Hey Siri – are you there?: Jamming of voice commands using the resonance effect

Taekkyung Oh, William Aiken, Hyoungshick Kim

Sungkyunkwan University
Suwon, Republic of Korea
{ohtk, billzo, hyoung}@skku.edu

***Abstract*—Micro Electro-Mechanical Systems (MEMS) microphones have become popularly used in portable devices thanks to their numerous advantages over other types of microphones. However, MEMS microphones introduce their own vulnerabilities, and in this paper we discuss the possibility of new attacks that impact devices equipped with a MEMS microphone. We found that these devices can be vulnerable to a new jamming attack based on the resonance effect inherent in the vibrating nature of MEMS devices. For example, a user's voice commands can be canceled by a jamming attack by broadcasting carefully crafted audio signals. When these signals are generated with a frequency that matches the inherent frequency of the target MEMS microphone, the microphone's membrane will resonant at a significantly large amplitude which prevents the device from receiving the victim's command.**

***Index Terms*—Sensor systems, Denial-of-service attack, Audio systems, Microelectromechanical systems.**

I. Introduction

Voice assistants are becoming more powerful and more capable of processing advanced commands. Popular applications like Siri (Apple), Alexa (Amazon), Google Now, Cortana (Microsoft), and Bixby (Samsung) allow people to shop online, place phone calls, send instant messages, schedule appointments, check emails, create to-do lists, control smart home appliances, and access banking services: all from a voice command. Therefore, it is paramount to secure voice command channels when interacting with voice assistants.

Micro Electro-Mechanical Systems (MEMS) are devices that contain at least one component between 0.001 and 0.1 millimeters in size and that are able to sense or affect their environment. Because MEMS technology provides inherent cost savings and size reduction, MEMS are popularly used in a wide range of devices including vascular system abnormality detectors [1], auto-pilot sensors for drones [2], and blood pressure monitors [3]. Recently, MEMS microphones have become increasingly popular in a variety of devices due to their numerous advantages over other microphones.

However, MEMS-based systems have been found to be vulnerable to several attacks. Son et al. [4] presented an attack using intentional sound noises to interfere with MEMS gyroscopes embedded in drones. Trippel et al. [5] showed that intentional acoustic interference can spoof MEMS accelerometer data. These attacks exploit a physical phenomenon, known as *resonance* where a source of vibration forces an object to oscillate with increasing amplitude at the object's inherent frequencies.

In this paper, we propose a novel Denial-of-Service (DoS) attack based on the resonance effect, aimed at disrupting the delivery of voice commands for devices equipped with a MEMS microphone.

II. Background

In this section, we introduce an attack model which aims to overwhelm a MEMS microphone via the resonance effect. We propose that while under the influence of specially crafted signals, a microphone cannot correctly transfer input (i.e., a voice signal) to the system (e.g., a speech recognition system of an AI). In order to conduct this attack, we follow the procedure used to implement denial-of-service attacks against drone gyroscopes [4]. Because a MEMS microphone and MEMS gyroscope both rely on the physical properties of the interaction between the membrane and a back plate, we believe that both are equally susceptible to malicious signals.

A. Resonance effect

Resonance is a phenomenon where the energy of a signal is amplified when an external signal's frequency and the system's inherent frequency coincide. The specific frequency that causes resonance differs from object to object, but an exact input frequency at any objects' natural frequency can cause maximum amplitude oscillations.

The resonance effect is a very famous and useful phenomenon in the scientific world. Many fields such as RF communication in the electrical layer or acoustic synthesis in the physical layer take advantage of its properties. For a further example, television and radio electronics rely on the resonance effect. Choosing a radio frequency or television channel requires matching the circuit frequency inside the radio or TV with the broadcast frequency of the station. In this way, many different broadcasts can be sent over the air simultaneously, and a user need only tune to the desired frequency to select one in particular.

However, contrary to the above positive aspects, there are also negative cases of the resonance effect. When the frequency of an external vibration and the natural frequency of an object coincide, the vibration of the object will become very large. When this occurs, the object may break or fail.

For instance, the famous suspension bridge of the Tacoma Narrows collapsed in part because the natural frequency of the bridge was consistent with the frequency of its shaking in strong winds. Although the bridge was designed to endure 190km/h strong wind, it collapsed at just 70km/h wind. Our proposal focuses on this negative side of resonance.

B. MEMS microphone

A MEMS microphone is a solid state integrated circuit (IC) which can sense vibrations caused by voice commands. Because MEMS microphones have many attractive qualities such as their small size, high signal-to-noise ratio, low power consumption, and high sensitivity, they are suitable for use in a wide variety of devices. They are becoming increasingly popular in modern devices like mobile phones, tablets, and laptops, and they frequently serve as the input for communicating with speech recognition artificial intelligence systems. More

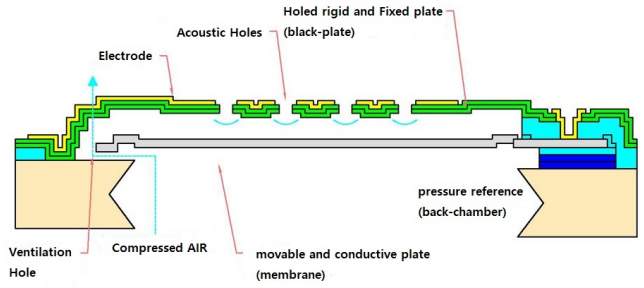

Fig. 1. Structure of MEMS microphone sensor.

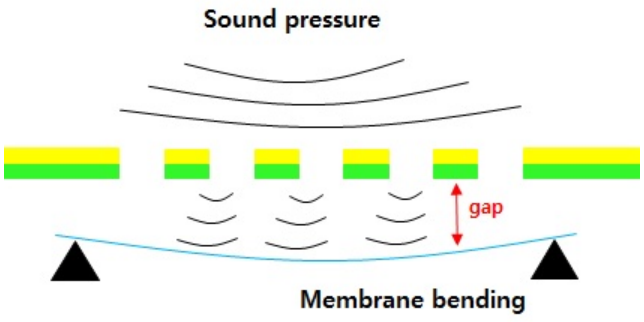

Fig. 2. Principle of MEMS microphone's operation.

specifically, a MEMS microphone-equipped device consists of not only a MEMS microphone sensor but also an Application Specific Integrated Circuit (ASIC). In Figure 1, you can see construction of a MEMS microphone sensor. Essentially, a MEMS microphone is an acoustic transducer. The following describes the construction of such an acoustic transducer:

- Signal transduction occurs when the coupled capacity changes between a fixed plate (back-plate) and a movable plate (membrane).
- The capacitive change is caused by the sound passing through the acoustic holes that moves the membrane. It modulates the air gap between the two conductive plates.
- The back-chamber serves as the acoustic resonator.
- The ventilation hole allows the air compressed in the back chamber to flow out and consequently allows the membrane to move back into place.

Figure 2 shows the principle means of operation. Changes in air pressure created by sound waves cause the thin membrane to flex while the thicker fixed plate remains stationary as the air moves through its acoustic holes. The movement of the membrane creates a change in the amount of capacitance between the membrane and the fixed plate, which is translated into an electrical signal by the ASIC. Then, the ASIC measures the voltage change that occurs when the capacitance between the membrane and the fixed plate changes due to membrane movement. As a result, the voice signal is interpreted by the movement of the membrane. We suppose that if we manipulate the movement of the membrane in MEMS microphone sensor maliciously, MEMS microphone will output incorrect signals to the device.

III. Jamming Voice Commands Based on the Resonance Effect

Our attack method proposes to confuse a MEMS microphone by obfuscating legitimate voice signals by playing the natural frequency of the target device. When the malicious signal and MEMS microphone inherent frequency match, the microphone's membrane will oscillate at a very large amplitude. When the user tries to give a command to the MEMS microphone, the membrane will already be oscillating due to the malicious interference, and the MEMS microphone will be unable to interpret the legitimate sound correctly. In this way, we can effectively perform a denial of service attack on parts of the system that rely on voice commands as shown in Figure 3.

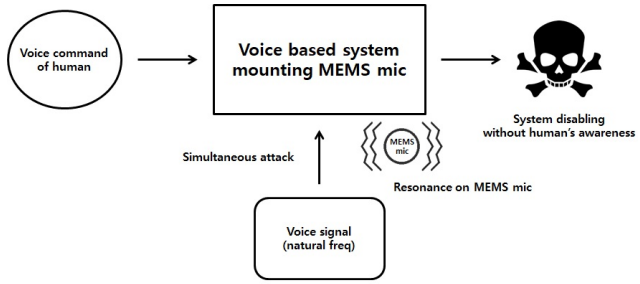

Fig. 3. Process of denial of service attack.

Previous work has shown that measuring the natural frequency of a membrane in a MEMS microphone can be performed by electrical impedance analysis experiments [6]. Here, the natural frequency is a resonance frequency in the physical or mechanical layer that is capable of generating a resonance effect. The resonance frequency of the membrane is measured at around 75kHz in the air or in a vacuum

environment. In addition, the resonance frequency of the fixed plate is measured at around 170kHz. Although our attack target is the membrane, if possible, we can inject additional frequencies that interfere with the fixed plate as well. Because 75kHz is a very high frequency, if we generate a malicious signal at this level, the victim user would not be able to hear it (the human hearing range is 20Hz to 20kHz). This has been illustrated in a similar manner in the Dolphin attack [7] where inaudible voice commands were injected into the victim's phone.

IV. Threat model

A traffic jamming attack is the process of disrupting traffic to a particular communication link between the sender and the recipient by injecting signals in an intentional manner. In order to successfully perform jamming attacks on voice commands, an attacker might be in close proximity (e.g., within a range of 3 meters) to either the victim or the victim's device. In many real world situations (in public places, workplaces and transportation environments), this assumption is reasonable.

We also assume that the attacker is equipped with an ultrasound speaker to generate sound samples (with the frequency band around 75kHz) that are inaudible to human listeners but can interrupt the transmission of voice commands by enhancing the resonance effect on the MEMS microphone in the target device.

Voice jamming attacks are difficult to detect because existing voice recognition systems are inherently prone to errors. Victims may tend to blame poor performance of the voice recognition system when their voices are not recognized. Because voice recognition can be negatively affected by environmental factors such as background noise, many users have already experienced failures in recognition.

V. Experiments

In order to demonstrate the feasibility of our approach, we implemented a proof-of-concept prototype to launch a voice jamming attack based on the resonance effect against voice commands for Siri and Alexa.

In our experiments, we used two voice assistant applications: Siri running on an iPhone 7 and Alexa running on Amazon's Echo Dot. A consumer-grade speaker [8] was used to broadcast the malicious audio signals. For experiments, we particularly generated two *attack sound signals*. Our first signal was a 16kHz frequency single tone sound with an echo effect and additive white noise. Our second signal was an aggressive growing sound of intense concentration like "Grrr". These *attack sound signals* are broadcasted at approximately 70dB. We note that the current prototype implementation just uses the audible frequency range (i.e., not resonance frequency range) even though our ultimate goal is to launch the voice jamming attacks using inaudible sound signals with the frequency band around 75kHz.

The experiment setup is shown in Figure 4. When a human participant utters some voice command (e.g., "Hey Siri" or "Alexa"), at the same time, the recorded attack sound signal is played through the speaker. We repeated this procedure 10 times for each application and each attack signal. The experiment results are shown in Table I and Table II.

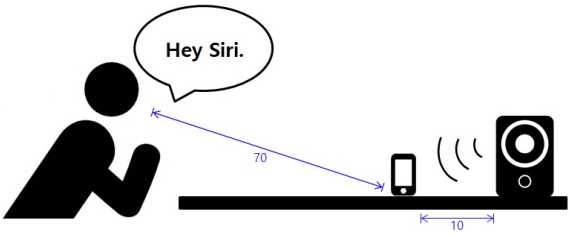

Fig. 4. Experiment Setup (unit : cm).

TABLE I
COMPARISON OF THE VOICE JAMMING ATTACK PERFORMANCE WITH A SINGLE TONE SOUND BETWEEN SIRI AND ALEXA.

	Siri	Alexa
# of total attempts	10	10
# of successful attacks	1	9
% of attacks	10%	90%

TABLE II
COMPARISON OF THE VOICE JAMMING ATTACK PERFORMANCE WITH A GROWING SOUND ("GRRR") BETWEEN SIRI AND ALEXA.

	Siri	Alexa
# of total attempts	10	10
# of successful attacks	10	10
% of attacks	100%	100%

The voice jamming attack success rates of the second experiment were higher than first experiment. This difference could be caused by noise filtering in the input circuit of the device, as the first *attack sound signal* contains significant white noise. In the second experiment, the voice jamming attack success rates against both Siri and Alexa were 100%. These preliminary findings demonstrate that voice assistant applications are usually susceptible to voice jamming attacks using a growing sound that can interfere with the transmission of human voice commands. In future work, we plan to extend our attack experiments by using inaudible sound signals broadcasted by an ultrasonic speaker to covertly launch the voice jamming attack using the resonance effect.

VI. Related work

Attacks that take advantage of the physical properties of resonance are not new, and they have demonstrated vulnerabilities across a wide range of devices and domains. Son et al. [4] conducted a series of tests to discover the resonant frequency in several MEMS gyroscopes, and from that information, they were able to successfully crash an aerial drone on every attempt by playing the resonant frequency at the target drone from a small Bluetooth speaker above the gyroscope. Trippel et al. [5] instead targeted a MEMS accelerometer. In their work, they effectively input fake steps to a FitBit tracker as well as took control of an RC car that was controlled by the smartphone's physical orientation.

There are also studies that exploit the effects of resonance in other environments. For example, in Wu et al.'s research [9], the authors demonstrate that by applying a modified input to a load frequency control power generation system, an attack can destabilize a smart grid while staying within admissible input boundaries. Recently, Sharad et al. [10] performed a resonance-based attack on Hard Disk Drives. In their work, the authors demonstrated that aiming well-crafted acoustic signals were able to stop security cameras from recording properly as well as inhibit file-copy processes in some operating systems.

VII. Conclusion and future work

In this work, we introduced an attack based on the resonance effect against MEMS microphones. Resonance is not only a useful scientific theory in many fields, but is also a very practical attack method in security. Because inaudible resonance attacks on voice-based systems using MEMS microphones have not been thoroughly investigated, our proposal can present a new direction of attack.

The audible approach in this paper can be applied to various devices such as the Amazon Echo, iPhone, Samsung Galaxy, and Google Home. This resonance attack on MEMS microphones poses a DoS threat to existing voice input systems. Moreover, a resonance attack on various sensors can be a major security threat if they are used in areas where real-time control is important such as automobile control sensors. If conspicuity is not a concern of the attacker, an audible denial of service attack as presented in this paper could be carried out against many MEMS speakers on existing systems using commodity hardware.

Acknowledgment

This research was supported by the MSIT(Ministry of Science, ICT),Korea, under the ITRC(Information Technology Research Center) support program (IITP-2018-2015-0-00403)supervised by the IITP(Institute for Information &communications Technology Promotion)

References

[1] L.-A. Louizos, P. G. Athanasopoulos, and K. Varty, "Microelectromechanical systems and nanotechnology: A platform for the next stent technological era," *Vascular and Endovascular Surgery*, vol. 46, no. 8, pp. 605–609, 2012.

[2] C. M. Ananda, P. Akula, and S. Prasad, "Mems sensor suites for micro air vehicle (mav) autopilot," in *International Conference on Circuits, Communication, Control and Computing*, 2014.

[3] N. A. M. Yunus, I. A. Halin, N. Sulaiman, N. F. Ismail, and O. K. Sheng, "Valuation on mems pressure sensors and device applications," in *International Conference on MEMS, Nano and Smart Systems*, 2015.

[4] Y. Son, H. Shin, D. Kim, Y. Park, J. Noh, K. Choi, J. Choi, and Y. Kim, "Rocking drones with intentional sound noise on gyroscopic sensors," in *Proceedings of the 24th USENIX Conference on Security Symposium*, 2015.

[5] T. Trippel, O. Weisse, W. Xu, P. Honeyman, and K. Fu, "Walnut: Waging doubt on the integrity of mems accelerometers with acoustic injection attacks," in *IEEE European Symposium on Security and Privacy*, 2017.

[6] G. Chandramohan, "Electrical characterization of mems microphones," Ph.D. dissertation, TU Delft, 2010.

[7] G. Zhang, C. Yan, X. Ji, T. Zhang, T. Zhang, and W. Xu, "Dolphinattack: Inaudible voice commands," in *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security*, 2017.

[8] Britz BR-2900. http://www.britz.co.kr/product.detail.php?category=4&product=63.

[9] Y. Wu, Z. Wei, J. Weng, X. Li, and R. H. Deng, "Resonance attacks on load frequency control of smart grids," *IEEE Transactions on Smart Grid*, 2017.

[10] M. Shahrad, A. Mosenia, L. Song, M. Chiang, D. Wentzlaff, and P. Mittal, "Acoustic denial of service attacks on hdds," *arXiv preprint arXiv:1712.07816*, 2017.