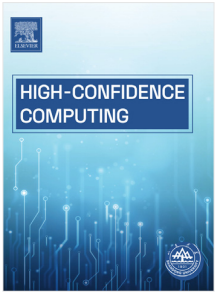
Journal Pre-proof



A survey of acoustic eavesdropping attacks: Principle, methods, and progress

Yiwei Chen, Wenhao Li, XiuZhen Cheng, Pengfei Hu

PII: S2667-2952(24)00044-8
DOI: https://doi.org/10.1016/j.hcc.2024.100241
Reference: HCC 100241

To appear in: *High-Confidence Computing*

Received date : 7 March 2024
Revised date : 5 April 2024
Accepted date : 8 April 2024

Please cite this article as: Y. Chen, W. Li, X. Cheng et al., A survey of acoustic eavesdropping attacks: Principle, methods, and progress, *High-Confidence Computing* (2024), doi: https://doi.org/10.1016/j.hcc.2024.100241.

A Survey of Acoustic Eavesdropping Attacks: Principle, Methods, and Progress

Yiwei Chen[a], Wenhao Li[a], XiuZhen Cheng[a] and Pengfei Hu[a,*]

[a]Institute of Intelligent Computing, the School of Computer Science and Technology, Shandong University, Qingdao, 266237, P.R. China

ARTICLE INFO

Keywords:
Acoustic Eavesdropping
Attack scenarios and threat models
Acoustic side-channel attacks

ABSTRACT

In today's information age, eavesdropping has been one of the most serious privacy threats in information security, such as exodus spyware[1] and pegasus spyware[2]. And the main one of them is acoustic eavesdropping. Acoustic eavesdropping[3] is a technology that uses microphones, sensors, or other devices to collect and process sound signals and convert them into readable information. Although much research has been done in this area, there is still a lack of comprehensive investigation into the timeliness of this technology, given the continuous advancement of technology and the rapid development of eavesdropping methods. In this article, we have given a selective overview of acoustic eavesdropping, focusing on the methods of acoustic eavesdropping. More specifically, we divide acoustic eavesdropping into three categories: motion sensor-based acoustic eavesdropping, optical sensor-based acoustic eavesdropping, and RF-based acoustic eavesdropping. Within these three representative frameworks, we review the results of acoustic eavesdropping according to the type of equipment they use and the physical principles of each. Secondly, we also introduce several important but challenging applications of these acoustic eavesdropping methods. In addition, we compared the systems that meet the requirements of acoustic eavesdropping in real-world scenarios from multiple perspectives, including whether they are non-intrusive, whether they can achieve unconstrained word eavesdropping, and whether they use machine learning, etc. The general template of our article is as follows: firstly, we systematically review and classify the existing eavesdropping technologies, elaborate on their working mechanisms, and give corresponding formulas. Then, these eavesdropping methods were compared and analyzed, and each method's effectiveness and technical difficulty were evaluated from multiple dimensions. In addition to an assessment of the current state of the field, we discuss the current shortcomings and challenges and give a fruitful direction for the future of acoustic eavesdropping research. We hope to continue to inspire researchers in this direction.

1. Introduction

In recent years, eavesdropping has had a relatively large security problem on smart homes, smartphones, and other devices. For example, the possibility of covert eavesdropping with a smartphone microphone[4], the KeyListener smartphone side channel attacks [5] and Lamphone[6]. Interestingly, most eavesdropping uses sound sensors to pick up the sound and then restore the sound and uses auditory or visual equipment to analyze and extract information, which can be specifically described as acoustic eavesdropping problems. Specifically, acoustic eavesdropping[3] is a surveillance technique that obtains information by capturing and analyzing sound or vibration[7]. This approach can involve listening to conversations[8], machine runs[7], keyboard clicks[9], etc., to quietly obtain sensitive information in the target environment without physical contact. Acoustic eavesdropping devices may include high-sensitivity microphones[10], sound amplifiers[11], vibration sensors[12], etc., which can be hidden in a variety of places[7] to collect audio evidence or intelligence undetected.

*Corresponding author
E-mail address: phu@sdu.edu.cn(P.Hu).
ywchen@mail.sdu.edu.cn (Y. Chen); li_wenhao@mail.sdu.edu.cn (W. Li); xzcheng@sdu.edu.cn (X. Cheng); phu@sdu.edu.cn (P. Hu)
ORCID(s):

Acoustic eavesdropping technology is essential for applications in the security sector[13]. Acoustic eavesdropping technology can be used to gather intelligence information to improve security and response capabilities. Acoustic eavesdropping technology also inspires and promotes scientific research and engineering applications. In the fields of acoustic signal processing, sensor networks[14], wireless communications[12], etc., the application of acoustic eavesdropping technology has promoted the innovation and development of related technologies and provided new ideas and methods for solving practical problems.

However, there is a lack of comprehensive investigation and analysis of the latest development and application of acoustic eavesdropping technology. Therefore, this article summarizes the current acoustic eavesdropping systems.

To a large extent, acoustic eavesdropping can be divided into the following categories: motion sensor-based acoustic eavesdropping, optical sensor-based acoustic eavesdropping, and RF-based acoustic eavesdropping. These vary depending on the device they are using. In particular, in the first type, motion sensor-based acoustic eavesdropping utilizes motion sensors on smart devices to capture the tiny movements of the device due to sound wave vibrations[15]. Motion sensors convert these vibrations into electrical signals that can be analyzed and reconstructed to identify sound information[16]. It does not require direct contact with the sound source and can be monitored remotely from a distance[17]. In the second, optical sensor-based acoustic

eavesdropping relies on optical devices to capture changes in light generated by vibrations on the surface of objects caused by sound waves[18]. This method allows for very precise measurement of the small movements of the surface of an object[19]. It can be used for acoustic eavesdropping at a greater distance[20]. In the third, RF-based acoustic eavesdropping uses the reflective properties of radio waves to detect and analyze the vibration of objects[21]. This method can penetrate certain substances and is beneficial for acoustic eavesdropping in non-line-of-sight situations[22]. For the use of different acoustic eavesdropping, it is necessary to cover multiple principles or theories such as the gyroscope principle, the acceleration principle, and the refraction and reflection theory of light, and for more detailed discussions, please refer to section 2.

Although there has been a lot of research in the field of acoustic eavesdropping, there is still a lack of comprehensive investigation into the timeliness of this technology due to the continuous advancement of technology and the rapid development of eavesdropping methods. With the rapid development of digital signal processing[23], artificial intelligence[24], wireless communication[25], and other fields, acoustic eavesdropping technology is also constantly evolving and updating.

In recent years, acoustic sensing technology has made remarkable progress in the fields of health monitoring[26], environmental monitoring[27], and smart home[28], which provides new possibilities for the application of acoustic eavesdropping technology[29]. [30] discusses the latest advances in acoustic emission technology and further demonstrates the potential and application prospects of acoustic technology in the field of eavesdropping.

In this article, we selectively elaborate on the methods of acoustic eavesdropping. More specifically, we classify acoustic eavesdropping into three categories based on the different devices used: motion sensor-based acoustic eavesdropping, optical sensor-based acoustic eavesdropping, and RF-based acoustic eavesdropping.

Motion sensor-based acoustic eavesdropping[31] exploits the sensitivity of motion sensors to sound-induced vibrations to capture sound signals. Acoustic eavesdropping of human voices is possible due to the partial overlap between the fundamental frequency of human speech and the sampling frequency of the sensor. Research on this attack began in 2014, and a variety of identification and reconstruction systems based on deep learning have been proposed. Acoustic eavesdropping based on optical sensors[32] is an attack method that uses lasers or other light sources to measure tiny vibrations on a target surface and restore sound signals. This attack exploits the high sensitivity of optical sensors to light signals and the nonlinear relationship between sound and light signals. Research on this attack began in the 1940s, and a variety of optical acoustic sensors based on MEMS[33], fiber optics[34], or micro-resonant cavities[35] have been proposed. RF-based acoustic eavesdropping[36] is a method that uses radio waves to measure tiny vibrations on the target surface and restore them. Sound signal attack method. This attack exploits the sensitivity of radio waves to changes in the electromagnetic field caused by sound, as well as the modulation and demodulation relationship between sound and radio waves. Research on this attack began in the 1960s[37]. At present, a variety of RF sound sensors based on ultrasonic[38], radar[39], or WiFi[19] have been proposed.

We discuss each of the three types of acoustic eavesdropping attacks. The representative eavesdropping methods are explained and the relevant formulas are given. We also compare three types of acoustic eavesdropping attack methods.

We compared them in terms of whether they were non-invasive, whether they could achieve unconstrained vocabulary, whether they could pass through opaque insulators, whether they needed ML help, whether they could detect mobile audio sources, whether they achieved high accuracy, and whether they could achieve low energy consumption. Each of these seven areas is briefly described below.

Research into non-intrusive techniques[40] allows for concealment, which in turn allows for covert acoustic eavesdropping. Non-intrusive techniques are usually more flexible and easily adapted to different acoustic eavesdropping needs[41]. Unconstrained vocabulary[42] means that the acoustic eavesdropping system needs to be able to understand and process any vocabulary without being constrained by presuppositions. This helps to understand more complex contexts and extract information broadly. Acoustic eavesdropping through an opaque insulator[10] allows sound to be captured without entering the target space, which increases the concealment of the action and the breadth of the sound signal collection. Unaided by ML[43] means that we don't need a lot of training data to train the model, which is good for acoustic eavesdropping in situations where data is difficult to obtain or where privacy is extremely demanding. In addition, the speed of sound signal processing can be increased without relying on complex machine learning models and large computing resources, making real-time monitoring and analysis possible. The ability of acoustic eavesdropping to listen to mobile audio sources[43] expands the range of applications for acoustic taps, allowing surveillance to be no longer limited to specific static locations, but to cover a wider area and different environmental conditions[44]. A high accuracy rate[3] can improve the overall performance of an acoustic eavesdropping system by reducing false alarms (incorrectly identifying non-target sounds as target sounds) and missed alarms (failing to detect target sounds that are present)[45]. It can also improve the quality of the collected acoustic signals. The realization of low energy consumptionc[46] can extend the life of the equipment, reduce costs, and facilitate the widespread use of acoustic eavesdropping methods.

For all articles that meet the requirements of acoustic eavesdropping in real-world scenarios, we have table 1 to compare their eavesdropping methods.

Finally, we discuss the shortcomings and challenges of current research and provide possible future directions.

Sensor Type	Acoustic Eavesdropping Attack	Competence						
		Non-Invasive	Unconstrained Vocabulary	Through Opaque Insulator	Unaided by ML	Mobile Audio Source	High accuracy	Low energy consumption
Motion Sensor	AccelWord[47]	✗	✗	✗	✗	✓	✓	✓
	PitchIn[48]	✗	✓	✗	✓	-	-	-
	AccelEve[49]	✗	✗	✗	✗	-	✓	✓
	AccEar[42]	✗	✓	✗	✗	-	✓	✓
	Speechless[50]	✗	✗	✗	✗	-	-	-
	Gyrophone[51]	✗	✗	✗	✗	-	✗	-
	HDD [52]	✗	✓	-	✓	-	-	-
	VibraPhone[49]	✗	✓	✗	✗	-	✓	-
	V-Speech[53]	✗	✓	✗	✓	-	✓	✓
Optical Sensor	Visual Microphone[54]	✓	✓	✗	✓	-	-	-
	LidarPhone[55]	✗	✓	✗	✗	-	✓	-
	Lamphone[6]	✓	✓	✗	✓	-	✓	-
Radio Receiver	WiHear[56]	✓	✗	✓	✗	✗	✓	-
	ART[10]	✓	✗	✓	✓	-	✓	-
	Tag-Bug[57]	✗	✓	✓	✗	-	✓	✓
	UWHear[58]	✓	-	✓	✓	✗	-	-
	WaveEar[59]	✓	✓	-	✗	✗	✓	✓
	MILLIEAR[60]	✓	✓	✓	✗	✗	✓	-
	mmSpy[61]	✓	✗	-	✗	✗	✓	-
	mmEcho[43]	✓	✓	✓	✓	✓	✓	-

Table 1
Comparison of acoustic eavesdropping attacks in the literature

The survey is presented as follows: First, we classify and summarize three types of acoustic eavesdropping and illustrate them with examples in section 2. Then we discuss the advantages and disadvantages of different acoustic eavesdropping attacks in section 3. Finally, section 4 and section 5 give future research directions. and summary.

2. Acoustic Eavesdropping Method

Acoustic eavesdropping is secretly or quietly listening to other people's private conversations without their permission to get information. Acoustic eavesdropping is a basic security and privacy threat in wireless networks, which can lead to the leakage of sensitive information, identity impersonation, data tampering, or other malicious behavior. Acoustic eavesdropping can be used in the military to obtain the strategies and plans of enemies or competitors. It can be used in business and finance to steal trade secrets, customer data, or market dynamics. It can be used for scientific and technical analysis and detection of signal features. Ans it can be used by individuals to obtain other people's private information, etc.

To improve privacy protection capabilities, we summarized existing acoustic eavesdropping methods. This section reviews existing acoustic eavesdropping methods and divides them into the following three categories based on different acoustic eavesdropping methods.

2.1. Motion Sensor-based acoustic eavesdropping

Motion Sensor-based acoustic eavesdropping [62] is a technology that uses motion sensors (such as accelerometers, gyroscopes, etc.) on smartphones or other devices to capture and analyze surrounding sound signals. This technology can bypass the permission restrictions on the microphone [63], thereby achieving privacy eavesdropping on the user. The basic principle of motion sensing-based acoustic eavesdropping is that when sound waves propagate in the air, they cause tiny vibrations in objects, which can be detected by motion sensors and converted into electrical signals. By performing signal processing and machine learning on these electrical signals, the spectral characteristics of the sound waves can be restored, thereby identifying the speaker's information and even parsing the speech content.

In the existing research, acoustic eavesdropping based on motion sensors can be divided into the following aspects.

2.1.1. Base on MEMS Gyroscopes

A standard-size(non-MEMS) gyroscope [64] generally consists of a rotor, an axis of rotation, and a gimbal coordinate system (inner and outer rings). When the rotor rotates at high speed, if no external moment acts on the gyroscope, the spindle of the rotor remains axially fixed, i.e. pointing in a fixed direction. If an external moment acts on the gyroscope, the spindle of the rotor rotates, i.e., around an axis perpendicular to the external moment. By measuring the direction of the rotor's axis of rotation and the angular velocity of the precession, the angular velocity and direction of the carrier can be obtained. However, acoustic eavesdropping uses a MEMS gyroscope [65]. All MEMS gyroscopes work with another physical phenomenon – the Coriolis force [66], which is a fictitious force (D'Alembert force) observed in a rotating frame of reference, which is oriented perpendicular to the axis of rotation of the frame of reference and the velocity of the object.

The formula for calculating the Coriolis force is

$$F_{cor} = 2m\vec{v} \times \omega \tag{1}$$

where m and v represent the mass and velocity of the object, respectively, and ω denotes the angular rate of the reference frame.

In general, the structure of a MEMS gyroscope consists of a vibrating detection mass and a fixed driving mass, and the MEMS gyroscope measures its angular rate (ω) by sensing the amount of Coriolis force acting on the moving detection mass within the gyroscope. When the gyroscope rotates around an axis perpendicular to the direction of vibration, the detected mass is subjected to the Coriolis force, which produces a secondary vibration perpendicular to the direction of vibration, and the Coriolis force is sensed by measuring the vibration it produces. Its vibrational frequency is also known as the gyroscope's resonant frequency[67].

Specifically, a MEMS gyroscope is used as a microphone for speech recognition and analysis in [51]. First, the authors extracted the features measured by the MEMS gyroscope, such as digital pronunciation and other information, and preprocessed the data. Then, they trained machine learning algorithms to identify specific speakers or number pronunciations. In this way, they could recognize and leak sensitive information such as numbers spoken near or above the phone, and achieve speaker (including speaker gender) identification and isolated word recognition, showing the potential eavesdropping risk of MEMS gyroscopes.

2.1.2. Based on Accelerometer

An accelerometer[68] is an inertial sensor that is capable of measuring the acceleration force of an object. The acceleration force is the force that acts on an object during its acceleration, such as the earth's gravitational pull, which is also known as gravity. The acceleration force can be a constant, such as g, or a variable. Accelerometers generally consist of a body, a spring, and an inertial body. Accelerometers use inertial forces and capacitance changes to detect acceleration. When the body of the accelerometer is accelerated, the inertial body will move relative to the body due to inertia, resulting in a change in the shape and length of the spring, and the capacitance between the body and the inertial body will change with distance. In this way, by measuring the change in capacitance, the magnitude and direction of the acceleration can be calculated.

The accelerometer sensors currently used in smartphones and other smart devices, such as smartwatches and smart glasses, are Micro-Electro-Mechanical Systems (MEMS)[69]. MEMS technology enables MEMS accelerometers to be small, light, and energy-efficient. This type of MEMS accelerometer is a device that captures the acceleration of its body along three sensing axes, each of which is typically handled by a sensing unit with three main components: inertial mass, spring legs, and stationary fingers[70]. When acceleration is applied, the inertial mass shifts in the opposite direction, resulting in a change in capacitance between the stationary fingers. This change produces an analog signal, which is then mapped to an acceleration measurement.

Specifically, in [71] the authors discussed a method of performing a side-channel attack on smartphone speakers using the smartphone's accelerometer. They propose AccelEve, a learning-based smartphone eavesdropping attack that can identify and reconstruct speech signals emitted by smartphone speakers, and demonstrate its ability to identify sensitive words in calls, as well as how to link this information to specific callers by cross-identifying sensitive words across multiple phone calls. [47]discusses how to identify some of the user's private information through the smartphone's accelerometer, and proposes a solution named AccelWord to address the energy consumption problem of voice control. The authors point out that if the attacker can adjust the sampling frequency of the accelerometer, they can even use the accelerometer readings to reconstruct part of the human speech, increasing the risk to user privacy.

The above are all eavesdropping studies that only target human speech. [50] focuses on the impact of machine-generated speech and actual human speech on smartphone motion sensors, as well as the potential privacy leakage risks. [42] performs eavesdropping attacks through accelerometer signals, which can reconstruct any audio signals played by smartphone speakers, and the vocabulary is unrestricted.

2.1.3. Other methods

Besides the above methods of eavesdropping using gyroscopes and accelerometers, there are also the following types of motion-sensing acoustic eavesdropping: [48] eavesdrops on sounds in a room by reconstructing intelligible speech signals from data fused from non-acoustic sensors (e.g. seismometers, gyroscopes, accelerometers). [52] explores the possibility of extracting and parsing human speech using the mechanical components in disk drives as miniature microphones, and demonstrates a method of eavesdropping using the mechanical objects in hard disk drives. [49] discusses the possibility of extracting and parsing human speech using the sound sensor with minimal hardware changes by transforming the vibration motor into a sound sensor using the back electromotive force generated by the ambient sound in the smartphone's vibration motor, and designs techniques to decode human speech from noisy, low-bandwidth signals, indicating that the vibration motor can be used to listen to human speech under certain conditions. [53] eavesdrops on audio using the vibration sensor in the glasses nose pad. [72] achieves eavesdropping by converting the speakers connected to the computer into microphones.

2.2. Optical sensor-based acoustic eavesdropping

Optical sensor-based acoustic eavesdropping is a secret technique that uses optical sensors[73] such as photodiodes, photomultiplier tubes, phototransistors, etc. to capture the vibration signals of the target object and restore the sound content from them. The principle of this technique is that sound is a vibration[74], which causes tiny deformations on the surface of the object, thereby changing the reflection or transmission of light from the object. By measuring these changes in light intensity, the vibration signal of the object can be obtained, and then the sound content can be restored

by signal processing and machine learning[75]. There are several types of acoustic eavesdropping based on optical sensors.

2.2.1. Based on Cameras

The principle of acoustic eavesdropping using a camera is to use the optical sensor of the camera to capture the vibration signal of the target object and restore the sound content from it. This is a secret technique. [54] used a high-speed camera for acoustic eavesdropping. A high-speed camera[76] is a device that can capture motion images at an exposure of less than 1/1000 second or a frame rate of more than 250 frames per second. Its working principle is to use the optical sensor to convert the light signal into an electric signal, and then process and store the signal, and finally playback or display it in slow motion. The article introduces a novel visual microphone technology, which uses a high-speed camera to capture the vibration mode of the object under the action of sound waves, and restores the sound signal by processing the recorded video with an algorithm.

Specifically, the technology is based on the movement of objects caused by sound waves, using high-speed cameras to record the movement patterns of objects, and then processing the video through algorithms to restore the sound signal.

Firstly, the input video is decomposed into spatial sub-bands with different directions and scales, and the local motion signals of each pixel, direction, and scale are calculated to obtain the local motion information of objects in different directions and scales. The specific calculation is as follows: Phase variations are used in the complex steerable pyramid[77] representation of video V to calculate local motion. Decompose each frame of video $V(x, y, t)$ into complex sub-bands corresponding to different scales and orientations. Every scale r and direction θ can be described as a complex image, which can be expressed in terms of amplitude A and phase ϕ as

$$A(r, \theta, x, y, t)e^{i\varphi(r,\theta,x,y,t)}. \tag{2}$$

By subtracting the local phase ϕ derived from this equation from the local phase of the reference frame t_0 (the first frame of the video in most cases), the phase change is calculated.

$$\varphi_v(r, \theta, x, y, t) = \varphi(r, \theta, x, y, t) - \varphi(r, \theta, x, y, t_0). \tag{3}$$

For small motions, the alterations in phase are roughly in line with the displacement of the image structure across the associated direction and scale.

Then, by combining local motion signals, aligning individual motion signals $\varphi(r, \theta 1, t)$, and performing weighted averaging, where r is the radial coordinate,θ is the angular coordinate, and t is the time. The aligned signals are given by $\phi(r_i, \theta_i, t - t_i)$, such that:

$$t_i = \arg\max_{t_i} \phi_0(r_0, \theta_0, t)^T \phi_i(r_i, \theta_i, t - t_i) \tag{4}$$

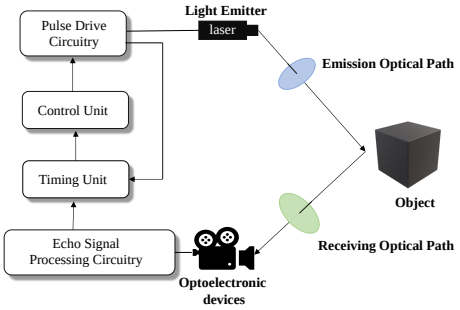


Figure 1: General process of acoustic eavesdropping based on optical sensors. The laser emitter emits a laser beam that irradiates the target object along the emitting optical path. The target object reflects the laser light and the laser returns to the system along the receiving optical path. The optoelectronic device on the receiving optical path receives the reflected laser signal. The echo signal processing circuit analyzes the received signal and calculates the distance between the target object and the system.

The weight of each local signal is measured by its (squared) amplitude:

$$\phi_i(r, \theta, t) = \sum_{x,y} A(r, \theta, x, y)^2 \varphi_v(r, \theta, x, y, t). \tag{5}$$

Then the global motion signal is

$$\hat{s}(t) = \sum_i \phi_i(r_i, \theta_i, t - t_i) \tag{6}$$

In this way, the overall movement pattern of the object is obtained, and the sound signal is finally restored.

2.2.2. Based on Laser Sensors

Lidar Sensor[78] is a technology that uses a laser beam to scan the surrounding environment and generate a distance map. It can be used in a variety of applications such as remote sensing, autonomous vehicles, meteorology, astronomy, etc.

Lidar Ranging

Lidar sensors are typically used to measure distance and detect targets. The ranging principle is shown in the figure 1. Specifically, there are several ways to measure the range by lidar[79].

Time-of-Flight(ToF). Time-of-Flight[80] is a method of measuring distance. Lidar emits a short pulse of laser light and records the elapsed time from the time the beam is transmitted to the time it is received. Since the speed of light is known, the distance between the target object and the lidar can be calculated by measuring the time of flight of the light. It is commonly used in applications that require fast and accurate distance measurement, such as autonomous driving[81] and intelligent transportation[82].

The formula for ToF ranging[83] is

$$D = \frac{c * \Delta t}{2} \tag{7}$$

where D is the distance of the measured object, c is the speed of light, and Δt is the time of flight of the laser.

Pulse Lidar. This method[84] is similar to the TOF, but the difference is that the lidar emits a series of short pulses of laser light and records the launch and return time of each pulse. By comparing the transmitting time and return time of each pulse, it is possible to calculate the distance to the target object. It is often used in long-distance and high-precision ranging scenarios, such as geological exploration[85], environmental monitoring[86], spatial mapping[87], etc.

Assuming that the number of count pulses when the echo pulse arrives is n and the repetition period of the count pulse is T, then the delay time of the echo pulse relative to the transmitted pulse is

$$\Delta t = n * T \tag{8}$$

Then the distance is

$$D = \frac{c * \Delta t}{2} = \frac{c * nT}{2} \tag{9}$$

Frequency-Modulated Continuous-Wave(FMCW). It uses a continuous laser beam and applies modulation at its frequency[88]. By measuring the frequency difference between the emitted frequency and the returned frequency, the distance between the target object and the lidar can be calculated. It is widely used in the automotive field, industrial-ranging applications, etc.

The following is a discussion of the case where the relative velocity v_r is 0. Assuming that the distance from the detected target to the lidar is R, the delay of the received signal of the lidar receiver is τ, and the speed of light is c, the relationship between the detection range R and the delay τ and the speed of light c is as follows when the target velocity is ignored

$$R = \frac{\tau * c}{2} \tag{10}$$

t_c is half of the swept period. f_c is the swept bandwidth. And *tau* is the time from transmission to acceptance. Let $f_s(t)$ and $f_e(t)$ be the frequency change functions of the sent and received signals, respectively, and assume that the relative velocity v_r is 0, then the rising edge of the signal has the following relationship:

$$\begin{aligned} f_S(t) &= f_0 + \frac{f_c}{t_c} * t \\ f_E &= f_S(t - \tau) \end{aligned} \tag{11}$$

And there is a beat frequency function:

$$f_b(t) = f_S(t) - f_E(t) \tag{12}$$

And because $R = \frac{\tau * c}{2}$, from the geometric relation $\frac{f_b}{\tau} = \frac{f_c}{t_C}$, we can deduce:

$$R = \frac{c}{2} * \frac{t_c}{f_c} * f_b \tag{13}$$

From the above formula 13, R is proportional to f_b. If the emitted signal is a cosine wave, the variation of its time domain is as follows:

$$u_S(t) = \hat{u}_S * \cos[2\pi * f_S(t) * t + \varphi_S] \tag{14}$$

The change of the received signal in the time domain is:

$$u_E(t) = \hat{u}_E * \cos[2\pi * f_E(t) * t + \varphi_E] \tag{15}$$

Bring in f_e to get:

$$u_E(t) = \hat{u}_E * \cos[2\pi * f_S(t) * t - 2\pi * f_b(t) * t + \varphi_E] \tag{16}$$

Phase-Based Lidar. This method[89] calculates the distance by measuring the phase difference between the emitted laser beam and the returned beam. The change in phase difference is associated with the distance of the target object. It is suitable for medium and close-range measurements, such as industrial measurements and medical applications.

Specifically, it is to modulate the intensity of the light wave emitted. The measured distance can be expressed as

$$D = \frac{c * \Delta\phi}{2f} \tag{17}$$

where D is the distance of the measured object, c is the speed of light, $\Delta\phi$ is the phase difference of the laser, and f is the modulation frequency of the laser.

Triangulation. It uses the geometric relationship between the angle of incidence and the angle of reflection of the laser to calculate the distance. It is suitable for low-cost sensors, such as those found in robotic cleaners.

Specifically, the laser emitted by the laser is incident at a certain angle with the normal of the object surface to the surface of the measured object, and the back(scattered) light is converged and imaged through the lens at one point B, and finally collected by the photosensitive unit. We suppose the angle between the incident light AO and the baseline AB is α. AB is the distance between the center of the laser and the center of the CCD. BF is the focal length f of the lens. D is the limit position of the reflected light imaging on the photosensitive unit when the measured object is far from the infinity of the baseline. DE is the displacement of the spot from the limit position on the photosensitive unit, denoted as x.When the optical path of the system is determined, the α, AB, and f are all known parameters. From this, we know that

$$AO = \frac{AB * F}{x * \alpha} \tag{18}$$

When the relative displacement of the measured object and the baseline AB occurs, x changes to x', and the distance of the measured object y can be obtained from the above conditions that

$$y = \frac{AB * f * (x - x')}{xx'} \tag{19}$$

Lidar-based Eavesdropping

Above, we have explained the method of using lidar for ranging. Because lasers can be used for fine-grained distance measurements, and the propagation of sound through a medium in the form of mechanical waves causes small physical vibrations of nearby objects, we can take advantage of this feature for remote audio eavesdropping. A laser microphone was designed in [90]. A laser microphone shines a laser beam on an object close to the place of origin and measures the vibrations caused by the sound to restore the audio. Laser microphones must be manually applied to their transmitters and receivers to obtain the required information through specular reflection. In [55], LidarPhone uses the diffuse reflection of the robot vacuum cleaner for eavesdropping. It uses triangulation to capture tiny vibrations. The signal-to-noise ratio of the final input signal is improved by preprocessing, followed by training and prediction. Supervised learning techniques were used to extract relevant features for classification. Deep learning techniques, especially convolutional neural networks, are used to extract privacy-sensitive information. Acoustic eavesdropping was finally achieved.

2.2.3. Based on Remote Electro-optical Sensor

In addition to the above-mentioned methods using high-speed cameras and lidar, some methods use remote photoelectric sensors[6] to analyze the vibrations caused by the sound of the bulb to reconstruct the audio. Lamphone uses the hanging light bulb as an information leakage channel and analyzes the light signal reflected by the vibration of the light bulb to steal the sound in the room. The principle is shown in the figure 2. It can improve the sensitivity of the system by increasing the internal gain of the sensor to optimize the signal-to-noise ratio of optical measurements.

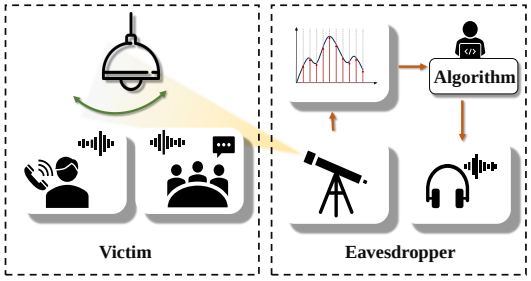


Figure 2: Lamphone's threat model. The sound creates fluctuations on the surface of the hanging bulb. Lamphone uses a remote electro-optical sensor to analyze the frequency response of the hanging bulb to the sound, and the algorithm processes it to obtain the recovered acoustic signal.

2.3. RF-based acoustic eavesdropping

A radio frequency signal[91] is a high-frequency alternating electromagnetic wave that can travel through the air and be reflected by the ionosphere at the outer edge of the atmosphere to form a long-distance transmission capability. The frequency range of RF signals is generally between 300kHz and 300GHz. It is widely used in radar and wireless communication. RF-based eavesdropping attacks are an attack method that uses radio frequency signals to listen to other people's communications or obtain sensitive information. The principle is to use a special wireless receiving device to intercept the radio frequency signal sent by the target device, and then convert the signal into readable data through a decoder, from which to extract sensitive information such as telephone call content, email, and text messages. The Great Seal Bug [92] was one of the first acoustic eavesdropping devices to use passive RF technology to transmit audio signals. The eavesdropping device is a silver-plated copper cylinder consisting of a central tuning column and a coupling disc inside the cylinder. The Great Seal Bug is eavesdropping on amplitude modulation signals triggered by tuning, coupling modes, and vibrations of the diaphragm.

There are different types and ways of eavesdropping based on RF technology, and we have summarized them.

2.3.1. Based on WiFi

[10] and [56] can identify specific words by analyzing WiFi Received Signal Strength (RSS) and Channel State Information (CSI), respectively.

[10] proposed a new acoustic eavesdropping method, ART. It uses a reflected or emitted wireless signal to penetrate a conventional soundproof device, stealing subtle vibrations from the target device and converting it into an audio signal. The authors used wireless vibrometry for remote sound capture or recovery. Because audio emissions cause tiny vibrations in the speaker itself, they can resonate with radio waves reflected from the speaker or a wireless transmitter located in the same location as the speaker. So contaminated radio waves can be captured and processed by a tampered receiver to restore the original audio played by the speaker. After the audio is collected, the required audio is obtained through the audio RF conversion and demodulation of the converted audio.

Specifically, the effect of vibration on radio RSS in audio RF conversion can be expressed by the following formula

$$RSS_L = \sigma[A^2(d_0)+2A(d_0)A'(d_0\hat{d}+\cdots+o(d_0)\hat{d}^k] \quad (20)$$

where $\hat{d} = d * cos\beta$ and β is the angle between the direction of vibration and the direction of reflection. σ is the reflectivity (reflection gain) of the speaker surface ($0 < \sigma < 1$). d_0 indicates the distance between the antenna and the speaker. $A(\cdot)$ is the channel gain function. The squared operation on $A(\cdot)$ models signal attenuation due to round-trip propagation.

The effect of vibration on the radio phase can be expressed by the following formula:

$$Phase_L = \frac{2\pi(d_0+2d)}{\lambda_0} + \gamma \quad (21)$$

where the $\frac{4\pi d}{\lambda_0}$ term contains the audio frequencies from the speaker. γ is the initial phase of the reflection path. This formula relates the phase of a radio signal to a vibration-induced displacement, which changes the length of the path the signal travels, thus changing its phase.

Audio signals are obtained by reverse demodulation by capturing audio-modulated radio samples, splitting samples, signal processing, controlling ranges using bandpass filters, and so on

[56] introduces a system called WiHear. The frame is shown in figure 3. WiHear doesn't need to deploy any devices and uses Wi-Fi signals to eavesdrop on people's conversations. It does this by detecting and analyzing tiny radiant reflexes from mouth movements. WiHear utilizes MIMO beamforming technology to locate and focus the mouth while hearing multiple conversations. It utilizes CSI (Channel State Information) for partial multipath removal of commercial OFDM-based Wi-Fi devices. Partial multipath removal and discrete wavelet packet transform were used to construct the mouth motion profile. Leverage machine learning to recognize pronunciations and translate them through classification and context-based error correction. WiHear enables the monitoring and recognition of human speech behavior.

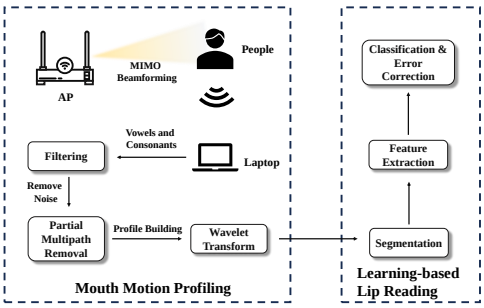


Figure 3: Frame diagram of WiHear. It consists of a transmitter and a receiver for single-user lip reading. The transmitter uses beamforming to send a Wi-Fi signal to the user's mouth. The receiver extracts and analyzes the reflexes of mouth movements. Then it explains the mouth movement in two steps. On the left is the mouth movement analysis using filtering, partial multipath removal, and wavelet transform to purify the signal and decompose the moh movement profile. On the right is learning-based lip reading, which applies machine learning to recognize pronunciations and translate them through classification and context-based error correction.

2.3.2. Based on Ultra-Wide Band

Ultra-wide Band (UWB)[93] is a wireless communication technology that enables short-range, high-bandwidth communications using very low energy levels over a large radio spectrum. UWB transmits information by emitting short pulses over a large bandwidth (>500 MHz), enabling pulse position or time modulation. Information can also be modulated on UWB signals (pulses) by encoding the polarity and amplitude of the pulses or by using quadrature pulses. UWB pulses can be sent intermittently at relatively low pulse rates to support time or position modulation, or they can be sent at rates up to the reciprocal of the UWB pulse bandwidth.

UWB wireless systems[94] can be used to measure the "time of flight" of transmissions at different frequencies, which helps to overcome multipath propagation, as some frequencies have direct paths, while others have longer delays. Using cooperatively symmetrical bidirectional metrology technology, distances can be measured with high resolution and high accuracy. In addition, UWB devices can also determine whether an object is stationary, near, or far away

Acoustic eavesdropping based on Ultra-Wide Band[95] utilizes ultra-wideband technology to transmit audio signals. It collects and converts the sound signal into a digital signal, and then uses UWB to transmit it to the receiver and decode it to restore the original sound signal. It utilizes the high-frequency band utilization and high-speed data transmission capabilities provided by UWB technology, enabling the eavesdropping device to transmit sound signals in a wide frequency band, and having a certain degree of concealment and anti-interference ability.

[58] proposed the UWHear system. It uses Impulse Radio Ultra-Wideband (IR-UWB) technology to build an enhanced audio perception system.

The following describes the data structure of UWHear. The X-axis is the fast time, which represents the round-trip ToF of the pulse so that the fast time can be converted to the distance bin. The Y-axis is slow time. The data collected during the response is called a frame. All frames are sorted chronologically, and placed along the Y-axis. After finding the distance bin corresponding to each sound source, a slice is taken from the 2D matrix, and the 1D time series is obtained to estimate the sound source. The IR-UWB radar uses ultra-wideband pulse signals for distance measurement and recovers the acoustic information of the target by receiving and analyzing the echo signal. The target audio is recovered by sending and receiving ultra-wideband pulse signals, processing pulse-echo signals, time domain analysis, inversion algorithms, etc.

The received signal in this article is $y(t)$, which is modeled as a convolution of the transmitted signal and the channel impulse response plus additive noise, i.e. the signal $y(t)$ is transmitted through a system (expressed as the system function $h(t)$) plus the noise $n(t)$

$$\begin{aligned} y(t) &= x(t) * h(t) + n(t) \\ &= \sum_{p=1)}^{P} \alpha_p g(t - kT_s - T_p - T_p^D(t)) \\ &\quad \times \cos(2\pi f_c(t - kT_s - T_p - T_p^D(t))) + n(t) \end{aligned} \tag{22}$$

where $x(t)$ is the input signal. P denotes the range of the sum. ap is the coefficient. $g(cdot)$ is a function. f_c

is the frequency of the signal. T_s, T_p, and T_p^D are time delay parameters in communication systems. The formula describes the transformation of a signal during transmission and takes into account both internal and external influences in the system.

UWHear is capable of sensing audio vibrations on the wall and operates in some non-line-of-sight (NLOS) conditions. The ability to simultaneously recover and separate sounds from multiple sources is achieved. Experiments have shown that UWHear is effective in separating the contents of two speakers that are only 25 cm apart. It achieves a good balance between signal penetration and ranging resolution. However, because its sampling rate is 1.6 kHz, which is less than the minimum sampling rate of 3 kHz (the sampling rate of landlines) for a sufficient understanding of human speech, it is impossible to understand the larger corpus by intelligently understanding the numbers read by humans.

X-band UWB[96] is an ultra-wideband wireless communication technology that uses the frequency range of 8 12 GHz. The advantages of X-band UWB are higher resolution, lower path loss, better penetration, and smaller antenna size. [97] demonstrated the feasibility of remotely sensing sound and recovering sound signals from vibration sources by using UWB radar technology. X-band UWB radar is used to observe multiple separated sound sources in different ranges, and their signals are separated and recovered, and the perspective ability of microwave signals is used for target monitoring blocked by obstacles. Three unique experimental setups are used to demonstrate the feasibility of this technique: a passive object in the proximity of the active source, the separation of multiple sound sources at different ranges of the radar, and the blocking of the sound source by a dielectric medium. However, it has a low audio response and can only recover audio below 400 Hz, so the recovery of human voice is not taken into account.

2.3.3. Based on RFID

RFID[98] is a wireless communication technology that uses electromagnetic waves to identify and track tags attached to objects, people, or animals. These tags are called RFID tags[99]. It stores digitized data related to the marked object and can be read by RFID readers. [57] demonstrates the possibility of using low-cost and easily overlooked RFID tags to effectively perform through-the-wall eavesdropping. A battery-free method called Tag-Bug is proposed. Tag-Bug extracts sound features in two ways: (i) vibration effect, in which the sound directly causes the label to vibrate; (ii) Reflection effect, in which the label does not vibrate, but perceives the reflected signal of nearby vibrating objects. In this article, the attack focused on the sound played by the speakers, not the sound of a real person speaking because real people are speaking mainly cause air movement, not air vibrations caused by sound.

The signal received by the RFID reader can be divided into three parts: leakage signal, multipath signal, and backscattered signal. Among them, leakage signal[100] refers to the leakage of a part of the signal from the transmitter to the receiver or from the receiver to the transmitter due to the incomplete isolation of the channel between the transmitter and the receiver in wireless communication, resulting in signal interference and performance degradation. These leaks to the receiver or transmitter become leaked signals. Multipath signal[101] refers to the phenomenon of reflection, diffraction, scattering, and other phenomena of the signal due to the signal encountering various objects in the process of propagation, resulting in the formation of several different paths to reach the receiver, and the signals of these different paths are called multipath signals. Backscattered signal[102] refers to a technology that uses the energy of an incident radio frequency signal to transmit information in wireless communication, also known as backscattered communication. Therefore, the received signal is expressed as

$$\begin{cases} S_L = S_{TX} h_L, \\ S_E = S_{TX} h_{E,d}, \\ S_0 = S_{TX} h_d h_{d'}, S_1 = S_{TX} h_d h_{d'} h_1. \end{cases} \tag{23}$$

where h_1 is the modulation gain of the tag. S_{TX} is the CW signal sent by the TX antenna. S_L is the leak signal. S_E is the multi-path signal, S_0 or S_0 is the backscattered signal. h_d is the signal attenuation caused by the uplink transmission distance. $h_{d'}$ is the signal attenuation caused by the downlink transmission distance. And $h_{E,d}$ is the overall signal attenuation due to the environment, which is also related to the distance d. To amplify the effects of vibrational signals, the authors devised a new signal signature called Modulated Signal Differential (MSD) to reconstruct sound from RF signals. To improve the quality of reconstructed voices for human speech recognition, the authors applied Conditional Generative Adversarial Networks (CGANs) to recover the full frequency band from part of the reconstructed voice. The side-channel attacks described in the article can be launched in three different ways: medium-based eavesdropping, aerial-based eavesdropping, and reflection-based eavesdropping. Experiments have shown that Tag-Bug can successfully capture monotonous sounds when the loudness is greater than 60 dB. Tag-Bug can effectively identify the number of human voices in free-space eavesdropping, thru-the-brick-wall eavesdropping, and thru-the-insulating-glass eavesdropping, as well as accurately identify letters in free-space eavesdropping.

2.3.4. Based on mmWave

Millimeter wave[103] is an electromagnetic wave with a wavelength between 1 mm and 10 mm and a frequency between 30 GHz and 300 GHz. Millimeter waves have a wide range of applications in communications, radar, remote sensing, and astronomy.

The general principle of millimeter wave for acoustic eavesdropping[104] is to use millimeter wave radar to measure the vibration of sounding objects, and then restore the sound signal through signal processing or deep learning.

Specifically, when sound waves propagate to sound-emitting objects (such as loudspeakers, glass windows, etc.), they cause them to produce small displacements that affect the echo phase of millimeter-wave radar. Millimeter-wave radar can capture these phase changes and convert them into electrical signals, which are then used to recognize human speech. A major advantage of this approach is that it can work in a variety of complex environments, including in the presence of noise or sound mitigation measures. This is because millimeter waves can accurately capture the air fluctuations generated by sound, and are highly robust to factors such as noise.

[60] proposed a system-level acoustic eavesdropping system - MILLIEAR, which integrates millimeter-wave FMCW and generative machine learning networks. It uses millimeter-wave radar to capture tiny vibrations caused by sound, extracting large amplitude vibrations based on coarse-grained phase estimation of millimeter-wave signals between chirps. However, all chirp-room-based methods have limitations in terms of low-frequency response and inaccuracy, so the authors used generative machine learning models to enhance the millimeter-wave radar signal reflected from the speaker to reconstruct the original audio. It does not require physical contact with the victim's device/sensor, nor does it require the installation of spyware on the victim's device. It uses frequency-modulated continuous wave (FMCW) to transmit signals (refer to Equations 14,15) to a vibrating speaker. Enhance captured speaker vibrations by generating machine-learning models. The model does not require prior knowledge of the words in the audio signal, and the eavesdropping vocabulary is not restricted by developing a Conditional Generative Adversarial Network (cGAN). It uses virtual sub-chirps to measure phase changes and uses millimeter-wave radar signals to extract speaker vibrations in the presence of multipath noise. The results and evaluations show that the attack is very effective under a range of real-world limitations, such as different angles and partitions.

[61] use millimeter-wave radars to eavesdrop on mobile phone voice content. The authors came up with an attack model called mmSpy. The model can be used to classify and reconstruct speech (such as words and numbers) within mobile phones through domain adaptation techniques based on synthetic and real radar data. mmSpy has created a model based on synthetic training data generated using popular speech datasets at scale. Synthetic training data is combined with small-scale training data from real radar to generate mmSpy's audio reconstruction and speech classification models. This model is used to classify and reconstruct speech content. mmSpy proposes a range of techniques, including statistical noise correction, machine learning-based modeling, and domain adaptation. To account for the discrepancies between synthetic and real radar data, mmSpy uses small-scale training data from real radar to domain adapt the model. mmSpy senses the tiny vibrations produced by the handset device that the user is listening to during a call by detecting phase changes in the millimeter-wave signal reflected from the phone's body. mmSpy can eavesdrop on audio content even if the audio is completely inaudible to humans and nearby microphones. mmSpy demonstrated the feasibility of eavesdropping on calls using headphones, and demonstrated its ability to detect tiny vibrations of headphones that can't be heard by microphones co-located with radar.

Both articles deal with techniques for eavesdropping using millimeter-wave radar. [60] adopted a combination of frequency-modulated continuous-wave (FMCW) ranging and conditional generative adversarial networks, which can accurately reconstruct audio even in different distances, angles, and through-wall scenarios. [61] achieved eavesdropping on phone calls by detecting the weak vibrations of headphones using millimeter-wave radar signals, and demonstrated the feasibility of reconstructing audio signals from radar data.

[43] introduces the voiceprint eavesdropping system mmEcho, which uses millimeter-wave radio signals to accurately measure the micron-level vibration of objects caused by sound waves. Compared to previous studies, mmEcho's eavesdropping method is highly accurate and does not require machine learning and prior knowledge. The mmEcho system consists of three modules: Reverberant Object Detection (ROD), Vibration Extraction and Audio Reconstruction (VA), and Audio Noise Reduction (ANR). Since the traditional FMCW distance calculation formula13 does not provide micron-level resolution for vibration measurements, the authors used the intra-chirp method to estimate the distance between the target object and the radar. The results show that mmEcho can accurately reconstruct audio from moving sources at various distances, orientations, reflective objects, soundproofing materials, different languages, and sound levels based on the textual information provided.

2.3.5. Other methods

[105] uses Doppler radar to identify human voices by capturing micro-Doppler features from laryngeal and oral vibrations. The authors succeeded in using Doppler radar to capture the echo signal produced when a person emitted seven notes from Do to Ti and letters from A to Z. Through spectrum analysis, the authors successfully classified these 26 letters using a deep convolutional neural network with an accuracy of 94%. To overcome the problem of insufficient data volume and improve classification accuracy, the authors introduced transfer learning and increased the accuracy to 97% using the VGG-16 model. However, the frequency response of the method proposed in this paper is limited to less than 200 Hz.

[106] mainly discusses a new side-channel attack method that obtains the content of a digital screen through liquid crystal nematic state sensing. The authors designed a portable, low-cost, and energy-efficient 24GHz millimeter-wave probe, and proposed a layered module based on end-to-end deep learning for identifying screen content types and retrieving sensitive information on digital screens. The authors have conducted a large number of experiments to show that the proposed WaveSpy is capable of achieving

more than 99% inference accuracy within 5 meters through the wall, with a centimeter-level solution for screen content. Experiments have shown that WaveSpy can carry out screen attacks in different open environments with good results. The paper ends with experimental results and a summary, which shows that the proposed WaveSpy system has reliability, robustness, and efficiency in practice.

[107] mainly introduces the preliminary work of vibration measurement using millimeter-wave radar. In this paper, we propose a Multi-Signal Integration Model (MSC) that describes the properties of reflected signals in the In-phase and Quadrature (IQ) domains and uses the intrinsic consistency between these signals to accurately recover the vibrational signature. In the experiments, the authors evaluated the performance based on the amplitude and frequency error of the vibration signal. The frequency error reflects the correctness of the measured vibration signal, while the amplitude error reflects the accuracy. They discuss the limitations of mmVib and show the implementation details and evaluation results. mmVib enables micron-level vibration measurements (below 500Hz) in industrial environments.

[108] mainly introduces radio frequency (RF) microphones, a sound recovery technology based on millimeter-wave radar systems. The authors propose this emerging sound recovery technique by using a millimeter-wave 120GHz interferometric radar system, which can track micron-scale displacements of vibration sources, such as working speakers, and other objects vibrated by sound waves, such as window glass, in real-time. By recovering the displacement information from the radar signal, the original audio information can be strictly correlated and recovered, making it possible to reproduce the sound information. This technology uses a novel algorithm based on trigonometric function transformation to perform millimeter-wave linear phase modulation to overcome the phase ambiguity caused by the displacement of the vibrating object by more than half a wavelength. The authors used millimeter-wave radar to detect vibrations below 1kHz, and the results showed that the RF microphone performed well at sensing precise vibrations, but the radar sensor needed to be very close to the speaker ($\leq 5cm$) to eavesdrop.

[109], the method of using non-contact voiceprint recognition technology for speaker verification and the effectiveness of anti-fraud attacks are discussed. This paper introduces the traditional method of using voiceprint recognition for speaker verification, discusses the feasibility of using voiceprint and language models as biometric features, and the design method of using millimeter-wave radar to perform speaker verification. Among them, the authors can not only verify the speaker in a non-contact and unobtrusive way by using millimeter-wave radar for speaker verification but also better detect various spoofing attacks. The system enables accurate and robust speaker verification in IoT smart home applications and is highly resistant to fraud attacks. However, the system does not focus on speech reconstruction and has a frequency response of less than 200Hz.

[110] mainly discusses the principle of using radar sensors to measure the small vibrations of objects caused by sound and its application in detecting sound signals. Radar sensors convert objects into microphones, which detect and identify sound signals by measuring the tiny vibrations caused by sound pressure waves on the surface of the object. This method can not only realize the detection of sound signals but also be applied to remote eavesdropping and security monitoring. In this paper, the 24GHz FMCW radar is used to reconstruct the audio, but the experimental evaluation is insufficient.

In [59], an end-to-end noise-resistant speech sensing system, WaveEar, is proposed. WaveEar uses a 24GHz mmWave probe and Wave-voice Net deep neural network technology to achieve noise-resistant voice perception. Wave-voice Net can recover noise-free speech from received mmWave signals. The system converts the reflected 1-D time-domain signal into a 2-D spectrogram, then uses a neural network based on the residual architecture to learn and establish the mapping between the millimeter-wave spectrogram and the speech spectrum, and finally uses the phase reconstruction algorithm to recover the speech. WaveEar can reconstruct high-quality sound from the user's throat using millimeter-wave radar, but it requires the subject to remain stationary and short (less than 2m) away from the radar probe.

[111] discusses a new method for measuring the displacement characteristics of mechanical vibrations using frequency-modulated continuous-wave (FMCW) radar systems. The authors used a method based on instantaneous phase evaluation rather than a spectrum estimator to achieve high accuracy and accuracy in distance measurements. Compared to previous FMCW radar vibration measurement methods, this method allows for the measurement of frequencies well above the slow repetition rate of the radar system, similar to continuous wave (CW) radar or six-port radar systems. The advantage of this method is that multiple targets at different distances can be measured simultaneously without the need to use ultrafast but mostly noisy and nonlinear frequency chirps.

[112]applies a speech enhancement algorithm to improve the voice signal captured by custom millimeter-wave radar. The performance of radar sensors in detecting speech signals was investigated through experiments, and its application in speech recognition and speech pathology was discussed. The authors successfully extracted the vibration signals of human speech organs using different technical means such as micropower pulse radar, ultra-wideband radar, and Doppler radar system, and carried out experimental evaluation. They experimentally verified the effectiveness of radar sensors in speech detection, including the detection of speech signals such as Mandarin and the evaluation of signal quality. Through field experiments and evaluation of the experimental results, the authors discussed the feasibility and effectiveness of using radar sensors for voice detection. Experimental results show that the proposed algorithm has achieved good results in the processing of radar speech, and

has better performance compared with other algorithms after analyzing the speech signals in different noise environments.

[113] discusses the technical aspects of multimodal automatic speech recognition (ASR) systems in public applications of voice user interfaces (VUIs). The authors used millimeter wave and audio signals to fuse to design a multi-mode ASR system called Wavoice to achieve accurate speech recognition in complex environments such as noise and motion interference. The system combines millimeter waves and microphones by using machine learning. The intrinsic correlation between millimeter wave and audio signal is studied through mathematical modeling, and a real-time and anti-interference method for acoustic activity detection and user localization is proposed. By optimizing the multimodal fusion network based on attention mechanism and using cross-modal calibration, the robustness and perception distance of Wavoice are improved, so that the character recognition error rate within 7 meters is less than 1% under unfavorable conditions. Overall, the technical aspects of this paper mainly involve the design and optimization of an automatic speech recognition system based on millimeter wave and audio signal fusion for application in voice interaction interfaces in public places. However, the research in the aforementioned papers ([105], [107], [108], [110], and [113]) lacks comprehensive coverage of the frequency response of the human speech spectrum (300 Hz to 3.4 kHz) ([114]). Papers [111] and [112] depend on specialized hardware for line-of-sight audio reconstruction. Papers [59], [105], [106], [109], and [113] utilize machine learning[115] and necessitate prior knowledge and extensive datasets for model training. Consequently, none of the aforementioned studies fulfill the criteria for acoustic eavesdropping in real-world scenarios.

3. Comparison

In the discussion below, we will compare the three different eavesdropping methods mentioned in Section 3 in detail to better understand their advantages and disadvantages.

Motion sensor-based acoustic eavesdropping can capture sound waves by analyzing the tiny movements of an object's surface, which allows it to be monitored without direct contact with the sound source, improving the concealment of eavesdropping. However, it may be affected by environmental factors. Factors such as air movement, movement of other objects, etc., can interfere with the sensor's accuracy. In addition, there are high requirements for the sensor's accuracy and the processing algorithm[116].

Optical sensor-based acoustic eavesdropping uses optical devices, such as lasers, to capture tiny vibrations on the surface of an object from a distance, enabling covert remote eavesdropping. This approach is particularly useful in situations where a safety barrier needs to be crossed or the target cannot be reached. The disadvantage is that it requires a direct look at the target, and the reflective properties of the target surface can affect the eavesdropping effect. Optical eavesdropping devices are often bulky, expensive, and susceptible to ambient light and weather conditions[117].

RF-based-based acoustic eavesdropping enables eavesdropping on wireless communication devices by capturing and analyzing sound information in wireless signals. Some RF-based acoustic eavesdropping methods are not even limited by physical isolation. RF eavesdropping technology can cover a wide range and is suitable for a variety of wireless communication protocols and devices. However, it may require sophisticated signal processing and decryption techniques to extract sound information from communications. In addition, RF eavesdropping can be hampered by encryption technology and anti-eavesdropping measures[118].

4. Future Work

In this section, we discuss the shortcomings of current research and possible ways to improve them.

Motion sensor-based acoustic eavesdropping systems such as [47] require higher sampling rates to improve the system's accuracy. However, since increasing the sampling frequency may increase the amount of energy consumed by sampling, a trade-off should be made between accuracy and energy efficiency. [50] mentions that motion sensors are only affected by speech signals in certain situations, so more efficient signal processing algorithms can be studied to improve the ability of motion sensors to capture small vibrations, thereby improving the accuracy of acoustic eavesdropping through ambient vibrations. [48] studied multi-sensor fusion. It does not specifically look at the impact of TI-ADC signals but only highlights the risks of ubiquitous IoT devices. In the future, it is possible to study data fusion algorithms combining different types of sensors, such as accelerometers and gyroscopes, to improve the quality and accuracy of sound reconstruction. Motion sensor-based acoustic eavesdropping may require the development of efficient noise reduction algorithms to reduce the impact of ambient noise on sound capture and improve acoustic eavesdropping in noisy environments[119].

Acoustic eavesdropping based on optical sensors, such as [6], demonstrates the possibility of using optical sensors for eavesdropping, but further research is needed to improve the performance of eavesdropping systems. In [54], while the authors show how to extract tiny vibrations from the surface of a body from high-speed video to partially restore the sound-producing sound, further research may still be needed to improve the quality and stability of the recovered sound. In addition, in the LidarPhone study, while the researchers successfully implemented an attack using lidar sensors for eavesdropping, more research is likely needed to improve classification accuracy and applicability to different environmental conditions. Therefore, future research directions may include improving technology to improve the accuracy of sound reconstruction, improving the stability of optical systems[120] to enable them to work stably under different environmental conditions, studying how to counter these side-channel attacks, researching and developing optical eavesdropping technologies that can work effectively

at longer distances, and exploring new defense mechanisms and privacy protection measures[121].

RF-based acoustic eavesdropping can remotely sense and recover sound[97], but further research is still needed on how to deal with complex sound environments and expand the monitoring range in practical applications. It can achieve silent lip recognition [We can], but more research is needed to improve the accuracy of oral narration and the detection of multi-person conversations. It can achieve silent lip recognition[56], but more research is needed to improve the accuracy of oral narration and the detection of multi-person conversations. In addition, it can recover and separate sounds from multiple sources at the same time[58], but further verification is required for sensitivity and accuracy for real-time applications. Therefore, future research directions may include but are not limited to, improving sound recovery and separation algorithms to improve accuracy and real-time, developing more covert RF eavesdropping techniques to reduce the risk of detection[122], and researching new RF signal processing technologies to improve the ability of signals to penetrate obstacles such as walls and the resolution of sound signals.

5. Conclusion

This work provides a systematic investigation of acoustic eavesdropping methods. We comprehensively introduce and classify the various current acoustic eavesdropping methods, namely motion sensor-based acoustic eavesdropping, optical sensor-based acoustic eavesdropping, and RF-based acoustic eavesdropping. We describe the representative methods in detail and give the relevant formulas. We then analyze the advantages and disadvantages of these methods and discuss several future research directions for the research community to explore. This work can help researchers understand the classification and characteristics of various acoustic eavesdropping methods, and help researchers choose appropriate acoustic eavesdropping methods for research or application. In addition, this paper also provides reference and enlightenment for further exploration in the field of acoustic eavesdropping.

References

[1] JD Rudie, Zach Katz, Sam Kuhbander, and Suman Bhunia. Technical analysis of the nso group's pegasus spyware. In *2021 International Conference on Computational Science and Computational Intelligence (CSCI)*, pages 747–752. IEEE, 2021.

[2] Khlopov Oleg Anatolyevich. The cyber security and its role to protect critical infrastructure. *International journal of professional science*, (3):5–13, 2020.

[3] A Shaji George and S Sagayarajan. Acoustic eavesdropping: How ais can steal your secrets by listening to your typing. *Partners Universal International Innovation Journal*, 1(4):1–14, 2023.

[4] Jacob Leon Kröger and Philip Raschke. Is my phone listening in? on the feasibility and detectability of mobile eavesdropping. In *Data and Applications Security and Privacy XXXIII: 33rd Annual IFIP WG 11.3 Conference, DBSec 2019, Charleston, SC, USA, July 15–17, 2019, Proceedings 33*, pages 102–120. Springer, 2019.

[5] Jiadi Yu, Li Lu, Yingying Chen, Yanmin Zhu, and Linghe Kong. An indirect eavesdropping attack of keystrokes on touch screen through acoustic sensing. *IEEE Transactions on Mobile Computing*, 20(2):337–351, 2019.

[6] Ben Nassi, Yaron Pirutin, Adi Shamir, Yuval Elovici, and Boris Zadov. Lamphone: Real-time passive sound recovery from light bulb vibrations. *Cryptology ePrint Archive*, 2020.

[7] Karin Bijsterveld. Eavesdropping by the eye: detecting sound events and the culture of acoustic intelligence. *Sound Studies*, 9(2):233–252, 2023.

[8] Weigao Su, Daibo Liu, Taiyuan Zhang, and Hongbo Jiang. Towards device independent eavesdropping on telephone conversations with built-in accelerometer. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(4):1–29, 2021.

[9] Alberto Compagno, Mauro Conti, Daniele Lain, and Gene Tsudik. Don't skype & type! acoustic eavesdropping in voice-over-ip. In *Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security*, pages 703–715, 2017.

[10] Teng Wei, Shu Wang, Anfu Zhou, and Xinyu Zhang. Acoustic eavesdropping through wireless vibrometry. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, pages 130–141, 2015.

[11] Nirupam Roy, Haitham Hassanieh, and Romit Roy Choudhury. Backdoor: Making microphones hear inaudible sounds. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*, pages 2–14, 2017.

[12] Tzipora Halevi and Nitesh Saxena. Acoustic eavesdropping attacks on constrained wireless device pairing. *IEEE Transactions on Information Forensics and Security*, 8(3):563–577, 2013.

[13] Simone Soderi. Acoustic-based security: A key enabling technology for wireless sensor networks. *International Journal of Wireless Information Networks*, 27(1):45–59, 2020.

[14] Qiu Wang, Hong-Ning Dai, Xuran Li, Hao Wang, and Hong Xiao. On modeling eavesdropping attacks in underwater acoustic sensor networks. *Sensors*, 16(5):721, 2016.

[15] Kirsten Crager and Anindya Maiti. Information leakage through mobile motion sensors: User awareness and concerns. In *Proceedings of the European Workshop on Usable Security (EuroUSEC)*, 2017.

[16] Fatih Erden, Senem Velipasalar, Ali Ziya Alkar, and A Enis Cetin. Sensors in assisted living: A survey of signal and image processing methods. *IEEE Signal Processing Magazine*, 33(2):36–44, 2016.

[17] Andrea Cherubini and David Navarro-Alarcon. Sensor-based control for collaborative robots: Fundamentals, challenges, and opportunities. *Frontiers in Neurorobotics*, page 113, 2021.

[18] Amit Kumar Sikder, Giuseppe Petracca, Hidayet Aksu, Trent Jaeger, and A Selcuk Uluagac. A survey on sensor-based threats to internet-of-things (iot) devices and applications. *arXiv preprint arXiv:1802.02041*, 2018.

[19] Yang Bai, Li Lu, Jerry Cheng, Jian Liu, Yingying Chen, and Jiadi Yu. Acoustic-based sensing and applications: A survey. *Computer Networks*, 181:107447, 2020.

[20] Zhuo Chen, Cheng-Cheng Zhang, Bin Shi, Tao Xie, Guangqing Wei, and Jun-Yi Guo. Eavesdropping on wastewater pollution: Detecting discharge events from river outfalls via fiber-optic distributed acoustic sensing. *Water Research*, 250:121069, 2024.

[21] Muhammed Zahid Ozturk, Chenshu Wu, Beibei Wang, and KJ Ray Liu. Radiomic: Sound sensing via radio signals. *IEEE Internet of Things Journal*, 10(5):4431–4448, 2022.

[22] Xiao Zhang, Griffin Klevering, Xinyu Lei, Yiwen Hu, Li Xiao, and Guan-hua Tu. The security in optical wireless communication: A survey. *ACM Computing Surveys*, 2023.

[23] Richard A Roberts and Clifford T Mullis. *Digital signal processing*. Addison-Wesley Longman Publishing Co., Inc., 1987.

[24] Stuart J Russell and Peter Norvig. *Artificial intelligence a modern approach*. London, 2010.

[25] David Tse and Pramod Viswanath. *Fundamentals of wireless communication*. Cambridge university press, 2005.

[26] Zhongyuan Fang, Fei Gao, Haoran Jin, Siyu Liu, Wensong Wang, Ruochong Zhang, Zesheng Zheng, Xuan Xiao, Kai Tang, Liheng

Lou, et al. A review of emerging electromagnetic-acoustic sensing techniques for healthcare monitoring. *IEEE Transactions on Biomedical Circuits and Systems*, 2022.

[27] Peter Prince, Andrew Hill, Evelyn Piña Covarrubias, Patrick Doncaster, Jake L Snaddon, and Alex Rogers. Deploying acoustic detection algorithms on low-cost, open-source acoustic sensors for environmental monitoring. *Sensors*, 19(3):553, 2019.

[28] Jae Mun Sim, Yonnim Lee, and Ohbyung Kwon. Acoustic sensor based recognition of human activity in everyday life for smart home services. *International Journal of Distributed Sensor Networks*, 11(9):679123, 2015.

[29] Ying Shang, Maocheng Sun, Chen Wang, Jian Yang, Yuankai Du, Jichao Yi, Wenan Zhao, Yingying Wang, Yanjie Zhao, and Jiasheng Ni. Research progress in distributed acoustic sensing techniques. *Sensors*, 22(16):6060, 2022.

[30] Gongtian Shen, Zhanwen Wu, and Junjiao Zhang. Advances in acoustic emission technology. In *Springer Proc. Phys*, volume 179, pages 257–8. Springer, 2014.

[31] Connor Bolton, Yan Long, Jun Han, Josiah Hester, and Kevin Fu. Characterizing and mitigating touchtone eavesdropping in smartphone motion sensors. In *Proceedings of the 26th International Symposium on Research in Attacks, Intrusions and Defenses*, pages 164–178, 2023.

[32] Supriyo Chakraborty, Wentao Ouyang, and Mani Srivastava. Lightspy: Optical eavesdropping on displays using light sensors on mobile devices. In *2017 IEEE International Conference on Big Data (Big Data)*, pages 2980–2989. IEEE, 2017.

[33] Mengying Zhang, Gaomi Wu, Dipeng Ren, Ran Gao, Zhi-Mei Qi, and Xingdong Liang. An optical mems acoustic sensor based on grating interferometer. *Sensors*, 19(7):1503, 2019.

[34] João GV Teixeira, Ivo T Leite, Susana Silva, and Orlando Frazão. Advanced fiber-optic acoustic sensors. *Photonic sensors*, 4:198–208, 2014.

[35] Bo Zhang, Yunjiang Jia, Benlei Zhao, Xiaosong Zhu, and Yiwei Shi. Highly sensitive photoacoustic gas sensor with micro-embedded acoustic resonator for gas leakage detection. *Optics Letters*, 48(16):4201–4204, 2023.

[36] Wei-Han Chen and Kannan Srinivasan. Acoustic eavesdropping from passive vibrations via mmwave signals. In *GLOBECOM 2022-2022 IEEE Global Communications Conference*, pages 4051–4056. IEEE, 2022.

[37] Shawn M Bullock. Radar, modems, and air defense systems: Noise as a data communication problem in the 1950s. *Perspectives on Science*, 24(1):73–92, 2016.

[38] Bert Cox, Liesbet Van der Perre, Stijn Wielandt, Geoffrey Ottoy, and Lieven De Strycker. High precision hybrid rf and ultrasonic chirp-based ranging for low-power iot nodes. *EURASIP Journal on Wireless Communications and Networking*, 2020(1):1–24, 2020.

[39] Xiangtian Shen, Yuyong Xiong, Songxu Li, and Zhike Peng. Rfmicphone: Robust sound acquisition combining millimeter-wave radar and microphone. *IEEE Sensors Letters*, 6(11):1–4, 2022.

[40] Davide Carboni, Alex Gluhak, Julie A McCann, and Thomas H Beach. Contextualising water use in residential settings: A survey of non-intrusive techniques and approaches. *Sensors*, 16(5):738, 2016.

[41] P Beyer. Non-intrusive detection, the way forward. Southern African Transport Conference, 2015.

[42] Pengfei Hu, Hui Zhuang, Panneer Selvam Santhalingam, Riccardo Spolaor, Parth Pathak, Guoming Zhang, and Xiuzhen Cheng. Accear: Accelerometer acoustic eavesdropping with unconstrained vocabulary. In *2022 IEEE Symposium on Security and Privacy (SP)*, pages 1757–1773. IEEE, 2022.

[43] Pengfei Hu, Wenhao Li, Riccardo Spolaor, and Xiuzhen Cheng. mmecho: A mmwave-based acoustic eavesdropping method. In *Proceedings of the ACM Turing Award Celebration Conference-China 2023*, pages 138–140, 2023.

[44] Marco Crocco, Marco Cristani, Andrea Trucco, and Vittorio Murino. Audio surveillance: A systematic review. *ACM Computing Surveys (CSUR)*, 48(4):1–46, 2016.

[45] John E Ball. Low signal-to-noise ratio radar target detection using linear support vector machines (l-svm). In *2014 IEEE Radar Conference*, pages 1291–1294. IEEE, 2014.

[46] Yao Ge and PC Ching. Energy efficiency for proactive eavesdropping in cooperative cognitive radio networks. *IEEE Internet of Things Journal*, 9(15):13443–13457, 2022.

[47] Li Zhang, Parth H Pathak, Muchen Wu, Yixin Zhao, and Prasant Mohapatra. Accelword: Energy efficient hotword detection through accelerometer. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*, pages 301–315, 2015.

[48] Jun Han, Albert Jin Chung, and Patrick Tague. Pitchln: 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*, pages 181–192, 2017.

[49] Nirupam Roy and Romit Roy Choudhury. Listening through a vibration motor. In *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*, pages 57–69, 2016.

[50] S Abhishek Anand and Nitesh Saxena. Speechless: Analyzing the threat to speech privacy from smartphone motion sensors. In *2018 IEEE Symposium on Security and Privacy (SP)*, pages 1000–1017. IEEE, 2018.

[51] Yan Michalevsky Dan Boneh and Gabi Nakibly. Gyrophone: Recognizing speech from gyroscope signals.

[52] Andrew Kwong, Wenyuan Xu, and Kevin Fu. Hard drive of hearing: Disks that eavesdrop with a synthesized microphone. In *2019 IEEE symposium on security and privacy (SP)*, pages 905–919. IEEE, 2019.

[53] Héctor A Cordourier Maruri, Paulo Lopez-Meyer, Jonathan Huang, Willem Marco Beltman, Lama Nachman, and Hong Lu. V-speech: noise-robust speech capturing glasses using vibration sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(4):1–23, 2018.

[54] Abe Davis, Michael Rubinstein, Neal Wadhwa, Gautham J Mysore, Fredo Durand, and William T Freeman. The visual microphone: Passive recovery of sound from video. 2014.

[55] Sriram Sami, Yimin Dai, Sean Rui Xiang Tan, Nirupam Roy, and Jun Han. Spying with your robot vacuum cleaner: eavesdropping via lidar sensors. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*, pages 354–367, 2020.

[56] Guanhua Wang, Yongpan Zou, Zimu Zhou, Kaishun Wu, and Lionel M Ni. We can hear you with wi-fi! In *Proceedings of the 20th annual international conference on Mobile computing and networking*, pages 593–604, 2014.

[57] Chuyu Wang, Lei Xie, Yuancan Lin, Wei Wang, Yingying Chen, Yanling Bu, Kai Zhang, and Sanglu Lu. Thru-the-wall eavesdropping on loudspeakers via rfid by capturing sub-mm level vibration. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(4):1–25, 2021.

[58] Ziqi Wang, Zhe Chen, Akash Deep Singh, Luis Garcia, Jun Luo, and Mani B Srivastava. Uwhear: through-wall extraction and separation of audio vibrations using wireless signals. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*, pages 1–14, 2020.

[59] Chenhan Xu, Zhengxiong Li, Hanbin Zhang, Aditya Singh Rathore, Huining Li, Chen Song, Kun Wang, and Wenyao Xu. Waveear: Exploring a mmwave-based noise-resistant speech sensing for voice-user interface. In *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*, pages 14–26, 2019.

[60] Pengfei Hu, Yifan Ma, Panneer Selvam Santhalingam, Parth H Pathak, and Xiuzhen Cheng. Milliear: Millimeter-wave acoustic eavesdropping with unconstrained vocabulary. In *IEEE INFOCOM 2022-IEEE Conference on Computer Communications*, pages 11–20. IEEE, 2022.

[61] Suryoday Basak and Mahanth Gowda. mmspy: Spying phone calls using mmwave radars. In *2022 IEEE Symposium on Security and Privacy (SP)*, pages 1211–1228. IEEE, 2022.

[62] Shijia Zhang, Yilin Liu, and Mahanth Gowda. I spy you: Eavesdropping continuous speech on smartphones via motion sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 6(4):1–31, 2023.

[63] S Abhishek Anand, Chen Wang, Jian Liu, Nitesh Saxena, and Yingying Chen. Motion sensor-based privacy attack on smartphones. *arXiv preprint arXiv:1907.05972*, 2019.

[64] Vittorio MN Passaro, Antonello Cuccovillo, Lorenzo Vaiani, Martino De Carlo, and Carlo Edoardo Campanella. Gyroscope technology and applications: A review in the industrial perspective. *Sensors*, 17(10):2284, 2017.

[65] Kai Liu, Weiping Zhang, Wenyuan Chen, Kai Li, Fuyan Dai, Feng Cui, Xiaosheng Wu, Gaoyin Ma, and Qijun Xiao. The development of micro-gyroscope technology. *Journal of Micromechanics and Microengineering*, 19(11):113001, 2009.

[66] Anders Persson. How do we understand the coriolis force? *Bulletin of the American Meteorological Society*, 79(7):1373–1386, 1998.

[67] Chihwan Jeong, Seonho Seok, Byeungleul Lee, Hyeonched Kim, and Kukjin Chun. A study on resonant frequency and q factor tunings for mems vibratory gyroscopes. *Journal of micromechanics and microengineering*, 14(11):1530, 2004.

[68] Patrick L Walter. The history of the accelerometer. *Sound and vibration*, 31(3):16–23, 1997.

[69] Alessandro Sabato, Christopher Niezrecki, and Giancarlo Fortino. Wireless mems-based accelerometer sensor boards for structural vibration monitoring: A review. *IEEE Sensors Journal*, 17(2):226–235, 2016.

[70] I Arun Faisal, T Waluyo Purboyo, and A Siswo Raharjo Ansori. A review of accelerometer sensor and gyroscope sensor in imu sensors on motion capture. *J. Eng. Appl. Sci*, 15(3):826–829, 2019.

[71] Zhongjie Ba, Tianhang Zheng, Xinyu Zhang, Zhan Qin, Baochun Li, Xue Liu, and Kui Ren. Learning-based practical smartphone eavesdropping with built-in accelerometer. In *NDSS*, volume 2020, pages 1–18, 2020.

[72] Mordechai Guri, Yosef Solewicz, Andrey Daidakulov, and Yuval Elovici. {SPEAKE (a) R}: Turn speakers to microphones for fun and profit. In *11th USENIX Workshop on Offensive Technologies (WOOT 17)*, 2017.

[73] José Luís Santos and Faramarz Farahi. *Handbook of optical sensors*. Crc Press, 2014.

[74] Thomas D Rossing and Neville H Fletcher. Principles of vibration and sound, 2004.

[75] Yifei Zou, Zuyuan Zhang, Congwei Zhang, Yanwei Zheng, Dongxiao Yu, and Jiguo Yu. A distributed abstract mac layer for cooperative learning on internet of vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 2024.

[76] Michael Vollmer and Klaus-Peter Möllmann. High speed and slow motion: the technology of modern high speed cameras. *Physics Education*, 46(2):191, 2011.

[77] Javier Portilla and Eero P Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *International journal of computer vision*, 40:49–70, 2000.

[78] J Kim, KK Kwon, and Su In Lee. Trends and applications on lidar sensor technology. *Electronics and Telecommunications Trends*, 27(6):134–143, 2012.

[79] Stephen E Reutebuch, Hans-Erik Andersen, and Robert J McGaughey. Light detection and ranging (lidar): an emerging tool for multiple resource inventory. *Journal of forestry*, 103(6):286–292, 2005.

[80] Sergi Foix, Guillem Alenya, and Carme Torras. Lock-in time-of-flight (tof) cameras: A survey. *IEEE Sensors Journal*, 11(9):1917–1926, 2011.

[81] Jingyun Liu, Qiao Sun, Zhe Fan, and Yudong Jia. Tof lidar development in autonomous vehicle. In *2018 IEEE 3rd Optoelectronics Global Conference (OGC)*, pages 185–190. IEEE, 2018.

[82] Masaharu Imaki, Shumpei Kameyama, Eitaro Ishimura, Masaharu Nakaji, Hideo Yoshinaga, and Yoshihito Hirano. Line scanning time-of-flight laser sensor for intelligent transport systems, combining wide field-of-view optics of 30 deg, high scanning speed of 0.9 ms/line, and simple sensor configuration. *Optical Engineering*, 56(3):031205–031205, 2017.

[83] Tufan C Karalar and Jan Rabaey. An rf tof based ranging implementation for sensor networks. In *2006 IEEE International Conference on Communications*, volume 7, pages 3347–3352. IEEE, 2006.

[84] James D Spinhirne. Micro pulse lidar. *IEEE transactions on geoscience and remote sensing*, 31(1):48–55, 1993.

[85] Karen Elizabeth Joyce, SV Samsonov, Shaun R Levick, J Engelbrecht, and S Belliss. Mapping and monitoring geological hazards using optical, lidar, and synthetic aperture radar image data. *Natural hazards*, 73:137–163, 2014.

[86] Guangyu Zhao, Ming Lian, Yiyun Li, Zheng Duan, Shiming Zhu, Liang Mei, and Sune Svanberg. Mobile lidar system for environmental monitoring. *Applied Optics*, 56(5):1506–1516, 2017.

[87] J Reitberger, Cl Schnörr, M Heurich, P Krzystek, and U Stilla. Towards 3d mapping of forests: A comparative study with first/last pulse and full waveform lidar data. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci*, 37:1397–1404, 2008.

[88] Andrew G Stove. Linear fmcw radar techniques. In *IEE Proceedings F (Radar and Signal Processing)*, volume 139, pages 343–350. IET, 1992.

[89] Mustafa Mert Bayer and Ozdal Boyraz. Ranging and velocimetry measurements by phase-based mtcw lidar. *Optics Express*, 29(9):13552–13562, 2021.

[90] Ralph P Muscatell. Laser microphone. *The Journal of the Acoustical Society of America*, 76(4):1284–1284, 1984.

[91] Daniel M Dobkin and Titus Wandinger. A radio oriented introduction to radio frequency identification. *RFID Tutorial, High Frequency Electronics*, pages 46–54, 2005.

[92] Graham Brooker and Jairo Gomez. Lev termen's great seal bug analyzed. *IEEE Aerospace and Electronic Systems Magazine*, 28(11):4–11, 2013.

[93] Kazimierz Siwiak. Ultra-wide band radio: introducing a new technology. In *IEEE VTS 53rd Vehicular Technology Conference, Spring 2001. Proceedings (Cat. No. 01CH37202)*, volume 2, pages 1088–1093. IEEE, 2001.

[94] G Roberto Aiello and Gerald D Rogerson. Ultra-wideband wireless systems. *IEEE microwave magazine*, 4(2):36–47, 2003.

[95] Angela Digulescu, Cristina Despina-Stoian, Denis Stănescu, Florin Popescu, Florin Enache, Cornel Ioana, Emanuel Rădoi, Iulian Rîncu, and Alexandru Șerbănescu. New approach of uav movement detection and characterization using advanced signal processing methods based on uwb sensing. *Sensors*, 20(20):5904, 2020.

[96] Mohammad Ahmad Salamin, Wael AE Ali, Sudipta Das, and Asmaa Zugari. Design and investigation of a multi-functional antenna with variable wideband/notched uwb behavior for wlan/x-band/uwb and ku-band applications. *AEU-International Journal of Electronics and Communications*, 111:152895, 2019.

[97] Yu Rong, Sharanya Srinivas, Adarsh Venkataramani, and Daniel W Bliss. Uwb radar vibrometry: An rf microphone. In *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, pages 1066–1070. IEEE, 2019.

[98] Nabil Khalid, Rashid Mirzavand, and Ashwin K Iyer. A survey on battery-less rfid-based wireless sensors. *Micromachines*, 12(7):819, 2021.

[99] M Ayoub Khan, Manoj Sharma, and Brahmanandha R Prabhu. A survey of rfid tags. *International Journal of Recent Trends in Engineering*, 1(4):68, 2009.

[100] K Mandal and DL Atherton. A study of magnetic flux-leakage signals. *Journal of Physics D: Applied Physics*, 31(22):3211, 1998.

[101] Jeffrey R Foerster. The effects of multipath interference on the performance of uwb systems in an indoor wireless channel. In *IEEE VTS 53rd Vehicular Technology Conference, Spring 2001. Proceedings (Cat. No. 01CH37202)*, volume 2, pages 1176–1180.

IEEE, 2001.

[102] Jin-Ping Niu and Geoffrey Ye Li. An overview on backscatter communications. *Journal of Communications and Information Networks*, 4(2):1–14, 2019.

[103] Cesar Iovescu and Sandeep Rao. The fundamentals of millimeter wave sensors. *Texas Instruments*, pages 1–8, 2017.

[104] Chao Wang, Feng Lin, Tiantian Liu, Ziwei Liu, Yijie Shen, Zhongjie Ba, Li Lu, Wenyao Xu, and Kui Ren. mmphone: Acoustic eavesdropping on loudspeakers via mmwave-characterized piezoelectric effect. In *IEEE INFOCOM 2022-IEEE Conference on Computer Communications*, pages 820–829. IEEE, 2022.

[105] Rohan Khanna, Daegun Oh, and Youngwook Kim. Through-wall remote human voice recognition using doppler radar with transfer learning. *IEEE Sensors Journal*, 19(12):4571–4576, 2019.

[106] Zhengxiong Li, Fenglong Ma, Aditya Singh Rathore, Zhuolin Yang, Baicheng Chen, Lu Su, and Wenyao Xu. Wavespy: Remote and through-wall screen attack via mmwave sensing. In *2020 IEEE Symposium on Security and Privacy (SP)*, pages 217–232. IEEE, 2020.

[107] Chengkun Jiang, Junchen Guo, Yuan He, Meng Jin, Shuai Li, and Yunhao Liu. mmvib: micrometer-level vibration measurement with mmwave radar. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, pages 1–13, 2020.

[108] Li Wen, Yuchen Li, Yangtao Ye, Changzhan Gu, and Jun-Fa Mao. Audio recovery via noncontact vibration detection with 120 ghz millimeter-wave radar sensing. In *2021 International Conference on Microwave and Millimeter Wave Technology (ICMMT)*, pages 1–3. IEEE, 2021.

[109] Yudi Dong and Yu-Dong Yao. Secure mmwave-radar-based speaker verification for iot smart home. *IEEE Internet of Things Journal*, 8(5):3500–3511, 2020.

[110] Eloi Guerrero, Josep Brugués, Jordi Verdú, and Pedro de Paco. Microwave microphone using a general purpose 24-ghz fmcw radar. *IEEE Sensors Letters*, 4(6):1–4, 2020.

[111] Lukas Piotrowsky, Jan Siska, Christian Schweer, and Nils Pohl. Using fmcw radar for spatially resolved intra-chirp vibrometry in the audio range. In *2020 IEEE/MTT-S International Microwave Symposium (IMS)*, pages 791–794. IEEE, 2020.

[112] Fuming Chen, Sheng Li, Chuantao Li, Miao Liu, Zhao Li, Huijun Xue, Xijing Jing, and Jianqi Wang. A novel method for speech acquisition and enhancement by 94 ghz millimeter-wave sensor. *Sensors*, 16(1):50, 2015.

[113] Tiantian Liu, Ming Gao, Feng Lin, Chao Wang, Zhongjie Ba, Jinsong Han, Wenyao Xu, and Kui Ren. Wavoice: A noise-resistant multi-modal speech recognition system fusing mmwave and audio signals. In *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*, pages 97–110, 2021.

[114] Ronald J Baken. Clinical measurement of speech and voice. *(No Title)*, 1987.

[115] Zuyuan Zhang, Hanhan Zhou, Mahdi Imani, Taeyoung Lee, and Tian Lan. Collaborative ai teaming in unknown environments via active goal deduction. *arXiv preprint arXiv:2403.15341*, 2024.

[116] Amit Kumar Sikder, Giuseppe Petracca, Hidayet Aksu, Trent Jaeger, and A Selcuk Uluagac. A survey on sensor-based threats and attacks to smart devices and applications. *IEEE Communications Surveys & Tutorials*, 23(2):1125–1159, 2021.

[117] Syed Agha Hassnain Mohsan, Alireza Mazinani, Hassaan Bin Sadiq, and Hussain Amjad. A survey of optical wireless technologies: Practical considerations, impairments, security issues and future research directions. *Optical and Quantum Electronics*, 54(3):187, 2022.

[118] Michael Eoin Buckley and Shirook M Ali. Method and apparatus for anti-eavesdropping in vunerable nfc applications, March 15 2016. US Patent 9,287,935.

[119] Huining Li, Chenhan Xu, Aditya Singh Rathore, Zhengxiong Li, Hanbin Zhang, Chen Song, Kun Wang, Lu Su, Feng Lin, Kui Ren, et al. Vocalprint: A mmwave-based unmediated vocal sensing system for secure authentication. *IEEE Transactions on Mobile Computing*, 22(1):589–606, 2021.

[120] Andres Guesalaga, Benoit Neichel, Maxime Boccas, Celine d'Orgeville, Francois Rigaut, Dani Guzman, and Jaime Anguita. Improving stability, robustness, and performance of laser systems. In *Adaptive Optics Systems III*, volume 8447, pages 1519–1529. SPIE, 2012.

[121] Chris L Willis. Boresight stability of an optical system, August 24 2004. US Patent 6,781,773.

[122] Mohammad Vahid Jamali and Hessam Mahdavifar. Covert millimeter-wave communication: Design strategies and performance analysis. *IEEE Transactions on Wireless Communications*, 21(6):3691–3704, 2021.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The author is an Editorial Board Member/Editor-in-Chief/Associate Editor/Guest Editor for *[Journal name]* and was not involved in the editorial review or the decision to publish this article.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: