Eavesdropping Speech with Non-sensing Devices

Tim Holzhey
Technische Universität Berlin
Berlin, Germany
holzhey@campus.tu-berlin.de

ABSTRACT

In recent years, numerous research papers have shown that air pressure waves produced by human speech or other sounds can induce vibrations into an array of non-acoustic sensors (e.g. motions sensors) or into externally measured objects (e.g. laser-based vibrometer) skewing sensor readings in a reversable manner, effectively turning them into undisclosed microphones. This allows for eavesdropping on private speech by maliciously altered devices and therefore posing a real threat to privacy when exploited.

This work will examine and compare different types of acoustic vibration-based side channel attacks employed on common IoT and Smart devices to recover speech or infer privacy-sensitive information about the speaker like their identity, political views or gender. We explore the steps taken to take control of the targeted device, gather the necessary data, and perform signal processing and machine learning techniques to extract audible information from the sensor readings. The overview established over the attacks then allows for a comprehensive feasability study for the respective attack methods and complexity required to perform such attacks in a real world scenario. We discuss possible countermeasures to mitigate the risk of such attacks and provide an outlook on future research directions in the field.

CCS CONCEPTS

• Security and privacy → Side-channel analysis and countermeasures; **Embedded systems security**; **• Computer systems organization → Sensors and actuators**.

KEYWORDS

Security, Privacy, Side-channel, Eavesdropping, Speech, Acoustic, Hardware Security, Privacy Leaks

1 INTRODUCTION

While the IoT market is on the rise and still growing exponentially, projected to exceed USD 4 trillion by 2032 [2], this opens up a new attack vector for adversaries to exploit in addition to traditional software vulnerabilities in computers. Latest surveys show that the American households had on average 21 connected devices [1], a relevant part of which are IoT and Smart Home devices. IoT devices are often equipped with a variety of sensors to interface with their physical environment, such as accelerometers, gyroscopes, microphones, and cameras. Many of these sensors can also be found in modern smartphones, which are carried around by most people[1]. Mobile operating systems provide zero-permission access to sensor data from the built-in accelerometer and gyroscope, therefore have been the subject of the majority of research done in this field. The findings from vulnerabilities found in smartphones can be projected onto IoT and Smart devices with similar sensors that do not have a primary function of audio recording i.e. do not have a built-in microphone ("non-sensing"). To execute a vibration-based eavesdropping attack, most of the previous papers took the approuch to exploited MEMS[2] motion sensors (accelerometers, gyroscopes and magnetometers) commonly found in smartphones and many smart devices including smartwatches, fitness trackers, gaming controllers, etc. Some of the more experimental approaches have also shown that other sensors like Lidar scanners in vacuum cleaners, the position error signal of write heads in hard drives or electro-optical sensors directed at ceiling lights can be exploited for similar attacks.

CONTRIBUTION

Altough parts of the available research material in this field is investigating keystroke recovery attacks [?][?][?] or is using sophisticated external setups (e.g. RFID-Tags [?], millimeter-waves [?], WiFi radio [?]), we limit the scope of this paper to on-device vibration-based speech and general sound recovery attacks. This includes attacks in theory possible without any modified or additional hardware assuming a compromised device or malicious software. This work aims to provide a comprehensive overview of the current state of research in the field of vibration-based eavesdropping attacks on non-sensing devices. We highlight notable research papers and their findings, compare the different attack methods and achieved results, and discuss the feasibility of such attacks in real-world scenarios.

2 BACKGROUND

2.1 Vibration-based Eavesdropping Attacks

Sound created by a human speaking or any other sound can be characterized as spatially and temporally propagating changes in air pressure in the audible frequency range (20 Hz - 20 kHz). Similarly to how sound waves induce vibrations into our eardrums to let us perceive sound, they can also couple vibrations into all other objects they encounter, more so into objects that are resonant at the frequency of the sound. In a typical microphone, an oscillating diaphragm is used to convert these vibrations into an electrical signal i.e. a change in voltage by varying the capacitance of a capacitor (condenser microphone) or by inducing a current into a coil (dynamic microphone). Even if unintended, the same phenomenon can be used to turn any other sensing electrical component into a microphone if it has a moving part capable of influencing the electrical properties of the component directly (e.g. MEMS, write

[1]Surveys from 2024 suggest that 91 % of Americans own a smartphone

[2]Abbr. Micro-electromechanical systems [3]

head of a hard drive) or observing the movement of another object (e.g. laser vibrometer, Lidar scanner, camera).

2.2 Previous Work

2.3 MEMS-based Eavesdropping Attacks

2.4 Laser-based Eavesdropping Attacks

2.5 Other Eavesdropping Attacks

3 THREAT MODEL

4 SPEECH RECONSTRUCTION

4.1 Data Collection

4.2 Signal Processing

4.3 Machine Learning

4.4 Automated Speech Recognition

5 FEASIBILITY STUDY

6 COUNTERMEASURES

7 CONCLUSION

REFERENCES

[1] PRNewswire 2023. *Deloitte: The Connected Consumer Paradox - Desire for Fewer Devices vs. More Virtual Experiences and Technology Innovation.* PRNewswire. Retrieved January 07, 2025 from https://www.prnewswire.com/news-releases/deloitte-the-connected-consumer-paradox---desire-for-fewer-devices-vs-more-virtual-experiences-and-technology-innovation-301919928.html

[2] Fortune Business Insights 2024. *Internet of Things [IoT] Market Size, Share, Growth, Trends, 2032.* Fortune Business Insights. Retrieved January 07, 2025 from https://www.fortunebusinessinsights.com/industry-reports/internet-of-things-iot-market-100307

[3] Pew Research Center 2024. *Mobile Fact Sheet.* Pew Research Center. Retrieved January 07, 2025 from https://www.pewresearch.org/internet/fact-sheet/mobile/

[4] S Abhishek Anand and Nitesh Saxena. 2018. Speechless: Analyzing the Threat to Speech Privacy from Smartphone Motion Sensors. In *2018 IEEE Symposium on Security and Privacy (SP).* 1000–1017. https://doi.org/10.1109/SP.2018.00004

[5] S Abhishek Anand, Chen Wang, Jian Liu, Nitesh Saxena, and Yingying Chen. 2021. Spearphone: a lightweight speech privacy exploit via accelerometer-sensed reverberations from smartphone loudspeakers. In *Proceedings of the 14th ACM Conference on Security and Privacy in Wireless and Mobile Networks* (Abu Dhabi, United Arab Emirates) *(WiSec '21).* Association for Computing Machinery, New York, NY, USA, 288–299. https://doi.org/10.1145/3448300.3468499

[6] Zhongjie Ba, Tianhang Zheng, Xinyu Zhang, Zhan Qin, Baochun Li, Xue Liu, and Kui Ren. 2020. Learning-based Practical Smartphone Eavesdropping with Built-in Accelerometer. https://doi.org/10.14722/ndss.2020.24076

[7] Jun Han, Albert Jin Chung, and Patrick Tague. 2017. PitchIn: eavesdropping via intelligible speech reconstruction using non-acoustic sensor fusion. In *Proceedings of the 16th ACM/IEEE International Conference on Information Processing in Sensor Networks* (Pittsburgh, Pennsylvania) *(IPSN '17).* Association for Computing Machinery, New York, NY, USA, 181–192. https://doi.org/10.1145/3055031.3055088

[8] Andrew Kwong, Wenyuan Xu, and Kevin Fu. 2019. Hard Drive of Hearing: Disks that Eavesdrop with a Synthesized Microphone. In *2019 IEEE Symposium on Security and Privacy (SP).* 905–919. https://doi.org/10.1109/SP.2019.00008

[9] Yunji Liang, Yuchen Qin, Qi Li, Xiaokai Yan, Zhiwen Yu, Bin Guo, Sagar Samtani, and Yanyong Zhang. 2022. AccMyrinx: Speech Synthesis with Non-Acoustic Sensor. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 3, Article 127 (Sept. 2022), 24 pages. https://doi.org/10.1145/3550338

[10] Richard Matovu, Isaac Griswold-Steiner, and Abdul Serwadda. 2019. Kinetic Song Comprehension: Deciphering Personal Listening Habits via Phone Vibrations. *CoRR* abs/1909.09123 (2019). arXiv:1909.09123 http://arxiv.org/abs/1909.09123

[11] Yan Michalevsky, Dan Boneh, and Gabi Nakibly. 2014. Gyrophone: recognizing speech from gyroscope signals. In *Proceedings of the 23rd USENIX Conference on Security Symposium* (San Diego, CA) *(SEC'14).* USENIX Association, USA, 1053–1067.

[12] Sriram Sami, Sean Rui Xiang Tan, Yimin Dai, Nirupam Roy, and Jun Han. 2020. LidarPhone: acoustic eavesdropping using a lidar sensor: poster abstract. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems* (Virtual Event, Japan) *(SenSys '20).* Association for Computing Machinery, New York, NY, USA, 701–702. https://doi.org/10.1145/3384419.3430430

[13] Weigao Su, Daibo Liu, Taiyuan Zhang, and Hongbo Jiang. 2022. Towards Device Independent Eavesdropping on Telephone Conversations with Built-in Accelerometer. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 4, Article 177 (Dec. 2022), 29 pages. https://doi.org/10.1145/3494969

[14] Lei Wang, Meng Chen, Li Lu, Zhongjie Ba, Feng Lin, and Kui Ren. 2023. VoiceListener: A Training-free and Universal Eavesdropping Attack on Built-in Speakers of Mobile Devices. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 1, Article 32 (March 2023), 22 pages. https://doi.org/10.1145/3580789

[15] Qingsong Yao, Yuming Liu, Xiongjia Sun, Xuewen Dong, Xiaoyu Ji, and Jianfeng Ma. 2024. Watch the Rhythm: Breaking Privacy with Accelerometer at the Extremely-Low Sampling Rate of 5Hz. In *Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security* (Salt Lake City, UT, USA) *(CCS '24).* Association for Computing Machinery, New York, NY, USA, 1776–1790. https://doi.org/10.1145/3658644.3690370

[16] Li Zhang, Parth H. Pathak, Muchen Wu, Yixin Zhao, and Prasant Mohapatra. 2015. AccelWord: Energy Efficient Hotword Detection through Accelerometer. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services* (Florence, Italy) *(MobiSys '15).* Association for Computing Machinery, New York, NY, USA, 301–315. https://doi.org/10.1145/2742647.2742658

[17] Shijia Zhang, Yilin Liu, and Mahanth Gowda. 2023. I Spy You: Eavesdropping Continuous Speech on Smartphones via Motion Sensors. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 4, Article 197 (Jan. 2023), 31 pages. https://doi.org/10.1145/3569486

Received 7 January 2025

Table 1: Test parameters and key results from previous publications on vibration-based speech recovery attacks exploiting different sensors

Year	Paper	Sensor	Attack Goal	Sampling Freq. (max)	Audio source	Transmission Medium	Distance from source	Dictionary Size	Accuracy (best)
2014	Gyrophone [11]	Gyroscope	Speech Recognition, Speaker Identification, Gender Identification	200 Hz	External Loudspeaker	Solid Surface	10 cm	11 digits	26 %
2015	AccelWorld [16]	Accelerometer	Speech Recognition, Speaker Identification	200 Hz	External Loudspeaker	Air	30 cm	1 hotword	85 %
2017	PitchIn [7]	Accelerometer, Gyroscope, Geophone	Speech Recognition	1 kHz	Human	Air	1 m	10 words	79 %
2018	Speechless [4]	Accelerometer, Gyroscope	Speech Recognition	8 kHz	External Loudspeaker	Solid Surface	10 cm	10 digits	0 %
2019	Kinetic Song Comprehension [10]	Accelerometer, Gyroscope	Song Recognition	100 Hz	Smartphone Loudspeaker	Solid Surface	On-Device	100 songs	80 %
2020	AccelEve [6]	Accelerometer	Speech Recognition, Speaker Identification	500 Hz	Smartphone Loudspeaker	Solid Surface	On-Device	8 hotwords	90 %
2021	Spearphone [5]	Accelerometer	Speech Recognition, Speaker Identification, Gender Identification	500 Hz	Smartphone Loudspeaker	Solid Surface	On-Device	58 words	67 %
2021	Vibphone [13]	Accelerometer	Speech Recognition	170 Hz	Smartphone Loudspeaker	Solid Surface	On-Device	10 hotwords + 10 digits	54.2 %
2022	AccMyrinx [9]	Accelerometer	Speech Recognition	500 Hz	Smartphone Loudspeaker	Solid Surface	On-Device	Synthesis	57.33 %
2023	ISpyU [17]	Accelerometer, Gyroscope	Continuous Speech Recognition	500 Hz	Smartphone Loudspeaker	Solid Surface	On-Device	9950 words	53.3 %
2023	VoiceListener [14]	Accelerometer, Gyroscope, Magnetometer	Speech Recognition	250 Hz	Smartphone Loudspeaker	Solid Surface	On-Device	10 digits	82.7 %
2024	Watch the Rhythm [15]	Accelerometer	Speech Recognition	200 Hz	Smartphone Loudspeaker	Solid Surface	On-Device	10 digits	77.79 %
2020	Lidarphone [12]	Lidar Scanner	Speech Recognition, Speaker Identification, Gender Identification	1.8 kHz	External Loudspeaker	Air	1.5 m	10 digits	91 %
2019	Hard Drive of Hearing [8]	Hard Drive PES	Speech Recognition	34.56 kHz	External Loudspeaker	Air	25 cm	-	-