in Redmond, Wash., is a full-service experience.

At the building's first-floor bank of elevators, situated on a busy corridor, a "smart elevator" opens its doors, sensing that you need a ride.

When you arrive on the third floor, a cute, humanoid robot makes eye contact, senses your interest in getting assistance, and cheerfully asks, "Do you need directions?" You tell him you're looking for Eric, and the robot gives you the office number and directions, gesturing with his

arms and hands to explain the turns and twists of the path to Eric's office.

"By the way," the robot adds, "I'll let Monica, his virtual assistant, know that you're on your way."

Down a couple of hallways, Monica is waiting for you, right outside Horvitz's office. She's an onscreen avatar with a British accent.

"I was expecting you," she says. "The robot told me you were coming."

Horvitz is in his office working on his computer, but Monica reflects about the costs and benefits of an unplanned interruption—considering nuances of Horvitz's current desktop activity, calendar, and past behavior. Using a model learned from this trove of data, the system determines that a short interruption would be OK and suggests that you go on in.

These experiences are all part of the Situated Interaction project, a research effort co-led by Horvitz, a Microsoft distinguished scientist and managing director of Microsoft Research Redmond, and his colleague Dan Bohus. They are creating a code base—an architecture—that enables many forms of complex, layered interaction between machines and humans.

The interaction is "situated" in the sense that the machine has situational awareness—it can take into account the physical optics of the space, the geometry of people's comings and goings, their gestures and facial expressions, and the give and take of conversation between multiple individuals—among other factors that most humans consider second nature.

"When it comes to language-based interaction, people immediately think

'speech recognition,'" Bohus says, "but it is much more than that."

The project relies on intensive integration of multiple computational competencies and methods, including machine vision, natural language processing, machine learning, automated planning, speech recognition, acoustical analysis, and sociolinguistics. It also pushes into a new area of research: how to automate processes and systems that understand multiparty interaction.

"For years," Horvitz says, "we've both been passionate about the question of how people, when they talk to each other in groups or in one-on-one settings, do so well to coordinate both meaning and turn taking—often building a mutual understanding of one another."

This includes discerning where attention is focused and understanding the differing roles of the people engaged in the conversation.

"We're addressing core challenges in artificial intelligence," Horvitz says. "The goal is to build systems that can coordinate and collaborate with people in a fluid, natural manner."
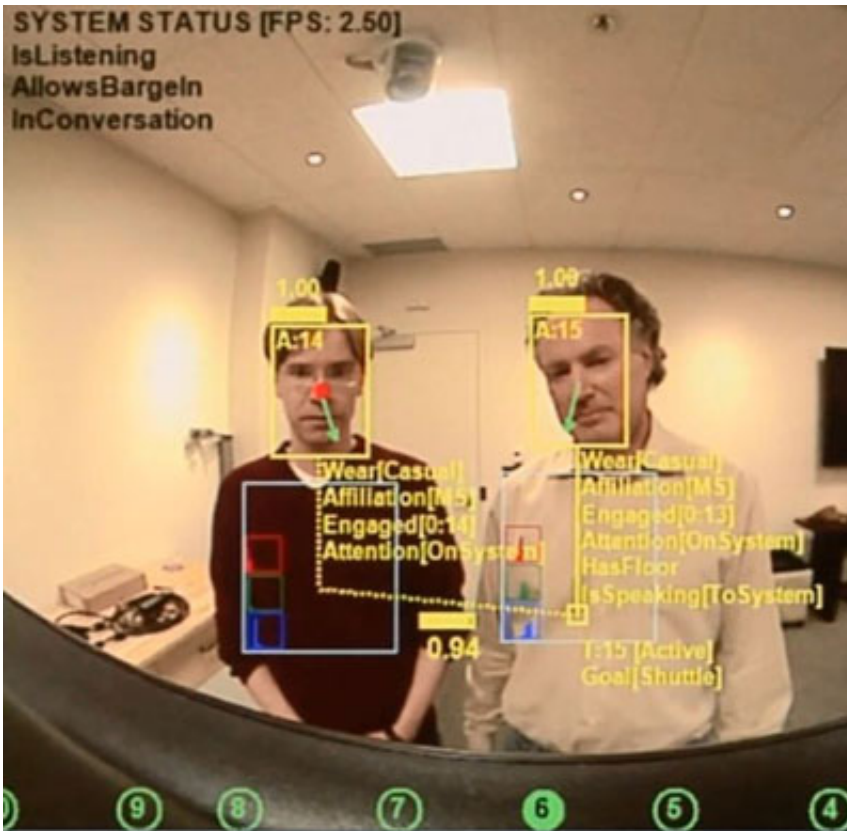
The capture system for the Situated Interaction project.

Enhancing Realism via Understanding Shared Context

One direction in making interactions with computers more natural is incorporating an awareness of time and space. This work includes efforts to endow systems with notions of short-term and long-term memory, and to have access to histories of experience similar to the abilities of people to recall.

Horvitz recalls a moment a couple of years back when he returned to his office from an all-day meeting and his virtual assistant said, "Hi, Eric, long time, no see!" He knew immediately that research intern, Stephanie Rosenthal, then a graduate student at Carnegie Mellon University, had

successfully integrated humanlike short- and long-term memory into the system that day.

"The interactions with the system, especially when we integrate new competencies, seem magical," Horvitz says, "with the magic emerging from a rich symphony of perception, reasoning, and decision-making that's running all the time."

The evolving platform has continued to fuel new research directions. As the system's overall level of competency and naturalness increases, previously unrecognized gaps in ability become salient—leading to new directions of research, such as the work on short- and long-term memory.

"As you increase the degree of realism," Bohus says, "and you get to something that acts and looks more human, it becomes uncanny when things that would be deficits in a human being show up."

Horvitz's virtual assistant has a rich set of skills that includes the ability to access his online calendar, detect his presence and absence in the office, infer how busy he is, predict when he'll finish a certain task based on his past habits, predict when he'll return to his office and when he'll next read email, and even predict when he'll conclude a conversation, based on the length of his past conversations.

The smart elevator, which uses a wide-angle camera attached to the ceiling to perceive the positions and movements of passers-by, takes the data and calculates the probability that someone needs a ride and isn't just walking past to get to the cafeteria or the building's atrium.

The directions robot makes real-time inferences about the engagement, actions, and intentions of the people who approach it, with the aid of multiple sensors, a wide-angle camera, the Kinect for Windows speech-

recognition engine, and knowledge of the building and the people who work there. The robot can direct people to anyone and any place in the building, using the building directory and the building map.

"Each of these projects," Bohus notes, "pushes the boundaries in different directions."

Other Situated Interaction projects have included a virtual receptionist that schedules shuttle rides between buildings on the Microsoft campus. Like Monica and the directions robot, it can detect multiple people, infer a range of things about them, and conduct a helpful interaction. All of the systems can be refined continually based on different streams of collected data.

All of the Situated Interaction work, Bohus adds, has benefited from contributions by multiple researchers and engineers at Microsoft Research. Horvitz agrees.

The directions robot, part of the Situated Interaction project, is a welcoming, informative presence at Microsoft Research headquarters.

"Beyond passionate researchers who have pursued these core challenges for decades," he says, "the projects rely on the contributions and creativity of a team of engineers."

Among those on the Situated Interaction team are Anne Loomis Thompson and Nick Saw.

The Next Frontiers of Situated Interaction

The array of possible applications for Situated Interaction technology is vast, from aerospace and medicine to education and disaster relief.

The magical capabilities of the systems developed at Microsoft Research provide a technical vision and long-term directions for what Cortana aspires to deliver. A personal assistant for the recently released Windows Phone 8.1, Cortana uses natural language, machine learning, and contextual signals in providing assistance to users.

Horvitz anticipates vigorous competition among technology companies to create personal assistants that operate across a person's various devices—and provide a unified, supportive experience across all of them over time and space. He sees this competition as fueling progress in artificial intelligence.

"It is going to be a cauldron of creativity and competition," he says. "Who's going to produce the best digital assistant?"

Such an assistant could coordinate with the assistants of other people, helping to schedule social engagements, work commitments, and travel. It could anticipate your needs based on past activities—such as where you have enjoyed dining—and coordinate with businesses that offer

special deals. It could help you select a movie based on which ones your friends liked.

"Intelligent, supportive assistants that assist and complement people are a key aspiration in computer science," Horvitz says, "and basic research in this space includes gathering data and observing people conversing, collaborating, and assisting one another so we can learn how to best develop systems that can serve in this role."

Horvitz also envisions deeper forms of collaboration between computers and people in the workplace. As you create presentations or documents, for example, computers could make suggestions about relevant graphical and informational content in a non-invasive manner. The interaction would be a fluid exchange, with both parties sharing and responding to ideas and making inferences about mutual goals.

Horvitz believes that we are in "the early dawn of a computational revolution." As he said in a recent TEDx talk, "[M]y view is that these methods will come to surround people in a very supportive and nurturing way. The intellects we're developing will complement our own human intellect and empower us to do new kinds of things and extend us in new kinds of ways according to our preferences and our interests."

Provided by Microsoft