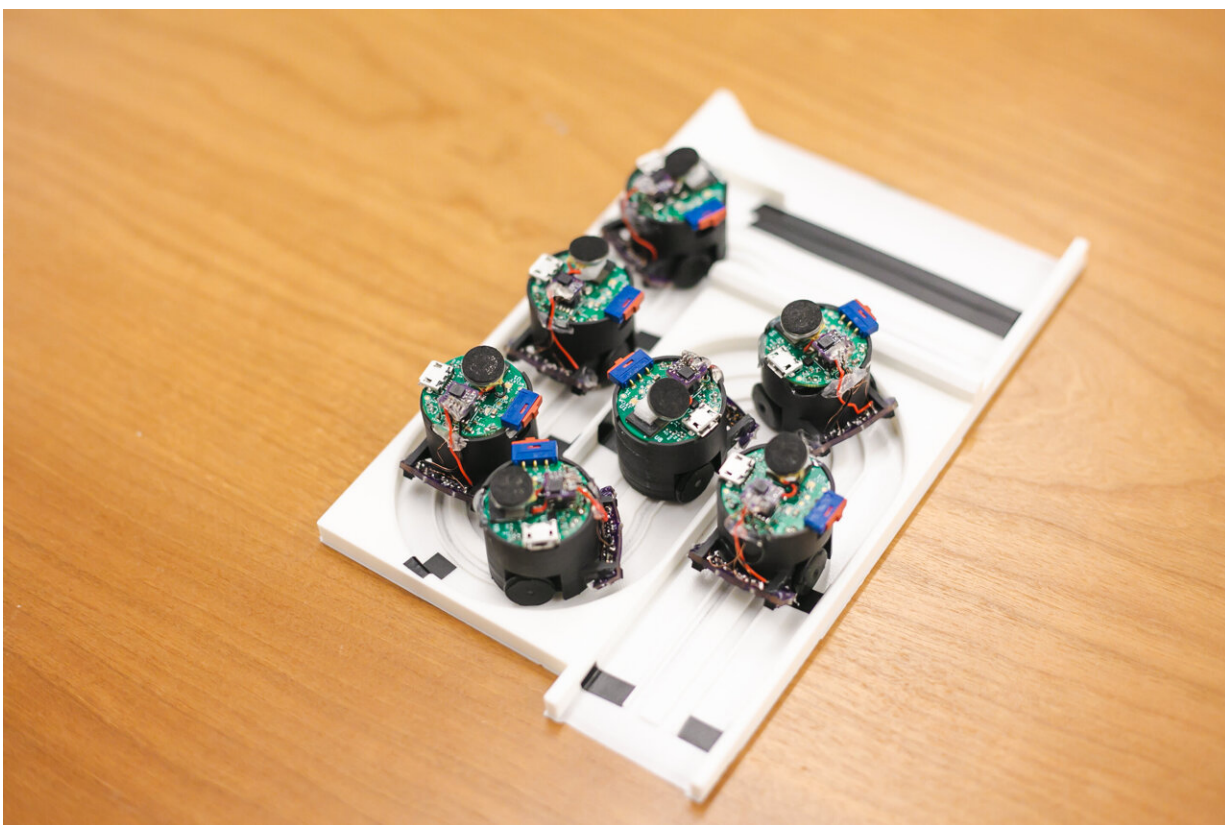


# Research team's shape-changing smart speaker lets users mute different areas of a room

September 21 2023

---



A team led by researchers at the University of Washington has developed a shape-changing smart speaker, which uses self-deploying microphones to divide rooms into speech zones and track the position of individual speakers. Here, the swarm of robots is shown in its charging station, which the robots can return to automatically. Credit: April Hong/University of Washington

In virtual meetings, it's easy to keep people from talking over each other. Someone just hits mute. But for the most part, this ability doesn't translate easily to recording in-person gatherings. In a bustling cafe, there are no buttons to silence the table beside you.

The ability to locate and control sound—isolating one person talking from a specific location in a crowded room, for instance—has [challenged researchers](#), especially without visual cues from cameras.

A team led by researchers at the University of Washington has developed a shape-changing [smart speaker](#), which uses self-deploying microphones to divide rooms into speech zones and track the positions of individual speakers. With the help of the team's deep-learning algorithms, the system lets users mute certain areas or separate simultaneous conversations, even if two adjacent people have similar voices.

Like a fleet of Roombas, each about an inch in diameter, the microphones automatically deploy from and then return to a charging station. This allows the system to be moved between environments and set up automatically. In a conference room meeting, for instance, such a system might be deployed instead of a central microphone, allowing better control of in-room audio.

The team has published [its findings](#) in *Nature Communications*.

"If I close my eyes and there are 10 people talking in a room, I have no idea who's saying what and where they are in the room exactly. That's extremely hard for the human brain to process. Until now, it's also been difficult for technology," said co-lead author Malek Itani, a UW doctoral student in the Paul G. Allen School of Computer Science & Engineering. "For the first time, using what we're calling a robotic 'acoustic swarm,' we're able to track the positions of multiple people talking in a room and

separate their speech."

Previous research on [robot swarms](#) has required using overhead or on-device cameras, projectors or special surfaces. The UW team's system is the first to accurately distribute a robot swarm using only sound.

The team's prototype consists of seven small robots that spread themselves across tables of various sizes. As they move from their charger, each robot emits a [high frequency sound](#), like a bat navigating, using this frequency and other sensors to avoid obstacles and move around without falling off the table.

The automatic deployment allows the robots to place themselves for maximum accuracy, permitting greater sound control than if a person set them. The robots disperse as far from each other as possible since greater distances make differentiating and locating people speaking easier. Today's consumer smart speakers have multiple microphones, but clustered on the same device, they're too close to allow for this system's mute and active zones.

"If I have one microphone a foot away from me, and another microphone two feet away, my voice will arrive at the microphone that's a foot away first. If someone else is closer to the microphone that's two feet away, their voice will arrive there first," said co-lead author Tuocho Chen, a UW doctoral student in the Allen School.

"We developed [neural networks](#) that use these time-delayed signals to separate what each person is saying and track their positions in a space. So you can have four people having two conversations and isolate any of the four voices and locate each of the voices in a room."

The team tested the robots in offices, living rooms and kitchens with groups of three to five people speaking. Across all these environments,

the system could discern different voices within 1.6 feet (50 centimeters) of each other 90% of the time, without prior information about the number of speakers. The system was able to process three seconds of audio in 1.82 seconds on average—fast enough for [live streaming](#), though a bit too long for real-time communications such as video calls.

As the technology progresses, researchers say, acoustic swarms might be deployed in smart homes to better differentiate people talking with smart speakers. That could potentially allow only people sitting on a couch, in an "active zone," to vocally control a TV, for example.

Researchers plan to eventually make [microphone](#) robots that can move around rooms, instead of being limited to tables. The team is also investigating whether the speakers can emit sounds that allow for real-world mute and active zones, so people in different parts of a room can hear different audio. The current study is another step toward science fiction technologies, such as the "cone of silence" in "Get Smart" and "Dune," the authors write.

Of course, any technology that evokes comparison to fictional spy tools will raise questions of privacy. Researchers acknowledge the potential for misuse, so they have included guards against this: The microphones navigate with sound, not an onboard camera like other similar systems.

The robots are easily visible and their lights blink when they're active. Instead of processing the audio in the cloud, as most smart speakers do, the acoustic swarms process all the audio locally, as a privacy constraint. And even though some people's first thoughts may be about surveillance, the system can be used for the opposite, the team says.

"It has the potential to actually benefit privacy, beyond what current smart speakers allow," Itani said. "I can say, 'Don't record anything

around my desk,' and our system will create a bubble three feet around me. Nothing in this bubble would be recorded. Or if two groups are speaking beside each other and one group is having a private conversation, while the other group is recording, one conversation can be in a mute zone, and it will remain private."

Takuya Yoshioka, a principal research manager at Microsoft, is a co-author on this paper, and Shyam Gollakota, a professor in the Allen School, is a senior author.

**More information:** Creating Speech Zones Using Self-distributing Acoustic Swarms, *Nature Communications* (2023). [DOI: 10.1038/s41467-023-40869-8](https://doi.org/10.1038/s41467-023-40869-8).  
[www.nature.com/articles/s41467-023-40869-8](https://www.nature.com/articles/s41467-023-40869-8)

Provided by University of Washington

Citation: Research team's shape-changing smart speaker lets users mute different areas of a room (2023, September 21) retrieved 6 May 2024 from <https://techxplore.com/news/2023-09-team-shape-changing-smart-speaker-users.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.