Are Alexa, Siri, and Cortana recording your private conversations?

July 15 2020, by Laura Castañón

Credit: Matthew Modoono/Northeastern University

Smart speakers act as our personal assistants, offering a hands-free way to get information, control our homes, and manage our schedules. And like any good assistant, they are always listening.

Each device has a "wake word" that triggers it to begin recording, processing, and transmitting what it hears.

"Alexa, is it going to rain?"

"OK Google, remind me to call my mom at noon tomorrow."

"Hey Siri, play 'Toxic' by Brittany Spears."

But how often are these devices waking up accidentally? Are they recording our private conversations and sending them to cloud-based storage? Is our personal information safe?

The good news is, they aren't recording all the time, says David Choffnes, an associate professor of computer sciences at Northeastern. And when they are activated accidentally, the recordings are typically short.

"But that doesn't mean there's no cause for concern," he says.

In a recent study, Choffnes and his colleagues set out to determine how often, and for how long, these speakers might be waking up and recording us. And because the inner workings of these devices are proprietary information, the best way to test them was to watch as they listened to conversations. A lot of conversations.

The researchers set up smart speakers of various types with the ultimate binge watch: 134 hours of audio from a dozen popular Netflix shows, including "Gilmore Girls," "Narcos," "The West Wing," and "Dear White People." A video of the scene (seven smart speakers listening attentively to one regular speaker) tracked activations based on each speaker's indicator light.

The researchers matched the light with router traffic, indicating that recordings were being sent over the internet, and also checked the cloud-based activation log provided by Amazon and Google for their respective devices.

"We had to create a lot of infrastructure and do something a little bit crazy," Choffnes says. "But that allowed us to get the devices exposed to a week's worth of dialogue, without it requiring me and Daniel Dubois [a research scientist at Northeastern] and other people on the team sitting in a room having conversations following a script."

Using TV shows also provided a greater diversity of speakers—each show might have ten or more characters speaking, with a variety of ages, accents, and occasionally languages.

The show that caused the most misactivations, 6.21 for every 10,000 words spoken, was Narcos. These activations seemed to happen primarily during dialogue in Spanish, or in heavily accented English, which raises an additional concern: Are these devices misactivating more frequently in, and intruding on the privacy of, people who speak multiple languages or don't have a "typical American" accent?

"That's one area we really want to look into," Choffnes says. "We did try to select a really wide range of shows in our study, but we did not do the kinds of additional tests and rigorous analysis to be able to tease apart these impacts and how there might be uneven behavior from the device towards certain groups of people."

Results varied depending on the individual speaker, with some misactivating at a rate of almost once per hour, or 1.43 times for every 10,000 words spoken. While these activations were typically short, for several devices 10 percent of their misactivations last 10 seconds or more.

Sometimes, it was clear why the speaker had made an error. Phrases like "I'm sorry" or "Hey, how are you feeling?" could be mistaken for "Hey Siri"; "head coach" and "pickle" sound enough like "Echo" to trigger Amazon's device.

But other activations seemed almost random.

"There's a lot of these cases where a device wakes up and we don't know why, and they're not repeatable," Choffnes says. "And that means in your home, in your work, wherever these devices happen to be located, there's just a whole bunch of random things that will cause them to wake up and record audio. And that can be a concern if you really want to make sure they're not recording when they shouldn't."

It may not seem like anyone could glean much from a 10 second audio clip, but we're not talking about just one recording. Every snippet of conversation is stored in the cloud, and that could add up to as substantial privacy risk over time, Choffnes says.

Even if this information isn't accessed by a malicious party, the companies that make smart speakers use the audio files to improve their voice-recognition software, by paying contractors to transcribe what is being said.

"A lot of us, when we think about being in our home, we think that's a private space where we can have conversations that are intended not to be shared," Choffnes says. "And now we have all these devices with microphones that could be taking those conversations and sharing them."

More information: moniotrlab.ccis.neu.edu/smart- … eakers-study-pets20/