Eavesdropping Speech with Non-sensing Devices

Tim Holzhey
Faculty of Electrical Engineering and Computer Science,
Technische Universität Berlin
Berlin, Germany
holzhey@campus.tu-berlin.de

ABSTRACT

In recent years, numerous research papers have shown that sound produced by human speech or other sources can induce minuscule vibrations into a variety of non-acoustic sensors (e.g. motion sensors) or into externally measured objects (e.g. vibrometer) influencing sensor readings in a reversible manner, effectively turning them into undisclosed microphones. This allows for eavesdropping on private speech by compromised unsuspicious devices and therefore posing a real threat to privacy when exploited.

This work will examine and compare different types of vibration-based eavesdropping attacks using a side channel 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. By highlighting notable previous research in this field, we explore the steps necessary 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 experimental setups and findings of 16 research papers published in the last decade are compared and discussed, to provide a timeline and trend of the development, limitations, and advancements. The overview established over the attacks then allows for a comprehensive feasibility 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

The IoT market is on the rise and is still growing exponentially - projected to exceed USD 4 trillion by 2032 [8] - this opens up a new attack vector for adversaries to exploit in addition to traditional software vulnerabilities in computers. The latest surveys show that American households had on average 21 connected devices [5], of which a significant part 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 on modern smartphones and tablets, which most people carry around[1]. Devices running on *Android* and alike 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 many IoT and smart devices that are equipped with similar sensors but do not have a primary function of audio recording, i.e., do not have a built-in microphone (in that regard "non-sensing"). To execute a vibration-based eavesdropping attack, most of the previous papers took the approach to exploit motion sensors implemented using micro-electromechanical systems (MEMS accelerometers, gyroscopes and magnetometers) commonly found in smartphones and many smart devices like smartwatches, fitness trackers or gaming controllers. 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 AND LIMITATIONS: Although a significant part of research done in this field is investigating keystroke recovery attacks [17][19][41], device fingerprinting [27][15] or is using sophisticated external setups (e.g. RFID-Tags [22], millimeter-waves [21], WiFi radio [39]), we limit the scope of this paper to **on-device vibration-based speech and general sound eavesdropping 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 development and current state of research (SOTA attacks) 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 and possible countermeasures of such attacks in real-world scenarios.

2 BACKGROUND

2.1 Vibration-based Eavesdropping Side-Channel 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

[1] Surveys from 2024 suggest that 91 % of Americans own a smartphone [10]

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 the circuit (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 head of a hard drive) or observing the spaciotemporal properties of another object (e.g. laser vibrometer, Lidar scanner, camera). As audible information was not intended to be captured by these sensors, an attacker who is able to recover this information from the sensor readings is exploiting a side channel vulnerability.

2.2 MEMS

Sensors manufactured using micro-electromechanical systems (MEMS) incorporate electronics and moving parts on a micrometer-scale chip to measure physical parameters like acceleration (accelerometer), orientation and angular velocity (gyroscope) or the magnetic field (magnetometer). The fabrication process makes use of semiconductor manufacturing techniques including lithography and etching on silicon wafers that allows for the production of small, low-cost sensors with high sensitivity and accuracy. They are widely used in consumer electronics to enable features like screen rotation, step counting, navigation and gaming feedback. On a physical level, MEMS sensors are most commonly realized by a spring-suspended proof mass that changes the capacitance of the circuitry when displaced (variable capacitance MEMS) or by a flexible piezoelectric material that changes its electrical resistance when bent (piezoresistive MEMS). The structures can be repeated and aligned in three orthogonal directions to measure the physical property in the three-dimensional spacial domain.

MEMS Accelerometer: An accelerometer measures the proper acceleration (change in velocity) of an object relative to a local inertial reference frame. In the gravitational field of the earth, the accelerometer's measurement is offset by the upwards acceleration of 1 g (9.81 m/s^2) relative to the free-falling reference frame. The basic mechanical structure of an accelerometer consists of a damped proof mass suspended by springs that is displaced when the sensor is accelerated in the opposite direction of movement. In a typical VC MEMS accelerometer, the proof mass moves between air-gapped fixed electrodes forming a variable capacitor as shown in Figure 1.

MEMS Gyroscope: A gyroscope measures the angular velocity (rate of rotation) of an object relative to a local inertial reference frame. Gyroscopes realized as a MEMS sensor are commonly Vibrating structure gyroscopes (VSG) that measure the Coriolis force acting on a vibrating proof mass when the sensor is rotated. As the vibrating mass tends to continue vibrating in the same plane, the Coriolis force deflects the mass in the direction perpendicular to the rotation axis. The deflection is measured by capacitive sensing or piezoresistive sensing and is proportional to the angular velocity of the rotation as shown in Figure 2.

MEMS Magnetometer: A magnetometer measures the strength and direction of the local magnetic field. MEMS-based magnetometers are often implemented using the Lorentz force acting on the current-carrying conductor in the magnetic field to move the mechanical structure. The displacement is then measured by capacitive, piezoresistive or optical sensing and is proportional to the magnetic field strength.

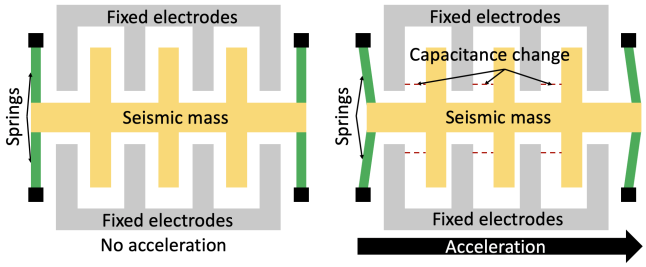

Figure 1: Accelerometer - Capacitive MEMS structure, Source: *AccelEve* 2020 [16]

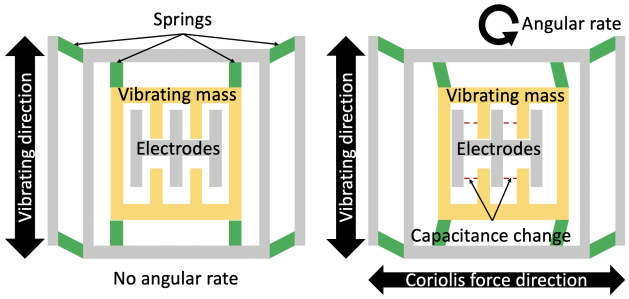

Figure 2: Gyroscope - Capacitive MEMS structure, Source: *AccelEve* 2020 [16]

2.3 Laser-based Sensors

Laser-based measurement devices (e.g. LiDAR scanners, interferometers, vibrometers) are used in various applications to measure precise distances, velocities and material properties contactlessly in various scientific, industrial and medical applications, but recently they made their way into consumer electronics. Prominently, LiDAR scanners consolidate themselves in our daily lifes as they are increasingly used in applications like autonomous vehicles (Waymo Self-Driving [4]), drones (DJI UAVs [25]), robotic vacuum cleaners, *Apple FaceID* or Augumented Reality enabled smartphones.

LiDAR Scanner: LiDAR (Light Detection and Ranging) is a remote sensing and imaging technique that uses a pulsed laser beam to measure the distance to a target object by measuring the time it takes for the light to reflect back to the sensor (Time of Flight) with known speed of light ($d = \frac{c \cdot t}{2}$). LiDAR systems can map the environment in all directions by rotating the laser beam in a horizontal plane (mechanical spinning LiDAR, Figure 3) or by other techniques (solid-state MEMS, optical phased array, Flash LiDAR) and measuring the distance at different angles. The system can be further extended to a 3D LiDAR by adding more vertical scanning layers. The cummulative distance measurements can be used to

create a point cloud representation of the environment. In order to not interfere with other optical sensors (e.g. camera, human eye), LiDAR's wavelength is mainly located in the near-infrared part of the electromagnetic spectrum (750 nm to 1.5µm).

Laser Doppler Vibrometer: A laser microphone uses a laser beam to detect sound vibrations in a distant object. The minute differences in the distance traveled by the light as it reflects from the vibrating object are detected interferometrically. The Laser Doppler Vibrometer (LDV) implements this principle of laser interferometry by splitting the laser into two beams, one of which is reflected off the vibrating object. The surface will module the phase and frequency of the light due to the Doppler effect. One of the beams is passed through a Bragg cell (acousto-optic modulator) to add a frequency shift and then recombined with the other beam to be directed to a photodetector (Figure 4). The electrical signal produced by the photodetector is equal to the carrier frequency produced by the Bragg cell modulated by the Doppler frequency of the vibrating object and proportional to the velocity of the object.

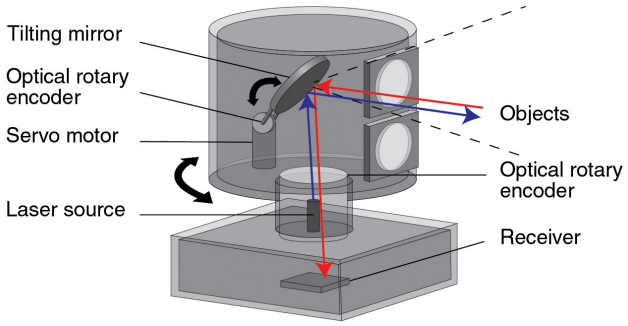

Figure 3: Mechanical spinning LiDAR [2]

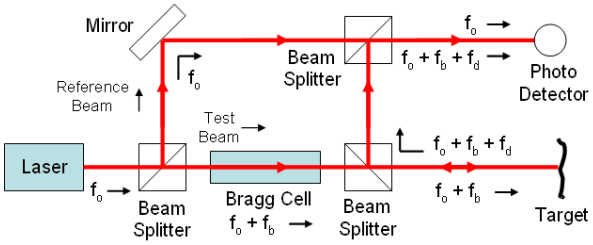

Figure 4: Laser Doppler Vibrometer [9]

2.4 Speech Intelligibility

A human speeking in a non-tonal language like English produces a complex waveform that is composed of various frequencies in the audible range. While the fundamental frequency f_0 (lowest frequency component) of the human voice is typically in the range of 100 Hz to 300 Hz (higher for women and children), overtones and consonant articulations can cover most of the audible frequency range of up to 17 kHz (Figure 5). Research has shown that frequencies between 1 kHz and 4 kHz are most important for speech intelligibility by surveying with masking band-pass filters [18]. Applying a low-pass filter to the speech signal at 1 kHz and below quickly degrades the intelligibility of the speech to near zero as perceived by humans (Figure 6). Usual digital audio recording and playback systems operate at a sampling rate of 44.1 kHz or 48 kHz to capture the full audible frequency range. Since most experiments conducted using motion sensors are limited to a sampling rate of 100-500 Hz, machine learning techniques have to be employed to recover (fill in) frequencies above the Nyquist frequency $f_N = \frac{1}{2} f_s$ that are essential for speech intelligibility but not encoded in the sensor readings directly. For intelligible speech it has been shown, a minimum peak signal-to-noise ratio (PSNR) of 25 - 30 is required [18].

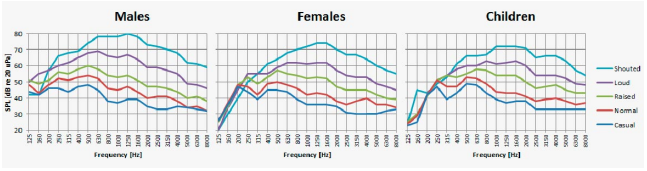

Figure 5: Frequency spectrum of a human voice for Males, Females and Children [18]

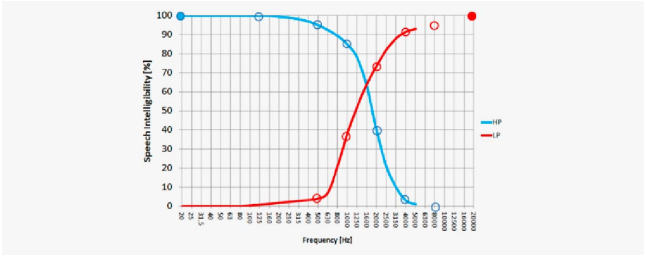

Figure 6: Speech intelligibility with low-pass and high-pass filters applied at various frequencies [18]

3 LITERATURE REVIEW

3.1 MEMS-based Eavesdropping Attacks

Although MEMS sensors are designed to be best insensitive to acoustic noise which would degrade their signal-to-noise ratio, they are still susceptible to sound waves that induce vibrations in the sensor structure. A MEMS-based eavesdropping attack exploits this vulnerability by recovering the sound-induced vibrations from the sensor readings and reconstructing the original sound. These types of attacks need to overcome several challenges that include the low sampling rate of the sensors (order of magnitude 100 Hz - 500 Hz), the poor signal-to-noise ratio (PSNR $\ll$ 25) and aliasing artifacts.

Early Work: The first paper able to demonstrate the feasibility of recovering speech from motion sensors in typical smartphones was *Gyrophone: Recognizing Speech from Gyroscope Signals* [32] in 2014 by Yan Michalevsky *et al.*, a joint effort of researchers from Stanford University and the National Research & Simulation Center (Rafael Advanced Defense Systems Ltd., Isreal). The authors showed that a

smartphone's gyroscope can be used to recover speech rendered by a nearby loudspeaker using sensor readings at a well below Nyquist sampling rate of 200 Hz. An Android app was developed to record the gyroscope data without requiring any special permissions. Afterwards, silcence removal and segmentation was applied to the signal that was then fed to the off-the-shelf *Sphinx* speech recognition system to recognize spoken digits of the *TIDIGITS* dataset using spectral statistical features. The authors also trained custom machine learning classifiers (SVM, GMM, DTW) in *Matlab* to identify the speaker and their gender. With a limited dictionary of spoken english digits (0-9), no assumptions about the speaker and a close distance of 10 cm between the loudspeaker and the smartphone on a solid table surface, a moderate recognition accuracy of at most 26 % was achieved. Still, they demonstrated speaker classification of up to 50 % and gender classification of up to 84 %. This opened up the field of research to many more papers to come building upon and considerably improving these types of attacks, primarily due to advances in machine learning.

Recent Work: Notably in 2023, Shijia Zhang *et al.* from The Pennsylvania State University leveraged speech eavesdropping attacks using motion sensors in their paper *I Spy You: Eavesdropping Continuous Speech on Smartphones via Motion Sensors* [44] to, for the first time, provide full continuous speech recognition by jointly using accelerometer and gyroscope data. With a large dictionary of 9950 words and a custom ASR (Automatic Speech Recognition) deep learning model, they achieved 53.3 % accuracy in recognizing spoken words on an Android smartphone at a sampling rate of 200-500 Hz. Previously, speech-related attacks were limited to classifying words from a small dictionary (e.g. dataset of 10 spoken english digits) [32][13][26][38] or recognizing a single hotword in a stream of words (e.g. dataset of 8 cities) [43][16][35].

Most recently in 2024, Qingsong Yao *et al.* from Xidian University, China published the paper *Watch the Rhythm: Breaking Privacy with Accelerometer at the Extremely-Low Sampling Rate of 5Hz* [42] in which they demonstrated that a smartphone's accelerometer can be used for eavesdropping attacks even when limited to a sampling rate of just 5 Hz. This was achieved by extending the machine learning algorithms to not only consider time-frequency features (spectral) but also the temporal dynamics of the signal (pause rhythm and energy intensity rhythm) that are much better preserved in the very-low frequency domain. To benchmark against previous papers, the authors showed that english spoken digits could be recognized with an accuracy of 32.70 % at 5 Hz with an on-device Android app reading accelerometer data (77.79 % at 200 Hz). Similar tests to determine typical places like bar, metro, bus and car yielded an average accuracy of 91.28 %.

3.2 Laser-based Eavesdropping Attacks

Recent Work: In 2020, a group of researchers from the University of Singapore and the University of Maryland sparked the interest of the security research community [23] and news outlets [40] with their paper *Spying with Your Robot Vacuum Cleaner: Eavesdropping via Lidar Sensors* ("LidarPhone") [34]. A method is introduced that repurposes lidar sensors in robot vacuum cleaners to function as laser-based microphones capable of capturing sound signals by detecting subtle vibrations in nearby objects (Figure 7). After reverse-engineering and modifying the firmware of a commercial robot vacuum cleaner (Xiaomi Roborock S5) to gain access to the raw lidar data, the authors tricked the device into activating the scanner without rotating its mirror and unknowingly duplicating the sensor data stream to be sent over the local network. Therefore, they could increase the sampling rate of a single point from 5 Hz to 1.8 kHz and later process the data on a remote computer offline. The robot's lidar scanner was aimed at different kinds of common household objects (e.g. trash can, cardboard box, plastic bag) and sounds were played back from a nearby loudspeaker. With an array of preprocessing steps (DC-offset correction, outlier removal, interpolation, normalization, high-pass filter, noise removal and equalization) the small signal-to-noise ratio of the sensor readings could be overcome and the authors were able to successfully identify spoken digits, the speaker's identity and gender and music snippets from popular news channels introductory jingles with accuracies of 91 %, 67.5 %, 96 % and 90 % respectively using classifying and correlating convolutional neural networks (CNN).

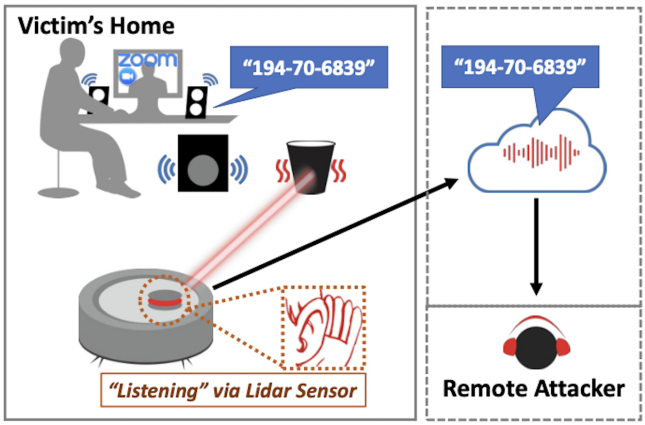

Figure 7: Experimental setup of the *LidarPhone* attack [34]

3.3 Other Eavesdropping Attacks

Other notable vibration-based speech eavesdropping side channel attacks explore methods to turn computer hardware and household objects into microphones.

The paper *Hard Drive of Hearing: Disks that Eavesdrop with a Synthesized Microphone* (2019) [29] by Andrew Kwong *et al.* from the University of Michigan and Zhejiang University demonstrated that the position error signal (PES) of the read/write head in a hard drive can be used to recover sound signals if subjected to sound vibrations. Although not exposed to the user, if the hard drive controller's firmware is compromised, the high-fidelity PES (16-bit, 34.56 kHz) can fully encode the recorded sound signal without the need for machine learning assistance. The authors put their experimental microphone to the test by reliably detecting music played from a nearby smartphone with the *Shazam* music recognition app fed with the recorded and processed PES data.

Other noteworthy research has focused on recovering speech with an electro-optical sensor directed at a ceiling light through a telescope (*Lamphone* [33]) which is not further discussed here.

Table 1: Test parameters and key results from a timeline of previous publications on vibration-based speech eavesdropping attacks exploiting different sensors and devices

Year	Paper	Sensor	Measuring Device	Attack Goal	Sampling Freq.	Audio source	Transmission Medium	Distance from source	Dictionary Size	Speech Rec. (best)
2014	Gyrophone [32]	Gyroscope	Android Smartphone	Speech Rec., Speaker Ident., Gender Ident.	200 Hz	External Loudspeaker	Solid Surface	10 cm	10 digits	26 %
2015	AccelWorld [43]	Accelerometer	Android Smartphone	Speech Rec., Speaker Ident.	200 Hz	External Loudspeaker	Air	30 cm	1 hotword	85 %
2017	PitchIn [28]	Accelerometer, Gyroscope, Geophone	Dedicated IMCU	Speech Rec.	1 kHz	Human	Air	1 m	10 words	79 %
2018	Speechless [13]	Accelerometer, Gyroscope	Android Smartphone	Speech Rec.	200 Hz	External Loudspeaker	Solid Surface	10 cm	10 digits	0 %
2019	Kinetic Song Comprehension [31]	Accelerometer, Gyroscope	Android Smartphone	Song Rec.	100 Hz	Smartphone Loudspeaker	Solid Surface	On-Device	100 songs	80 %
2020	AccelEve [16]	Accelerometer	Android Smartphone	Speech Rec., Speaker Ident.	100-500 Hz	Smartphone Loudspeaker	Solid Surface	On-Device	8 hotwords, 36 alphanum.	90 % (hotwd), 55 % (alnum), 78 % (digits)
2021	Spearphone [14]	Accelerometer	Android Smartphone	Speech Rec., Speaker Ident., Gender Ident.	120-500 Hz	Smartphone Loudspeaker	Solid Surface	On-Device	10 digits, 58 words	67 % (words), 71 % (digits)
2021	Vibphone [35]	Accelerometer	Android Smartphone	Speech Rec.	225-425 Hz	Smartphone Loudspeaker	Solid Surface	On-Device	10 hotwords + 10 digits	54.2 %
2022	AccMyrinx [30]	Accelerometer	Android Smartphone	Speech Rec.	100-500 Hz	Smartphone Loudspeaker	Solid Surface	On-Device	Synthesis	42.7 % SWER[2]
2022	InertiEAR [26]	Accelerometer, Gyroscope	Smartphone	Speech Rec.	40-200 Hz	Smartphone Loudspeaker	Solid Surface	On-Device	10 digits	78.8 %
2023	ISpyU [44]	Accelerometer, Gyroscope	Android Smartphone	Automatic Speech Rec.	200-500 Hz	Smartphone Loudspeaker	Solid Surface	On-Device	9950 words, 2000 words	53.3 % (big), 59.9 % (small)
2023	VoiceListener [38]	Accelerometer, Gyroscope, Magnetometer	Android Smartphone	Speech Rec.	100-500 Hz	Smartphone Loudspeaker	Solid Surface	On-Device	10 digits, 10 sentences	82.7 % (digits) 10 % (senten)
2023	StealthyIMU [36]	Accelerometer	Android Smartphone	Speech Rec. (Voice Assistant)	100-500 Hz	Smartphone Loudspeaker	Solid Surface	On-Device	23 voice commands	8.46 % SEER[3]
2024	Watch the Rhythm [42]	Accelerometer	Android Smartphone	Speech Rec., Scene Rec.	5-200 Hz	Smartphone Loudspeaker	Solid Surface	On-Device	10 dig. (en+ch), 28 scenes, 30 cities	51 % (dig-en), 47 % (dig-ch), 78.7 % (scene), 58.7 % (cities)
2020	LidarPhone [34]	Lidar Scanner	Robot Vacuum Cleaner	Speech Rec., Song Rec., Speaker Ident., Gender Ident.	1.8 kHz	External Loudspeaker	Air	1.5 m	10 digits, 10 snippets	91 % (digits), 90 % (snippets)
2019	Hard Drive of Hearing [29]	Hard Drive PES	HDD Controller	Speech Rec., Song Rec.	34.56 kHz	External Loudspeaker	Air	25 cm	-	Verified, no metric

4 FEASIBILITY STUDY

In the following, 16 of the most cited research papers on device-local vibration-based speech eavesdropping attacks exploiting different sensors and devices published between 2014 and 2024 are compared giving comprehensive insight over the research done in the field from its early pioneers to the very recent advances [32][43][28][13][31][16][14][35][30][26][44][38][36][42][34][29]. The test parameters used in the experimental setups of the papers and the key results the authors could achieve (metric for best speech recognition accuracy) are compiled in Table 1 in chronological order of publication grouped by the type of sensor the data was collected from. Most work focus on MEMS-based attacks using accelerometers, gyroscopes and magnetometers in zero-permission Android or HarmonyOS smartphones, while two papers explore high-fidelity laser-based and miscellaneous attacks for comparison.

It can be observed that attacks using sensors in smartphones are heavily restricted by the low sampling rate, usually in the realm of 100-500 Hz. This value is inherently hardware-bound by the sensor circuitry and hardware access overhead, and often times deliberately limited by the operating system to prevent privacy breaches as well as excessive power consumption. Still, many papers have shown that strikingly high speech recognition accuracies can be achieved with very low sampling rates. The development towards using larger word dictionaries (9950 words [44]), using a combination of different sensors in conjunction (ACC x GYRO [26]), proving the privacy concerns still exist with even lower sampling rates (5 Hz [42]) and experimenting with more sophisticated processing and machine learning pipelines is evident over the years. Noteworthy is also the achievement of continuous speech recognition in 2023 by *ISpyU* [44] with more than 50 % word accuracy and almost 10 thousand distinct recognizable words.

4.1 Limitations

The results from previous research all mention that the extremely low signal-to-noise ratio (low amplitude signal in relation to sensor

noise) is a limiting factor. To produce meaningful speech recognition the sound-producing source (generally a loudspeaker or subwoofer with 70+ dB) needs to be close and well sound-conducted to the sensor over a solid surface. In many cases the speaker even shares the same housing with the sensor e.g. in the case of a smartphone that shares both components on one PCB (< 10 cm copper and resin). Comparative studies have shown accelerometers pick up sounds significantly better than gyroscopes [16] because sound waves generally produce linear acceleration in the direction of travel rather than angular velocity. The detailed meta analysis conducted in *SoK: Assessing the Threat Potential of Vibration-based Attacks against Live Speech using Mobile Sensors* [37] by Payton Walker *et al.* assessing the real-world thread potential of some of the here mentioned attack methods concludes that the likelihood of a successful attack is fairly low due to the many limiting factors and the need for a controlled testing environment with favorable conditions.

5 COUNTERMEASURES

In the following, we will discuss possible countermeasures to mitigate the risk of vibration-based eavesdropping attacks on non-sensing devices.

5.1 Limit Sensor Quality

One way to lower the success rate of speech recovery attacks is to limit quality metrics of the sensor data exposed to the user. This can be achieved by reducing the sampling rate of the sensors further (e.g. to 50 Hz), lowering the resolution (bit depth) or by worsening the signal-to-noise ratio deliberately. However, this would also degrade the performance of the device's primary function and user experience (e.g. slower response to phone rotation in mobile games, aliasing artifacts in video stabilization or faults in lidar mapping).

In the past, mobile operating systems have updated their APIs to limit the sampling rate of motion sensors to 200 Hz (Android [12]) and 100 Hz (iOS [6][7]) in response to privacy concerns. While substantially limiting malicious apps from recording high-fidelity audio with motion sensors, research has shown that speech still can be leaked at very low sampling rates (e.g. 5 Hz [42]) to some extent.

Smartphone sensors are also exposed to web applications by the *Sensor Web API* [3] available in many modern web browsers (Chrome, Edge, Opera). The World Wide Web Consortium (W3C) has acknowledged the privacy concerns of exposing raw inertial sensor data to the web content and have proposed an *Accelerometer reading quantization algorithm* [24] that limits the resolution of the accelerometer's XYZ data points to a fixed $0.1 m/s^2$ granularity that is still sufficient for many practical applications. The W3C Editor's Draft [24] reads:

> **§ 4. Security and Privacy Considerations**
> Sensor readings provided by inertial sensors, such as accelerometer, could be used by adversaries to exploit various security threats, for example, key-logging, location tracking, fingerprinting and user identifying.

5.2 Constrain Sensor Access

Another way to lower the chance of a malicious app recording audio with motion sensors is to further constrain the access to the sensor data. While iOS is already strictly enforcing obligatory prompts for developers that tell the user why the app is requesting access to the device's motion data (iOS 7.0+ NSMotionUsageDescription [1]), Android is granting sensor access to apps by default for up to 200 Hz. Apps need to declare the *HIGH_SAMPLING_RATE_SENSORS* permission only to be able to sample the device's accelerometer, gyroscope, and geomagnetic field sensor at higher rates (Android 12+ [12]). Since Android 9 apps also cannot access sensors when running in background and users are able to manage permission preferences for each app individually in the settings. Google has acknowledged the privacy concerns of sensor-based eavesdropping attacks in the Android Developer's documentation [12] by linking sensor rate-limiting to the microphone access:

> **Sensor Rate-Limiting**
> [...]
> **Note:** If the user turns off microphone access using the device toggles, the motion sensors and position sensors are always rate-limited, regardless of whether you declare the HIGH_SAMPLING_RATE_SENSORS permission.

IoT devices should also introduce strict hardware-interlocks that e.g. would have prevented the robot vacuum cleaner' LiDAR investigated in *LidarPhone* [34] from being activated without rotating the mirror.

5.3 Minimize Acoustic Coupling

Motion sensors are designed to minimize the acoustic coupling to the sensor structure. This can be achieved by isolating the sensor from the sound source (e.g. by using a soundproof casing) or by damping the sensor structure in its embedding (e.g. by using dampening pads around the sensor's suspension). Other proposals include acoustic masking by e.g. introducing white noise into the sensor or band-pass filtering the sensor readings to remove speech-relevant frequencies.

5.4 Preprocess Audio Data

Recently, the authors of *EveGuard: Defeating Vibration-based Side-Channel Eavesdropping with Audio Adversarial Perturbations* [20] proposed a software-driven defense mechanism which protects voice privacy by introducing adversarial perturbations into the audio signal produced by a loudspeaker. They develop a perturbation generator model (PGM) that effectively suppresses sensor-based eavesdropping by introducing inaudible distortions while maintaining high audio quality using machine learning techniques. This audio preprocessing step could be integrated into smartphones, effectively reducing the ability of the motion sensors to recover sound from the speaker-induced vibrations.

5.5 Software Security

Since all eavesdropping attacks previously discussed require the attacker to have control over the device or to have installed malicious software on the device, manufacturers and developers need

to actively design their software and firmware to be secure against unauthorized access. For embedded and IoT devices, this concerns implementing mechanisms like secure boot, signed firmware update, encrypted data storage and locked-down debug interfaces. Operating systems offering app stores should also enforce strict security policies for apps that request access to sensitive data like motion sensor data. Special attention should be given to the Android platform, as it is possible to side-load apps from third-party sources without the need to be verified by the Google Play Protect service. The Android OS is globally being used by the vast majority of smartphones and tablets (market share of 73.52 % in 12/2024 [11]) and is therefore a prime target for malicious software.

6 CONCLUSION

In this work, we have provided a comprehensive overview of the state-of-the-art research in the field of vibration-based speech eavesdropping attacks using non-acoustic sensors. We have highlighted notable research papers and their findings, compared the different attack methods and achieved results, and discussed the feasibility and possible countermeasures of such attacks in real-world scenarios. The findings show that speech recovery is likely to be very limited in real-world environments due to multiple restricting factors including very low sampling rates, poor signal-to-noise ratios, the need for close proximity to the sound source and solid sound-conducting transmission medium. Fortunately for the users, no large-scale privacy breach of this kind has been reported to have been exploited in real-world applications yet. However, analysis show indications of a trend in research to overcome these limitations by using more sophisticated machine learning techniques, sensor fusion and temporal dynamics in conjunction with spectral features. Although limited, the risk of eavesdropping attacks is still present and manufacturers, developers and users should be made aware of the potential threat and take appropriate countermeasures to protect their privacy looking forward.

REFERENCES

[1] Apple Developer [n. d.]. *NSMotionUsageDescription*. Apple Developer. Retrieved January 07, 2025 from https://developer.apple.com/documentation/BundleResources/Information-Property-List/NSMotionUsageDescription

[2] Renishaw plc [n. d.]. *Optical encoders and LiDAR scanning*. Renishaw plc. Retrieved January 07, 2025 from https://www.renishaw.com/en/optical-encoders-and-lidar-scanning--39244

[3] MDN Web Docs [n. d.]. *Sensor APIs*. MDN Web Docs. Retrieved January 07, 2025 from https://developer.mozilla.org/en-US/docs/Web/API/Sensor_APIs

[4] Waymo LLC [n. d.]. *Waymo Driver*. Waymo LLC. Retrieved January 07, 2025 from https://waymo.com/waymo-driver/

[5] 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

[6] Apple Developer 2024. *Getting raw accelerometer events*. Apple Developer. Retrieved January 07, 2025 from https://developer.apple.com/documentation/coremotion/getting-raw-accelerometer-events

[7] Apple Developer 2024. *Getting raw gyroscope events*. Apple Developer. Retrieved January 07, 2025 from https://developer.apple.com/documentation/coremotion/getting-raw-gyroscope-events

[8] 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

[9] Wikipedia 2024. *Laser Doppler vibrometer*. Wikipedia. Retrieved January 07, 2025 from https://en.wikipedia.org/wiki/Laser_Doppler_vibrometer#/media/File:LDV_Schematic.png

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

[11] StatCounter 2024. *Mobile Operating System Market Share Worldwide*. StatCounter. Retrieved January 07, 2025 from https://gs.statcounter.com/os-market-share/mobile/worldwide

[12] Android Developers 2024. *Sensors Overview*. Android Developers. Retrieved January 07, 2025 from https://developer.android.com/develop/sensors-and-location/sensors/sensors_overview#sensors-rate-limiting

[13] 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

[14] 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

[15] Aydin Aysu, Nahid Farhady Ghalaty, Zane Franklin, Moein Pahlavan Yali, and Patrick Schaumont. 2013. Digital fingerprints for low-cost platforms using MEMS sensors. In *Proceedings of the Workshop on Embedded Systems Security* (Montreal, Quebec, Canada) *(WESS '13)*. Association for Computing Machinery, New York, NY, USA, Article 2, 6 pages. https://doi.org/10.1145/2527317.2527319

[16] 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

[17] Connor Bolton, Yan Long, Jun Han, Josiah Hester, and Kevin Fu. 2023. Characterizing and Mitigating Touchtone Eavesdropping in Smartphone Motion Sensors. In *Proceedings of the 26th International Symposium on Research in Attacks, Intrusions and Defenses* (Hong Kong, China) *(RAID '23)*. Association for Computing Machinery, New York, NY, USA, 164–178. https://doi.org/10.1145/3607199.3607203

[18] Eddy Bøgh Brixen. [n. d.]. *Facts about speech intelligibility*. DPA Microphones A/S. Retrieved January 07, 2025 from https://www.dpamicrophones.com/mic-university/background-knowledge/facts-about-speech-intelligibility/

[19] Liang Cai and Hao Chen. 2011. TouchLogger: inferring keystrokes on touch screen from smartphone motion. In *Proceedings of the 6th USENIX Conference on Hot Topics in Security* (San Francisco, CA) *(HotSec'11)*. USENIX Association, USA, 9.

[20] Jung-Woo Chang, Ke Sun, David Xia, Xinyu Zhang, and Farinaz Koushanfar. 2024. EveGuard: Defeating Vibration-based Side-Channel Eavesdropping with Audio Adversarial Perturbations. arXiv:2411.10034 [cs.CR] https://arxiv.org/abs/2411.10034

[21] Wei-Han Chen and Kannan Srinivasan. 2022. Acoustic Eavesdropping from Passive Vibrations via mmWave Signals. In *GLOBECOM 2022 - 2022 IEEE Global Communications Conference*. 4051–4056. https://doi.org/10.1109/GLOBECOM48099.2022.10001108

[22] Yunzhong Chen, Jiadi Yu, Linghe Kong, Hao Kong, Yanmin Zhu, and Yi-Chao Chen. 2023. RF-Mic: Live Voice Eavesdropping via Capturing Subtle Facial Speech Dynamics Leveraging RFID. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 2, Article 49 (June 2023), 25 pages. https://doi.org/10.1145/3596259

[23] Graham CLULEY. 2020. *Robot vacuum cleaners can eavesdrop on your conversations, researchers reveal*. Bitdefender. Retrieved January 07, 2025 from https://www.bitdefender.com/en-us/blog/hotforsecurity/robot-vacuum-cleaners-can-eavesdrop-conversations-researchers-reveal

[24] Editor's Draft. 2024. *Accelerometer*. W3C. Retrieved January 07, 2025 from https://w3c.github.io/accelerometer/#accelerometer-reading-quantization-algorithm

[25] DJI Enterprise. 2022. *LiDAR Drone Systems: Using LiDAR Equipped UAVs*. DJI. Retrieved January 07, 2025 from https://enterprise-insights.dji.com/blog/lidar-equipped-uavs

[26] Ming Gao, Yajie Liu, Yike Chen, Yimin Li, Zhongjie Ba, Xian Xu, Jinsong Han, and Kui Ren. 2022. Device-Independent Smartphone Eavesdropping Jointly Using Accelerometer and Gyroscope. *IEEE Transactions on Dependable and Secure Computing* PP (01 2022), 1–14. https://doi.org/10.1109/TDSC.2022.3193130

[27] Tom Goethem, Wout Scheepers, Davy Preuveneers, and Wouter Joosen. 2016. Accelerometer-Based Device Fingerprinting for Multi-factor Mobile Authentication. In *Proceedings of the 8th International Symposium on Engineering Secure Software and Systems - Volume 9639* (London, UK) *(ESSoS 2016)*. Springer-Verlag, Berlin, Heidelberg, 106–121. https://doi.org/10.1007/978-3-319-30806-7_7

[28] 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

[29] 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

[30] 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

[31] 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

[32] 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.

[33] Ben Nassi, Yaron Pirutin, Raz Swisa, Adi Shamir, Yuval Elovici, and Boris Zadov. 2022. Lamphone: Passive Sound Recovery from a Desk Lamp's Light Bulb Vibrations. In *USENIX Security Symposium*. https://api.semanticscholar.org/CorpusID:252402906

[34] 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

[35] 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

[36] Ke Sun, Chunyu Xia, Songlin Xu, and Xinyu Zhang. 2023. StealthyIMU: Stealing Permission-protected Private Information From Smartphone Voice Assistant Using Zero-Permission Sensors. https://doi.org/10.14722/ndss.2023.24077

[37] Payton Walker and Nitesh Saxena. 2021. SoK: assessing the threat potential of vibration-based attacks against live speech using mobile sensors. 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, 273–287. https://doi.org/10.1145/3448300.3467825

[38] 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

[39] Teng Wei, Shu Wang, Anfu Zhou, and Xinyu Zhang. 2015. Acoustic Eavesdropping through Wireless Vibrometry. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking* (Paris, France) *(MobiCom '15)*. Association for Computing Machinery, New York, NY, USA, 130–141. https://doi.org/10.1145/2789168.2790119

[40] Davey Winder. 2020. *Hacked Vacuum Cleaner Can Record Your Conversations—No Microphone Required.* Forbes. Retrieved January 07, 2025 from https://www.forbes.com/sites/daveywinder/2020/11/22/how-this-hacked-vacuum-cleaner-can-listen-to-your-conversations-using-lidar

[41] Zhi Xu, Kun Bai, and Sencun Zhu. 2012. TapLogger: inferring user inputs on smartphone touchscreens using on-board motion sensors. In *Proceedings of the Fifth ACM Conference on Security and Privacy in Wireless and Mobile Networks* (Tucson, Arizona, USA) *(WISEC '12)*. Association for Computing Machinery, New York, NY, USA, 113–124. https://doi.org/10.1145/2185448.2185465

[42] 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

[43] 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

[44] 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