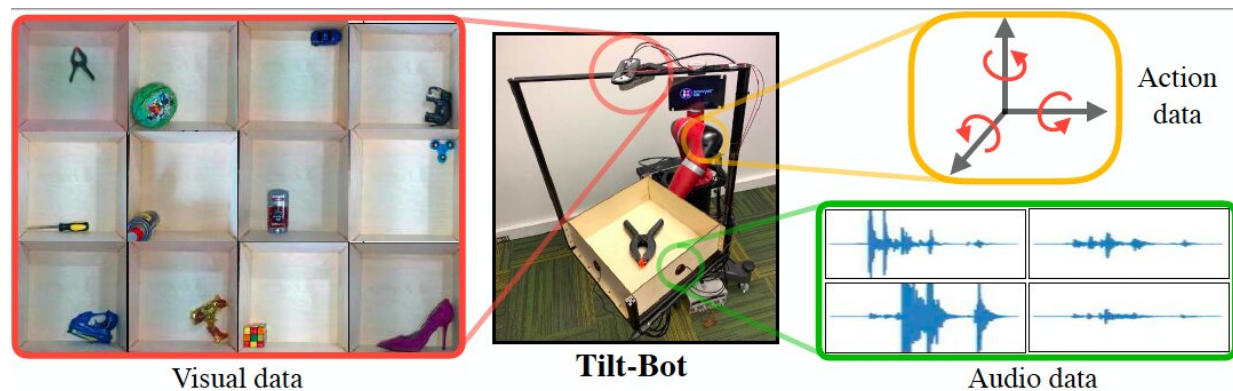


# Exploring the interactions between sound, action and vision in robotics

August 10 2020, by Ingrid Fadelli



Credit: Gandhi, Gupta & Pinto.

In recent years, researchers have developed a growing amount of computational techniques to enable human-like capabilities in robots. Most techniques developed so far, however, merely focus on artificially reproducing the senses of vision and touch, disregarding other senses, such as auditory perception.

A research team at Carnegie Mellon University (CMU) have recently carried out a study exploring the possibility of using sound to develop robots with more advanced sensing capabilities. Their paper, published in *Robotics: Science and Systems*, introduces the largest sound-action-vision dataset compiled up to date—which was collected as a [robotic](#)

[platform](#) called Tilt-Bot—and interacted with a wide variety of objects.

"In [robot](#) learning, we often only use visual inputs for perception, but humans have more sensory modalities than just vision," said Lerrel Pinto, one of the researchers who carried out the study, to TechXplore. "Sound is a key component of learning and understanding our physical environment. So, we asked the question: What can sound buy us in robotics? To answer this question, we created Tilt-Bot, a robot that can interact with objects and collect a large-scale audio-visual dataset of interactions."

Essentially, Tilt-Bot is a robotic tray that tilts objects until they hit one of the tray's walls. Pinto and his colleagues placed contact microphones on the robotic tray's walls to record the sounds produced when objects hit the wall and used an overhead camera to visually capture each [object](#)'s movements.

The researchers collected both visual and [audio data](#) for over 15,000 Tilt-Bot interactions with 60 different objects. This allowed them to compile a new image and audio dataset that could help to train robots to make associations between actions, images, and sounds.

In their paper, Pinto and his colleagues used this dataset to explore the relationship between sound and action in robotics applications, collecting a number of interesting findings. Firstly, they found that analyzing [sound recordings](#) of objects moving and hitting surfaces could allow machines to tell different objects apart, for instance differentiating between a metal screwdriver and a metal wrench.

"One exciting preliminary result of our study was that from sound alone you can recognize the type of object with close to 80% accuracy," Pinto explained. "We also showed that a machine can learn audio-based representations of objects that can help solve robotic tasks later on. For

example, when identifying the sound of an empty wine glass, a robot could understand that manipulating it will require different actions than those it would perform when handling a full wine glass."

Interestingly, Pinto and his colleagues showed that sound recordings can sometimes provide more valuable information than visual representations for solving robotics tasks, as they can also be used to effectively predict the future motions of an object. In a series of experiments using objects that the robot had not encountered during training, they found that the audio embeddings collected as their robot interacted with these objects could predict forward models (i.e., how to best manipulate an object in the future) 24% better than passive visual embeddings.

The dataset compiled by this team of researchers could ultimately help to develop robots that can select their actions and object manipulation strategies based on both audio recordings and images collected in their surroundings. Pinto and his colleagues are now planning further studies exploring the potential of sound analysis for creating robots with more advanced capabilities.

"This work is only a first step in holistically integrating sound in robotics," Pinto said. "In our future work, we will be looking at more practical applications of [sound](#) and action."

**More information:** Swoosh! Rattle! Thump! – Actions that sound. arXiv:2007.01851 [cs.RO]. [arxiv.org/abs/2007.01851](https://arxiv.org/abs/2007.01851)

[dhiraj100892.github.io/swoosh/](https://dhiraj100892.github.io/swoosh/)

© 2020 Science X Network

Citation: Exploring the interactions between sound, action and vision in robotics (2020, August 10) retrieved 25 March 2023 from <https://techxplore.com/news/2020-08-exploring-interactions-action-vision-robotics.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.