

Using AI, people who are blind are able to find familiar faces in a room

January 30 2020, by John Roach



Theo, left, a 12-year-old boy who is blind, interacts with Microsoft senior research software development engineers Tim Regan, middle, and Martin Grayson, right, during user testing of the Project Tokyo system. Credit: Jonathan Banks

Theo, a 12-year-old boy who is blind, is seated at a table in a crowded

kitchen on a gray and drippy mid-December day. A headband that houses cameras, a depth sensor and speakers rings his sandy-brown hair. He swivels his head left and right until the camera in the front of the headband points at the nose of a person on the far side of a counter.

Theo hears a bump sound followed by the name "Martin" through the headband's speakers, which are positioned above his ears.

"It took me like five seconds to get you, Martin," Theo says, his head and body fixed in the direction of [Martin Grayson](#), a senior research software development engineer with Microsoft's research lab in Cambridge. Grayson stands next to a knee-high black chest that contains computing hardware required to run the machine learning models that power the prototype system Theo used to recognize him.

Elin, Theo's mother, who is standing against a wall on the opposite side of Theo, says, "I love the way you turned around to find him. It is so nice."

As Theo begins to turn to face his mother, the speakers sound another bump and the name "Tim."

"Tim, there you are," says Theo with delight as his gaze lands on [Tim Regan](#), another senior research software development engineer at the lab, who took Theo under his wing to teach him advanced computer coding skills. Theo and his mother were at Regan's house for a bi-monthly coding lesson. They met while working on a research project that led to the [development of Code Jumper](#), a physical programming language that's inclusive of children with all ranges of vision.

Theo is now one of several members of the blind and low-vision community who are working with Regan, Grayson, researcher [Cecily Morrison](#) and her team on Project Tokyo, a multipronged research effort

to create intelligent personal agent technology that uses artificial intelligence to extend people's existing capabilities.

For Theo, that means tools to recognize who is around him.

"It is so exciting to be able to find out where the people are in my environment," Theo said. "Not just who chooses to talk, but all of the people who are silent that you can see by their face, but I can't."

But ultimately, noted Morrison, Project Tokyo is a research effort with a long-term goal of demonstrating how to build intelligent personal agents that extend the capabilities of all users. Rather than building end-to-end systems that can accomplish specific tasks, she sees the future of AI as a set of resources that people use in whatever way they see fit.

"All of a sudden we don't have to say, 'Hey you are blind and I just made this accessible to you.' We say, 'Hey, you are you and I have just built a system that works for you,'" she said. "I don't need to know anything about you. I don't need a label on you. I can make something that is right for you because I have a system that you can take and adapt to yourself."

Paralympics in Brazil

Project Tokyo was born out of a challenge, in early 2016, from senior leaders at Microsoft to create AI systems that would go beyond completing tasks such as fetching sports scores and weather forecasts or identifying objects. Morrison said creating tools for people who are blind and with low vision was a natural fit for the project, because people with disabilities are often early adopters of new technology.

"It is not about saying, 'Let's build something for [blind people](#),'" Morrison said. "We are working with blind people to help us imagine the future, and that future is about new experiences with AI."

Morrison and her colleague [Ed Cutrell](#), a senior principal researcher at Microsoft's research lab in Redmond, Washington, were tapped to lead the project. Both have expertise in designing technologies with people who are blind or with low vision and decided to begin by trying to understand how an agent technology could augment, or extend, the capabilities of these users.



Ed Cutrell, a senior principal researcher with Microsoft's research organization in Redmond, Washington, is a co-leader of Project Tokyo. On his desk are several modified Microsoft HoloLenses that project researchers are using to help people who are blind and with low vision learn who is where in their social environment. Credit: Dan DeLong

To start, they followed a group of athletes and spectators with varying levels of vision on a trip from the United Kingdom to the 2016 Paralympic Games in Rio de Janeiro, Brazil, observing how they interacted with other people as they navigated airports, attended sporting venues and went sightseeing, among other activities. A key learning, noted Cutrell, was how an enriched understanding of social context could help people who are blind or with low vision make sense of their environment.

"We, as humans, have this very, very nuanced and elaborate sense of social understanding of how to interact with people—getting a sense of who is in the room, what are they doing, what is their relationship to me, how do I understand if they are relevant for me or not," he said. "And for blind people a lot of the cues that we take for granted just go away."

This understanding spurred a series of workshops with the blind and low vision community that were focused on potential technologies that could provide such an experience. Peter Boshier, an audio engineer in his mid-50s who has been blind most of his life and worked with the Project Tokyo team, said the concept of a technology that provided information about the people around him resonated immediately.

"Whenever I am in a situation with more than two or three people, especially if I don't know some of them, it becomes exponentially more difficult to deal with because people use more and more eye contact and body language to signal that they want to talk to such-and-such a person, that they want to speak now," he said. "It is really very difficult as a blind person."

A modified HoloLens

Once the Project Tokyo researchers understood the type of AI experience they wanted to create, they set out to build the enabling

technology. They started with the original [Microsoft HoloLens](#), a mixed reality headset that projects holograms into the real world that users can manipulate.

"HoloLens gives us a ton of what we need to build a real time AI agent that can communicate the social environment," said Grayson during a demonstration of the technology at Microsoft's research lab in Cambridge.

For example, the device has an array of grayscale cameras that provide a near 180-degree view of the environment and a high-resolution color camera for high-accuracy facial recognition. In addition, the speakers above the user's ears allow for spatialized audio—the creation of sounds that seem to be coming from specific locations around the user.

Machine learning experts on the Project Tokyo team then developed computer vision algorithms that provide varying levels of information about who is where in the user's environment. The models run on graphical processing units, known as GPUs, that are housed in the black chest that Grayson carted off to Regan's house for the user testing with Theo.

One model, for example, detects the pose of people in the environment, which provides a sense of where and how far away people are from the user. Another analyzes the stream of photos from the high-resolution camera to recognize people and determine if they have opted to make their names known to the system. All this information is relayed to the user through audio cues.

For example, if the device detects a person one meter away on the user's left side, the system will play a click that sounds like it is coming from one meter away on the left. If the system recognizes the person's face, it will play a bump sound, and if that person is also known to the system, it

will announce their name.

When the user only hears a click but wants to know who the person is, a second layer of sound that resembles an elastic band stretching guides the user's gaze toward the person's face. When the lens' central camera connects with the person's nose, the user hears a high-pitched click and, if the person is known to the system, their name.

"I particularly like the thing that gives you the angle of gaze because I'm never really sure what is the sensible angle for your head to be at," said Boshier, who worked with the Project Tokyo team on the audio experience early in the design process and returned to the Cambridge lab to discuss his experience and check out the latest iteration. "That would be a great tool for learning body language."

Prototyping with adults

As the Project Tokyo team has developed and evolved the technology, the researchers routinely invite adults who are blind or with low vision to test the system and provide feedback. To facilitate more direct social interaction, for example, the team removed the lenses from the front of the HoloLens.



Peter Boshier, middle, an audio engineer who is blind who worked with the Project Tokyo team early in the design process, checks out the latest iteration of the system at Microsoft's research lab in Cambridge, UK, with researchers Martin Grayson, left, and Cecily Morrison, right. Credit: Jonathan Banks

Several users expressed a desire to unobtrusively get the information collected by the system without constantly turning their heads, which felt socially awkward. The feedback prompted the Project Tokyo team to work on features that help users quickly learn who is around them by, for example, asking for an overview and getting a spatial readout of all the names of people who have given permission to be recognized by the system.

Another experimental feature alerts the user with a spatialized chime

when someone is looking at them, because people with typical vision often establish eye contact to initiate a conversation. Unlike the bump, however, the chime is not followed by a name.

"We already use the name when you look at somebody," Grayson explained to Emily, a tester in her 20s who has low vision and visited the Cambridge lab to learn about the most recent features. "But also, by not giving the name, it might draw your attention to turn to somebody who is trying to get your attention. And by turning to them, you find out their name."

"I totally agree with that. That is how sighted people react. They capture someone out of the corner of their eye, or you get that sense, and go, 'Cecily,'" Emily said.

The modified HoloLens the researchers showed to Emily also included an LED strip affixed above the band of cameras. A white light tracks the person closest to the user and turns green when the person has been identified to the user. The feature lets communication partners or bystanders know they've been seen, making it more natural to initiate a conversation.

The LED strip also provides people an opportunity to move out of the device's field of view and not be seen, if they so choose. "When you know you are about to be seen, you can also decide not to be seen," noted Morrison. "If you know when you are being seen, you know when you are not being seen."

A tool for teaching social interaction skills

As the technical research continues, Project Tokyo is exploring an avenue revealed in the research process: using the technology to help children who are blind or with low vision develop social interaction

skills.

Two-thirds of children who are blind or with low vision exhibit social behaviors that are consistent with children who are on the autism spectrum, [according to academic research](#). For example, many children who are blind or with low vision appear disengaged from conversation partners, often resting their head on a table with an ear exposed.

Morrison and Cutrell pivoted Project Tokyo to explore whether a scaled-down version of the system could be used to help children who are blind or with low vision understand how they can use their bodies to initiate and maintain interactions with people.

Because the Microsoft researchers already had a relationship with Theo, they recruited him to help adapt the system to function with children, such as accounting for the tendency of children to sit close together and, at the same time, seldom sit still.

"When it was announcing people's names, it was trying to announce two names at once and I asked for that to be changed because, basically, it was very, very hard to hear anybody's name," Theo recalled.

The researchers also explored how Theo used the system. For example, during a family meal he started to subtly, but repeatedly, shift his head from side to side to force the system to read out the names of the people he was speaking to.

"We believe he was using that to support his spatial attention toward a person by refreshing his working memory of where they were," Morrison said. "That's something we could never have predicted, but a very powerful strategy for helping him maintain his attention, and if he can maintain his attention, he can maintain a topic of conversation."

Other uses of the technology were more in line with the researchers' hypothesis that it would help him build skills for socially interacting in a world dominated by people who are sighted.



Martin Grayson, left, a senior research software development engineer with Microsoft's research lab in Cambridge, UK, registers Microsoft researcher Cecily Morrison, right, in the Project Tokyo system for user testing. Credit: Jonathan Banks

For example, like other children who are blind or with low vision, Theo would put his head on the table during social situations, one ear cocked to the world. The researchers played a series of games with Theo designed to highlight the social power that could come when using his

body and head to engage in conversation with people who are sighted.

In a game played at the lab, the researchers had a group problem to solve. Theo knew the answer. The researchers only knew the topic and they could only talk when Theo looked at them. When Theo looked away, they had to stop talking.

"All of the sudden he realized he can manage a conversation," Morrison said. "He came to understand the power of being able to look at somebody, the power that gave him in a conversation and by that he's then enabled a whole new set of social capabilities that he hadn't been able to achieve before."

Today, Theo seldom speaks with his head on the table. Whether wearing the modified HoloLens or not, he turns his body and face toward the person he wants to engage. Whether the change will persist long term is unknown, nor are the researchers certain if other children who are blind or with low vision will respond similarly.

"From what we are seeing with Theo, we have a good feeling about it, because we have seen it with him, but that is a case of one. And who knows if that would have happened anyway," Cutrell said. "That is why we are spinning up to this next phase, which will be looking at considerably more children and a broader age range as well."

Future of Tokyo

The broader Project Tokyo research effort continues, including new directions in machine learning that allow users to adapt the system to their personal preferences. [Sebastian Tschitschek](#), a machine learning researcher at the Cambridge lab, is working on features that enable users to show the system the kind and amount of information they want to hear.

The development of personalization is requiring Tschatschek to take an unconventional approach to machine learning.

"What we like to do is formalize a problem in some mathematical form," he said. "You cannot do that so easily in this problem. Lots of the development comes through trying out things, having this interaction with people, seeing what they like, don't like, and enhancing the algorithms."

The desire for personalization, he explained, exists because people who are blind or with low vision have different levels of vision and thus different information needs. What's more, users of the system get frustrated when it provides information they already know.

"To get the vision of Project Tokyo done, you have to combine so many things that are not solved themselves," Tschatschek said.

Ultimately, Project Tokyo is about demonstrating how to build intelligent personal agents that extend the capabilities of all users. To get there, Morrison, Cutrell and their colleagues will continue to work with people who are blind or with low vision, including more children.

"What we saw with Theo is pretty powerful," Morrison said in her office the day after the system testing at Regan's house. "It was powerful because he was in control of his world in a way that he couldn't be before."

Among the expanding cohort of children to participate in Project Tokyo is Morrison's 7-year-old son, Ronan, who has been blind since birth.

"I think we are going to see that with Ronan," she added. "I'm super excited to try."

Provided by Microsoft

Citation: Using AI, people who are blind are able to find familiar faces in a room (2020, January 30) retrieved 25 April 2024 from <https://techxplore.com/news/2020-01-ai-people-familiar-room.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.