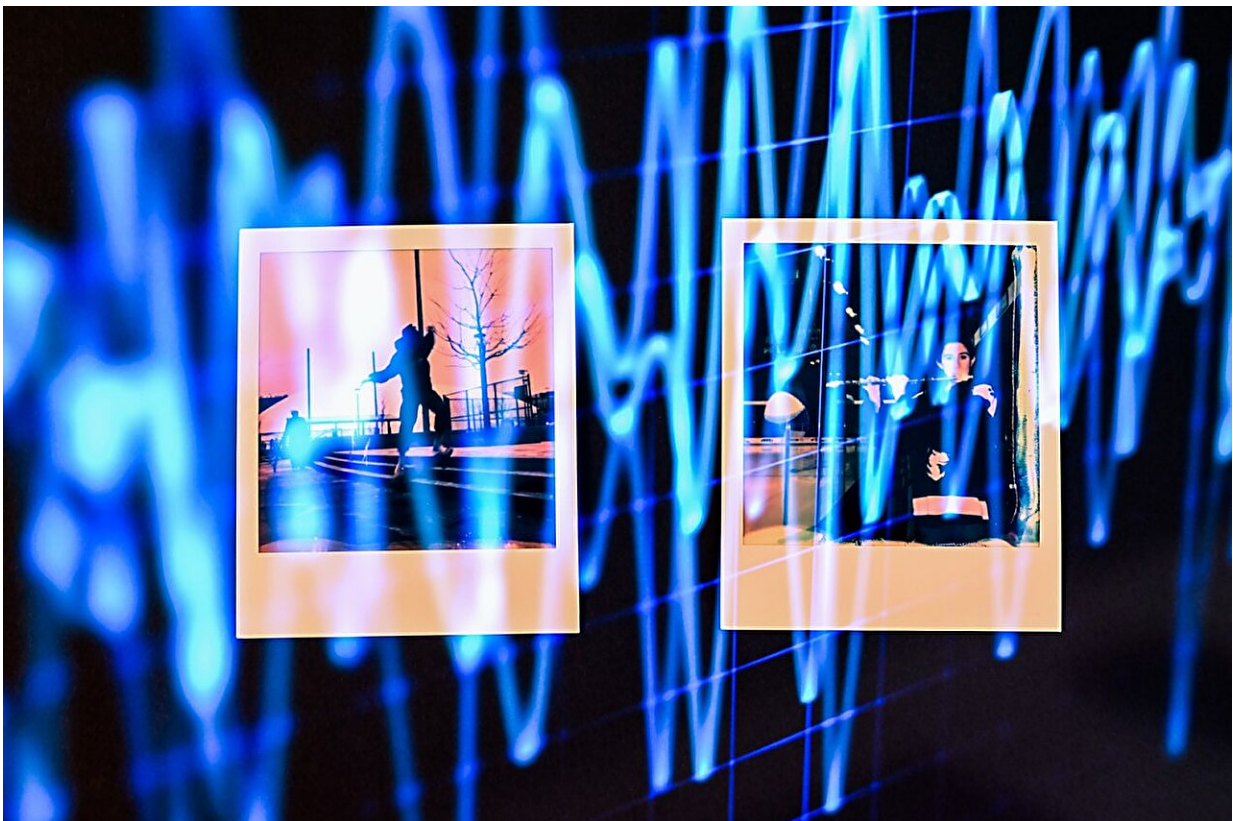


Researcher finds way to get audio from still images and silent videos

September 25 2023, by Cody Mello-Klein



When you take a photo on your phone, the vibrations of your voice can create tiny bends in the light that are enough to extract audio, according to Kevin Fu, a professor of engineering and computer science at Northeastern University. Credit: by Matthew Modoono/Northeastern University

With video calls becoming more common in the age of remote and

hybrid workplaces, "mute yourself" and "I think you're muted" have become part of our everyday vocabularies. But it turns out muting yourself might not be as safe as you think.

Kevin Fu, a professor of electrical and [computer engineering](#) and [computer science](#) at Northeastern University, has figured out a way to get audio from pictures and even muted videos. Using Side Eye, a [machine learning](#) assisted tool that Fu and his research team created, Fu can determine the gender of someone speaking in the room where a photo was taken—and even the exact words they spoke.

"Imagine someone is doing a TikTok video and they mute it and dub music," Fu says. "Have you ever been curious about what they're really saying? Was it 'Watermelon watermelon' or 'Here's my password?' Was somebody speaking behind them? You can actually pick up what is being spoken off camera."

It sounds like the stuff of science fiction—and it is. The idea for Side Eye was inspired by an episode of the sci-fi show "Fringe" that saw the main characters, a team of fringe science investigators working for the FBI, extracting audio from a melted pane of glass.

When the episode aired, one critic for Den of Geek called it a "ridiculous pseudo science technique." Fu disagreed.

"I was like, 'I bet we can do that,'" Fu says. "My lab specializes in the impossible. We usually expect the first reaction to anything we do to be 'You can't do that,' and we say, 'Well, we already did.'"

Side Eye takes advantage of the image stabilization technology that is now virtually standard across most phone cameras. To ensure a shaky hand doesn't make for a blurry photo, cameras have small springs that hold the lens suspended in liquid. An electromagnet and sensors then

push the lens in equal and [opposite directions](#) to reduce camera shake.

However, Fu says whenever someone speaks near a [camera lens](#), it causes tiny vibrations in the springs and bends the light ever so slightly. The angle of the light changes almost imperceptibly—"unless you're looking for it," Fu says.

Normally, it would be hard to extract sonic frequency from those microscopic vibrations. But Fu says rolling shutter, a method of photography most phone cameras use today, actually makes it easier to achieve the impossible.

"The way cameras work today to reduce cost basically is they don't scan all pixels of an image simultaneously — they do it one row at a time," Fu says. "[That happens] hundreds of thousands of times in a single photo. What this basically means is you're able to amplify by over a thousand times how much frequency information you can get, basically the granularity of the audio."

As long as there is even a little bit of light, Side Eye will work, although the more imagery it has access to, the better. Fu says even a photo pointed at a ceiling would let Side Eye do its thing.

The end result of this process is audio that, even at its best, sounds more like the muffled sound of adults in the Peanuts cartoons. But by using machine learning and training Side Eye on certain words and audio, Fu is able to extract a lot of information.

"If you want to know if I said yes or no, you can train [Side Eye] on people saying yes and no and then look at the patterns and with high confidence when I get an image later know if someone said yes or no," Fu says.

Side Eye can even identify the exact person who is speaking if it's been trained on that person's voice, although Fu says it's not as accurate when it comes to that just yet.

From a cybersecurity perspective, Side Eye opens up an entirely new world of threats that people and cybersecurity experts should be aware of. However, Fu says the most interesting application for Side Eye could be as a new form of digital evidence for lawyers and others working in the criminal legal system.

"Maybe there's an alibi and it's being admitted to court and somebody wants to prove somebody was or wasn't there," Fu says. "You might be able to use this technique if you have an authenticated video with a known timestamp to confirm one way or the other. If you hear the person's voice, they're more than likely there."

Provided by Northeastern University

Citation: Researcher finds way to get audio from still images and silent videos (2023, September 25) retrieved 13 May 2024 from

<https://techxplore.com/news/2023-09-audio-images-silent-videos.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.
