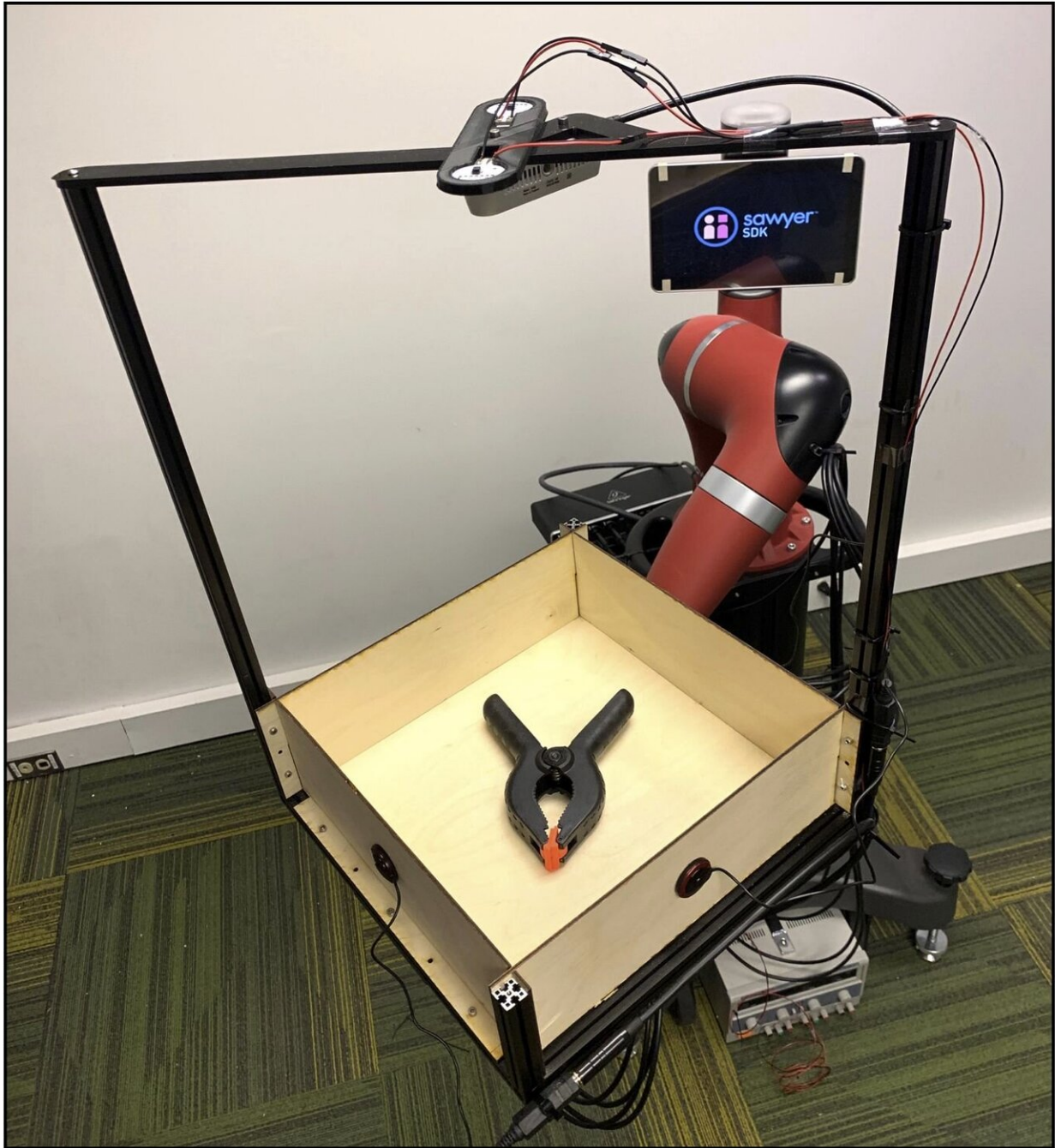# Sounds of action: Using ears, not just eyes, improves robot perception

August 14 2020

Carnegie Mellon University researchers devised an apparatus called Tilt-Bot to build a collection of actions, video and sound to improve robot perception. Objects were placed in a tray attached to a robot arm, which then moved the tray randomly while recording video and audio. Credit: Carnegie Mellon University

People rarely use just one sense to understand the world, but robots usually only rely on vision and, increasingly, touch. Carnegie Mellon University researchers find that robot perception could improve markedly by adding another sense: hearing.

In what they say is the first large-scale study of the interactions between sound and robotic action, researchers at CMU's Robotics Institute found that sounds could help a robot differentiate between objects, such as a metal screwdriver and a metal wrench. Hearing also could help robots determine what type of action caused a sound and help them use sounds to predict the physical properties of new objects.

"A lot of preliminary work in other fields indicated that sound could be useful, but it wasn't clear how useful it would be in robotics," said Lerrel Pinto, who recently earned his Ph.D. in robotics at CMU and will join the faculty of New York University this fall. He and his colleagues found the performance rate was quite high, with robots that used sound successfully classifying objects 76 percent of the time.

The results were so encouraging, he added, that it might prove useful to equip future robots with instrumented canes, enabling them to tap on objects they want to identify.

The researchers presented their findings last month during the virtual Robotics Science and Systems conference. Other team members included Abhinav Gupta, associate professor of robotics, and Dhiraj Gandhi, a former master's student who is now a research scientist at Facebook Artificial Intelligence Research's Pittsburgh lab.

To perform their study, the researchers created a large dataset, simultaneously recording video and audio of 60 common objects—such as toy blocks, hand tools, shoes, apples and tennis balls—as they slid or rolled around a tray and crashed into its sides. They have since released

this dataset, cataloging 15,000 interactions, for use by other researchers.

The team captured these interactions using an experimental apparatus they called Tilt-Bot—a square tray attached to the arm of a Sawyer robot. It was an efficient way to build a large dataset; they could place an object in the tray and let Sawyer spend a few hours moving the tray in random directions with varying levels of tilt as cameras and microphones recorded each action.

They also collected some data beyond the tray, using Sawyer to push objects on a surface.

Though the size of this dataset is unprecedented, other researchers have also studied how intelligent agents can glean information from sound. For instance, Oliver Kroemer, assistant professor of robotics, led research into using sound to estimate the amount of granular materials, such as rice or pasta, by shaking a container, or estimating the flow of those materials from a scoop.

Pinto said the usefulness of sound for robots was therefore not surprising, though he and the others were surprised at just how useful it proved to be. They found, for instance, that a robot could use what it learned about the sound of one set of objects to make predictions about the physical properties of previously unseen objects.

"I think what was really exciting was that when it failed, it would fail on things you expect it to fail on," he said. For instance, a robot couldn't use sound to tell the difference between a red block or a green block. "But if it was a different object, such as a block versus a cup, it could figure that out."

Provided by Carnegie Mellon University