The New York Times

Alexa and Siri Can Hear This Hidden Command. You Can't.

Researchers can now send secret audio instructions undetectable to the human ear to Apple's Siri, Amazon's Alexa and Google's Assistant.

By Craig S. Smith

May 10, 2018

BERKELEY, Calif. — Many people have grown accustomed to talking to their smart devices, asking them to read a text, play a song or set an alarm. But someone else might be secretly talking to them, too.

Over the last two years, researchers in China and the United States have begun demonstrating that they can send hidden commands that are undetectable to the human ear to Apple's Siri, Amazon's Alexa and Google's Assistant. Inside university labs, the researchers have been able to secretly activate the artificial intelligence systems on smartphones and smart speakers, making them dial phone numbers or open websites. In the wrong hands, the technology could be used to unlock doors, wire money or buy stuff online — simply with music playing over the radio.

A group of students from University of California, Berkeley, and Georgetown University showed in 2016 that they could hide commands in white noise played over loudspeakers and through YouTube videos to get smart devices to turn on airplane mode or open a website.

This month, some of those Berkeley researchers published a research paper that went further, saying they could embed commands directly into recordings of music or spoken text. So while a human listener hears someone talking or an orchestra playing, Amazon's Echo speaker might hear an instruction to add something to your shopping list.

"We wanted to see if we could make it even more stealthy," said Nicholas Carlini, a fifth-year Ph.D. student in computer security at U.C. Berkeley and one of the paper's authors.

[*Read more on what Alexa can hear when brought into your home*]

Mr. Carlini added that while there was no evidence that these techniques have left the lab, it may only be a matter of time before someone starts exploiting them. "My assumption is that the malicious people already employ people to do what I do," he said.

These deceptions illustrate how artificial intelligence — even as it is making great strides — can still be tricked and manipulated. Computers can be fooled into identifying an airplane as a cat just by changing a few pixels of a digital image, while researchers can make a self-driving car swerve or speed up simply by pasting small stickers on road signs and confusing the vehicle's computer vision system.

With audio attacks, the researchers are exploiting the gap between human and machine speech recognition. Speech recognition systems typically translate each sound to a letter, eventually compiling those into words and phrases. By making slight changes to audio files, researchers were able to cancel out the sound that the speech recognition system was supposed to hear and replace it with a sound that would be transcribed differently by machines while being nearly undetectable to the human ear.

The Amazon Echo in a kitchen. Christie Hemm Klok for The New York Times

The proliferation of voice-activated gadgets amplifies the implications of such tricks. Smartphones and smart speakers that use digital assistants like Amazon's Alexa or Apple's Siri are set to outnumber people by 2021, according to the research firm Ovum. And more than half of all American households will have at least one smart speaker by then, according to Juniper Research.

Amazon said that it doesn't disclose specific security measures, but it has taken steps to ensure its Echo smart speaker is secure. Google said security is an ongoing focus and that its Assistant has features to mitigate undetectable audio commands. Both companies' assistants employ voice recognition technology to prevent devices from acting on certain commands unless they recognize the user's voice.

Apple said its smart speaker, HomePod, is designed to prevent commands from doing things like unlocking doors, and it noted that iPhones and iPads must be unlocked before Siri will act on commands that access sensitive data or open apps and websites, among other measures.

Yet many people leave their smartphones unlocked, and, at least for now, voice recognition systems are notoriously easy to fool.

There is already a history of smart devices being exploited for commercial gains through spoken commands.

Last year, Burger King caused a stir with an online ad that purposely asked 'O.K., Google, what is the Whopper burger?" Android devices with voice-enabled search would respond by reading from the Whopper's Wikipedia page. The ad was canceled after viewers started editing the Wikipedia page to comic effect.

[Read more on how we may soon be living in Alexa's world]

A few months later, the animated series South Park followed up with an entire episode built around voice commands that caused viewers' voice-recognition assistants to parrot adolescent obscenities.

There is no American law against broadcasting subliminal messages to humans, let alone machines. The Federal Communications Commission discourages the practice as "counter to the public interest," and the Television Code of the National Association of Broadcasters bans "transmitting messages below the threshold of normal awareness." Neither say anything about subliminal stimuli for smart devices.

Courts have ruled that subliminal messages may constitute an invasion of privacy, but the law has not extended the concept of privacy to machines.

Now the technology is racing even further ahead of the law. Last year, researchers at Princeton University and China's Zhejiang University demonstrated that voice-recognition systems could be activated by using frequencies inaudible to the human ear. The attack first muted the phone so the owner wouldn't hear the system's responses, either.

The technique, which the Chinese researchers called DolphinAttack, can instruct smart devices to visit malicious websites, initiate phone calls, take a picture or send text messages. While DolphinAttack has its limitations — the transmitter must be close to the receiving device — experts warned that more powerful ultrasonic systems were possible.

That warning was borne out in April, when researchers at the University of Illinois at Urbana-Champaign demonstrated ultrasound attacks from 25 feet away. While the commands couldn't penetrate walls, they could control smart devices through open windows from outside a building.

This year, another group of Chinese and American researchers from China's Academy of Sciences and other institutions, demonstrated they could control voice-activated devices with commands embedded in songs that can be broadcast over the radio or played on services like YouTube.

More recently, Mr. Carlini and his colleagues at Berkeley have incorporated commands into audio recognized by Mozilla's DeepSpeech voice-to-text translation software, an open-source platform. They were able to hide the command, "O.K. Google, browse to evil.com" in a recording of the spoken phrase, "Without the data set, the article is useless." Humans cannot discern the command.

The Berkeley group also embedded the command in music files, including a four-second clip from Verdi's "Requiem."

How device makers respond will differ, especially as they balance security with ease of use.

"Companies have to ensure user-friendliness of their devices, because that's their major selling point," said Tavish Vaidya, a researcher at Georgetown. He wrote one of the first papers on audio attacks, which he titled "Cocaine Noodles" because devices interpreted the phrase "cocaine noodles" as "O.K., Google."

Mr. Carlini said he was confident that in time he and his colleagues could mount successful adversarial attacks against any smart device system on the market.

"We want to demonstrate that it's possible," he said, "and then hope that other people will say, 'O.K. this is possible, now let's try and fix it.' "

Follow Craig S. Smith on Twitter: @craigss

A version of this article appears in print on May 10, 2018, on Page B1 of the New York edition with the headline: How Your Smart Speaker Will Be Hijacked By Dog Whistle