Protecting your smartphone from voice impersonators

July 19 2017, by Kui Ren

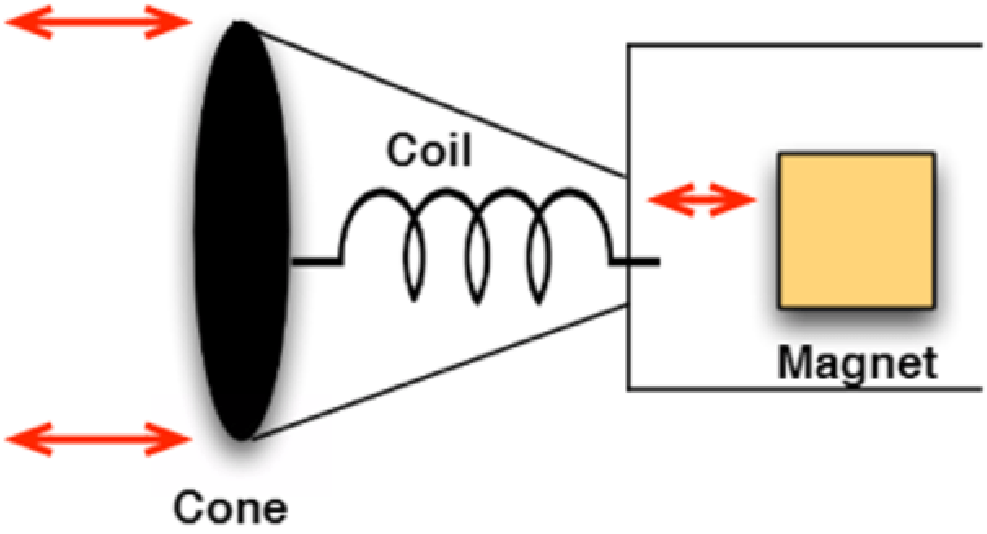
Credit: AI-generated image (disclaimer)

It's a lot easier to talk to a smartphone than to try to type instructions on its keyboard. This is particularly true when a person is trying to log in to a device or a system: Few people would choose to type a long, complex secure password if the alternative were to just say a few words and be authenticated with their voice. But voices can be recorded, simulated or even imitated, making voice authentication vulnerable to attack.

The most common methods for securing voice-based authentication involve only ensuring that analysis of a spoken passphrase is not tampered with; they securely store the passphrase and the authorized user's voiceprint in an encrypted database. But securing a voice authentication system has to start with the sound itself.

The easiest attack on voice authentication is impersonation: Find someone who sounds enough like the real person and get them to respond to the login prompts. Fortunately, there are automatic speaker verification systems that can detect human imitation. However, those systems can't detect more advanced machine-based attacks, in which an attacker uses a computer and a speaker to simulate or play back recordings of a person's voice.

If someone records your voice, he can use that recording to create a computer model that can generate any words in your voice. The consequences, from impersonating you with your friends to dipping into your bank account, are terrifying. The research my colleagues and I are doing uses fundamental properties of audio speakers, and smartphones' own sensors, to defeat these computer-assisted attacks.



The architecture of conventional loudspeaker showing the magnet, coil and cone used for loudspeaker operations.

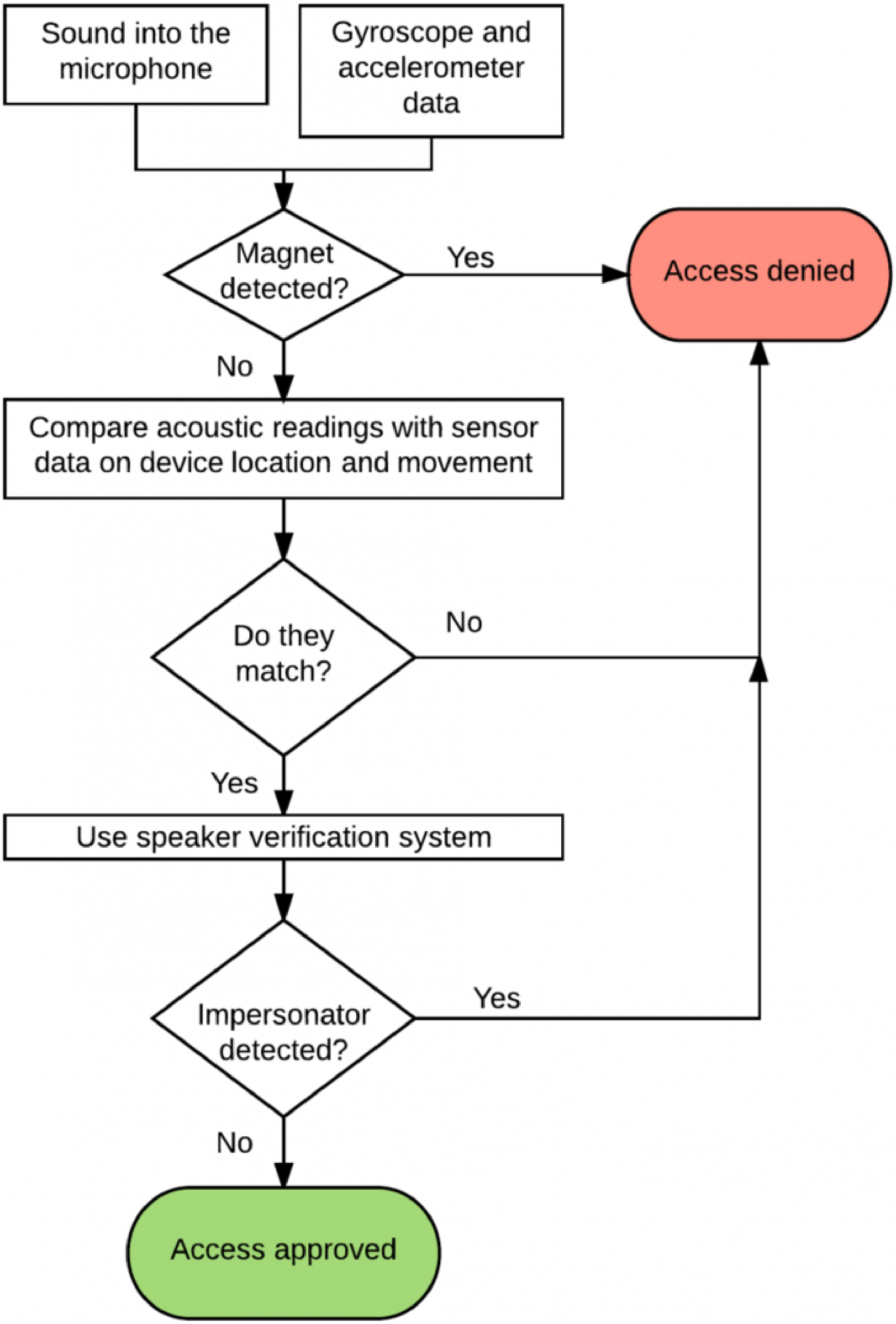
How speakers work

Conventional speakers contain magnets, which vibrate back and forth according to fluctuations of electrical or digital signals, converting them into sound waves in the air. Putting a speaker up against the microphone of a smartphone, for example, means moving a magnet very close to the smartphone. And most smartphones contain a magnetometer, an electronic chip that can detect magnetic fields. (It comes in handy when using a compass or navigation app, for example.)

If the smartphone detects a magnet nearby during the process of voice authentication, that can be an indicator that a real human might not be doing the talking.

Making sure it's a person talking

That's just one part of our system. If someone uses a smaller speaker, like a set of headphones, the magnetometer might not detect its smaller magnets. So we use machine learning and advanced mathematics to examine physical properties of the sound as it arrives at the microphone.



Sound into the microphone
Gyroscope and accelerometer data
Magnet detected?
Yes
Access denied
No
Compare acoustic readings with sensor data on device location and movement
Do they match?
No
Yes
Use speaker verification system
Impersonator detected?
Yes
No
Access approved

An outline of how our process works. Credit: Kui Ren et al., CC BY-ND

Our system requires a user to hold the smartphone in front of his or her face and move it from side to side in a half-circle while speaking. We combine the sound captured by the microphone with movement data from gyroscopes and accelerometers inside the smartphone – the same sensors apps use to know when you're walking or running, or changing direction.

Using that data, we can calculate how far away from the microphone the sound is being generated – which lets us identify the possibility of someone using speakers at a distance so its magnets wouldn't be detected. And we can compare the phone's movement to the changes in the sound to discover whether it is created by a sound source roughly the size of a human mouth near the phone.

All of this, of course, could be defeated by a skilled impersonator – an actual human who mimics a user's voice. But recall that existing speaker verification methods can catch impersonators, using machine learning techniques that identify whether a speaker is modifying or disguising his or her normal voice. We include that capability in our system as well.

Does detection work?

When we put our system to the test, we found that when the sound source is 6 centimeters (2 inches) from the microphone, we can always distinguish between a human and a computer-controlled speaker. At that distance, the magnet in a normal loudspeaker is strong enough to clearly interfere with the phone's magnetometer. And if an attacker is using earphone speakers, the microphone is close enough to the sound source to detect it.

When the sound source is farther from the microphone, it's harder to detect magnetic interference from a speaker. It's also more difficult to analyze the movement of the sound source in relation to the phone when the distances are greater. But by using multiple lines of defense, we can defeat the vast majority of speaker- and human-based attacks and significantly improve the security of voice-based mobile apps.

At the moment, our system is a stand-alone app, but in the future we'll be able to integrate it into other voice authentication systems.

This article was originally published on The Conversation. Read the original article.

Provided by The Conversation

Citation: Protecting your smartphone from voice impersonators (2017, July 19) retrieved 10 April 2024 from https://phys.org/news/2017-07-smartphone-voice-impersonators.html