Glowworm Attack: Optical TEMPEST Sound Recovery via a Device's Power Indicator LED

Ben Nassi[1], Yaron Pirutin[1], Tomer Galor[2], Yuval Elovici[1] , Boris Zadov[1]
[1] Ben-Gurion University of the Negev, [2] Weizmann Institute of Science
Website: https://www.nassiben.com/glowworm-attack
{nassib, yaronpir, tomercg, elovici, zadov}@post.bgu.ac.il

Abstract

Two main classes of optical TEMPEST attacks against the confidentiality of information processed/delivered by devices have been demonstrated in the past two decades; the first class includes methods for recovering content from monitors, and the second class includes methods for recovering keystrokes from physical and virtual keyboards. In this paper, we identify a new class of optical TEMPEST attacks: recovering sound by analyzing optical emanations from a device's power indicator LED. We analyze the response of the power indicator LED of various devices to sound and show that there is an optical correlation between the sound that is played by connected speakers and the intensity of their power indicator LED due to the facts that: (1) the power indicator LED of various devices is connected directly to the power line, (2) the intensity of a device's power indicator LED is correlative to the power consumption, and (3) many devices lack a dedicated means of countering this phenomenon. Based on our findings, we present the Glowworm attack, an optical TEMPEST attack that can be used by eavesdroppers to recover sound by analyzing optical measurements obtained via an electro-optical sensor directed at the power indicator LED of various devices (e.g., speakers, USB hub splitters, and microcontrollers). We propose an optical-audio transformation (OAT) to recover sound by isolating the speech from the optical measurements obtained by directing an electro-optical sensor at a device's power indicator LED. Finally, we test the performance of the Glowworm attack in various experimental setups and show that an eavesdropper can apply the attack to recover speech from a speaker's power LED indicator with good intelligibility from a distance of 15 meters and with fair intelligibility from 35 meters.

Keywords

Privacy, TEMPEST, Sound Recovery

1 Introduction

Optical TEMPEST attacks [23], which are methods aimed at recovering information from systems through optical side effects, pose a great risk to privacy. In the past two decades, various studies have demonstrated novel techniques for recovering/extracting information from victim devices using optical sensors by exploiting the correlation between the optical side effects of the information and the device that is used to deliver/process the information. In this context, two main classes of attacks were introduced; the first class includes methods for recovering content from monitors [15, 17, 38]. The second class includes methods for recovering keystrokes from physical and virtual keyboards [18, 46, 52, 53, 56, 64, 65]. In these studies, optical data that was obtained directly from the victim device's optical emanations (e.g., [38]) or indirectly from reflections of the victim device's optical emanations on nearby objects (e.g., [15, 17]) was used to recover the desired information from a victim device. These studies have contributed to improved understanding regarding the risks posed by optical TEMPEST attacks.

In this paper, we identify a new class of optical TEMPEST attacks: sound recovery by analyzing optical emanations obtained from a device's power indicator LED. We show that the power indicator LED of various devices leaks information regarding the sound played by connected speakers. This occurs in devices whose power indicator LED is connected directly to the device's power line and lack integrated voltage stabilizers. As a result, the optical response (intensity) of the power indicator LED of such devices is correlative to the power consumed by the device. This fact can be exploited to recover sound from the connected speakers directly, by obtaining optical measurements via an electro-optical sensor directed at the speakers' power indicator LED, or indirectly, by obtaining optical measurements via an electro-optical sensor directed at the power indicator LED of the device used to supply power to the speakers (e.g., USB hub, microcontrollers).

Previous studies have discussed the risks a device's power indicator LED can pose to the information delivered/processed by the device due to the linear response of the power indicator LED [35, 43]. This fact was exploited in some studies to establish covert channels by using a preinstalled malware that modulated the data via a device's power indicator LED [29–31], however no prior work was able to demonstrate end-to-end sound recovery from a commercial device's power indicator LED without the use of malware. Other studies [22, 47, 48, 55] presented optical methods for recovering sound by turning nearby objects into diaphragms (e.g., a hanging light bulb [48], bag of chips [22], trash can [55], glass window [47]). In these studies, sound was recovered by obtaining optical measurements from vibrating objects (objects vibrate when sound waves hit their surface). However, each of these methods [22, 47, 48, 55] suffer from one or more of the following limitations: (1) they are limited in range (the nearby vibrating object must be within five centimeters of the sound source [22, 48]), (2) their application can be detected by an optical sensor (because they require

CCS '21, November 14–19, 2021, Seoul, South Korea

ACM ISBN YYYY…$15.00
https://doi.org/XXXX/XXXXXX

the eavesdropper to direct a laser beam into the victim's room [55]), (3) they require the eavesdropper to compromise a device with malware (to stop the LiDAR from turning so it can target a specific point or to exifltrate the collected data via the Internet [55]). In addition, all of the methods target the optical changes resulting from minuscule vibrations of nearby objects that were affected by sound hitting their surface. We consider these methods to be optical side-channel attacks rather than optical TEMPEST attacks, because they do not target the optical correlation between the information and the device used to deliver/process the information. A recent paper presented a state-of-the-art sound recovery method [21] using an EMR TEMPEST attack against devices that contain an MSOC (mixed-signal system-on-chip) with an integrated switching regulator from a distance of 15 meters, however, to the best of our knowledge, no study has proposed a method capable of recovering sound from a device at distances greater than 15 meters using a TEMPEST attack.

In this paper, we present the Glowworm attack, an optical TEMPEST attack aimed at recovering sound played by commercial speakers. First, we analyze speakers and show that the intensity of their power indicator LED is affected by played audio. Then, we analyze various devices (USB hub splitter, micro-controller) used to supply power for the speakers and show that the intensity of the devices' power indicator LED is also affected by audio played by the speakers. Then, we suggest an optical-audio transformation (OAT) to isolate the sound from the optical signal obtained via an electro-optical sensor directed at the power indicator LED of the devices. Finally, we examine the performance of the Glowworm attack in various experimental setups. We show that it can be used by an eavesdropper to recover speech from virtual meetings by obtaining optical measurements directly from the power LED indicator of speakers with good intelligibility from a distance of 15 meters and with fair intelligibility from 35 meters.

In this paper, we make the following contributions: (1) We reveal a new class of optical TEMPEST attack that violates the confidentiality of the information processed/delivered by devices; the attack should be addressed by hardware manufacturers and considered by consumers. (2) We show that optical TEMPEST attack can recover sound from a device at greater distances (35 meters) than an existing SOTA method that used an EMR TEMPEST attack (whose range was limited to 15 meters) [21]. (3) We propose a new method for recovering speech that is external (obtains data without compromising a device in the target room), passive (does not rely on an active sensor), and does not depend on the distance between a sound source and a nearby object.

The remainder of the paper is structured as follows: In Section 2, we review related work. In Section 3, we present the threat model. In Section 4, we analyze the response of a device's power indicator LED to sound played by speakers. In Section 5, we present an optical audio transformation (OAT) for recovering sound, and in Section 6, we evaluate Glowworm's performance on the task of recovering sound. In Section 7, we discuss potential improvements that can be made to optimize the quality of the sound recovered by an eavesdropper. In Section 8, we suggest countermeasure methods that can be applied to prevent the proposed Glowworm attack. In Section 9, we present the responsible disclosure we made. In Section 10, we discuss the limitations of the attack and mention future work directions.

2 Related Work

In this section, we review related work in the area of optical data leakage and sound eavesdropping. TEMPEST attacks have attracted the interest of many researchers since Van Eck's paper was published in the mid-1980s [60]. The last three and a half decades have seen the development of various methods for extracting information from devices by exploiting the correlation between the information delivered/processed by a device and its: EMR emanations (e.g., [20, 21, 26]), acoustic emanations (e.g., [13, 16, 25, 32, 57, 67]), vibrations (e.g., [12, 14, 42, 45, 62, 66]), and power consumption (e.g., [36, 37, 44]). In the past two decades, two main classes of optical TEMPEST attacks were introduced; the first class includes methods for recovering content from monitors [15, 17, 38], and the second class includes methods for recovering keystrokes from physical and virtual keyboards [18, 46, 52, 53, 56, 64, 65]. In these studies, optical data that was obtained directly from the victim device's optical emanations (e.g., [38]) or indirectly from reflections of the victim device's optical emanations on nearby objects (e.g., [15, 17]) was used to recover the desired information from a victim device.

The risks posed by a device's power indicator LED were discussed by [35, 43]. However, prior research demonstrating methods capable of exploiting a device's power indicator LED for data exifltration relied on preinstalled malware to establish optical covert channels [29–31]. The proposed methods leak data from devices that are connected to air-gapped networks by using preinstalled malware that modulated data optically via the integrated LED of a device (e.g., a keyboard [29], router [30], hard drive [31]).

Recent studies have investigated sound eavesdropping [12, 14, 28, 39, 45, 54, 66], suggesting various methods for recovering sound by analyzing the side effects of sound waves that caused nearby lightweight objects (e.g., a bag of chips, a window) and devices (e.g., motion sensors) to vibrate (turning such objects/devices to diaphragms). In this context, malware was used to recover sound by: (1) obtaining data from a device's motion sensors [12, 14, 45, 66], (2) reprogramming a computer's audio port from output to input [28], (3) inverting the process of a vibration motor [54], and (4) analyzing magnetic data obtained from a hard disk head [39]. These methods pose a serious threat to privacy, but they require the eavesdropper to compromise a device (with malware) located near the victim (sound source) in order to obtain data and exfiltrate it to the eavesdropper.

Optical methods for sound recovery were introduced by [22, 47, 48, 55]. A recent study demonstrated a method capable of classifying words from a precollected dictionary, by analyzing the vibrations of a trash can using optical data obtained via a robotic vacuum cleaner's LiDAR. This method requires the eavesdropper to compromise the robotic vacuum cleaner in order to: (1) prevent the LiDAR from turning and fix the LiDAR on a specific object to increase the amount of data collected from the vibrating object (because the frequency of a robotic vacuum cleaner's 360° LiDAR is limited to 7 Hz), and (2) exfiltrate the data from the robotic vacuum cleaner. Three studies [22, 47, 48] presented external optical methods to recover sound that rely on data obtained via optical sensors, without the use of malware. The laser microphone [47, 47]

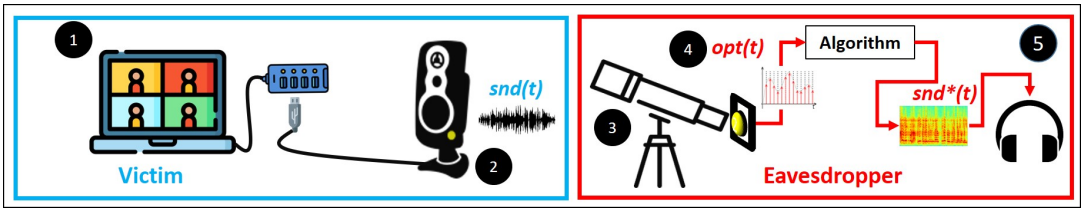

Figure 1: Glowworm's threat model: The sound $snd(t)$ of the virtual meeting (1) which is played by the connected speakers creates changes in the power consumption of the power indicator LED of a (2) connected peripheral (e.g., the speakers themselves, a USB hub splitter). The eavesdropper directs an electro-optical sensor at the power indicator LED of a connected peripheral using a telescope (3). The optical signal $opt(t)$ is sampled from the electro-optical sensor via an ADC (4) and processed, using an algorithm to recover the acoustic signal $snd^*(t)$ (5).

is a well-known method that recovers sound using an external laser transceiver in which a laser beam is directed through a window into a target room; the laser beam is reflected off an object and returned to the transceiver which then converts the beam to an audio signal. The visual microphone [22] recovers sound by analyzing the vibrations of material inside the victim's room (e.g., a bag of chips, water) using video obtained from a high-speed video camera (2200 FPS) to recover speech. Lamphone [48] recovers sound using a remote electro-optical sensor by exploiting the vibrations of a hanging light bulb; the vibrations cause optical changes due to the non-uniform intensity of lighting, which varies at each angle. These methods [22, 47, 48] pose a great privacy threat, however from an eavesdropper's perspective, they are limited in one of the following ways: they rely on (1) a very high sound level (over 100 dB) which is beyond the sound level of speech and meetings (e.g., [22, 48]), (2) active sensors that use a laser beam (e.g., [47]), a fact that increases the likelihood of detection (compared to passive sensors), (3) hanging light bulbs, which are not commonly used in office settings today (e.g., Lamphone [48]), or (4) specialized equipment for spying [47], a fact that may limit their availability in some countries (limiting the sale of such equipment to, e.g., police departments).

3 Threat Model

In this section, we describe the threat model and explain its significance with respect to other methods. The Glowworm attack targets the speech of participants in virtual meeting platforms (e.g., Zoom, Google Meet, Skype, Microsoft Teams). During the COVID-19 pandemic, these platforms became a popular way for people to meet and share information; personal and valuable information is routinely exchanged when these platforms are used for personal and business meetings.

We assume that an individual is located inside a room or office and using his/her computer to conduct a virtual meeting with another person (or a group) using a virtual meeting platform. The purpose of the conversation can vary, for example, the individuals may want to discuss business (e.g., sharing something with a client or colleague) or something of a personal nature (e.g., talking about medical test results with a doctor).

We consider an eavesdropper that is a malicious entity interested in recovering speech from meetings and using the valuable information discussed in the meeting for a malicious purpose that may include spying on individuals (e.g., to obtain sensitive information that can be used for blackmail) or spying on an organizations (e.g., to obtain a company's IP and use it to give a competitor an advantage). We assume the eavesdropper is located within 35 meters of the target room. The eavesdropper can be: (1) a person located in a room in an adjacent building, (2) a person in a nearby car. We consider this threat as highly probable in the COVID-19 era due to the number of personal and business meetings being held in unsecured settings, including home offices.

In order to recover the sound in this scenario, the eavesdropper performs the Glowworm attack. We assume that the a power indicator LED of a vulnerable device is visible from outside the room/office. We consider two types of attacks: (1) a direct attack, where the eavesdropper recovers sound from the power indicator LED of the speakers, and (2) an indirect attack, where the eavesdropper recovers sound from the power indicator LED of the device used to provide the power to the speakers (e.g., a connected USB hub, a microcontroller). Note that the Glowworm attack can be applied by eavesdroppers to recover: (1) the speech of any person speaking to the victim during a virtual meeting, and (2) any sound (e.g., music from YouTube, videos from the Internet) that is played by the speakers during the virtual meeting, which may or may not be related to the meeting; in this paper, we present the attack in the context of recovering speech from a virtual meeting.

The main components used to perform the Glowworm attack are: (1) A telescope - This piece of equipment is used to focus the field of view on a device's power indicator LED from a distance. (2) An electro-optical sensor - This sensor is mounted on the telescope and consists of a photodiode that converts light into an electrical current; the current is generated when photons are absorbed in the photodiode. (3) A sound recovery system - This system receives an optical signal as input and outputs the recovered acoustic signal. The eavesdropper can implement such a system with: (a) dedicated hardware (e.g., using capacitors, resistors), or (b) the use of ADC to sample the electro-optical sensor and process the data using a sound recovery algorithm running on a laptop. In this study, we use the latter digital approach.

Fig. 1 outlines the threat model: The sound $snd(t)$ played by the speakers in the victim's room results in changes in the power consumption due to the direct connection of the power indicator LED to the input power line and the device's lack of voltage stabilizers. These changes in power consumption influence the intensity of the light produced by the device's power indicator LED, resulting in a pattern of changes over time that the eavesdropper measures with an optical sensor which is directed at a device's power indicator

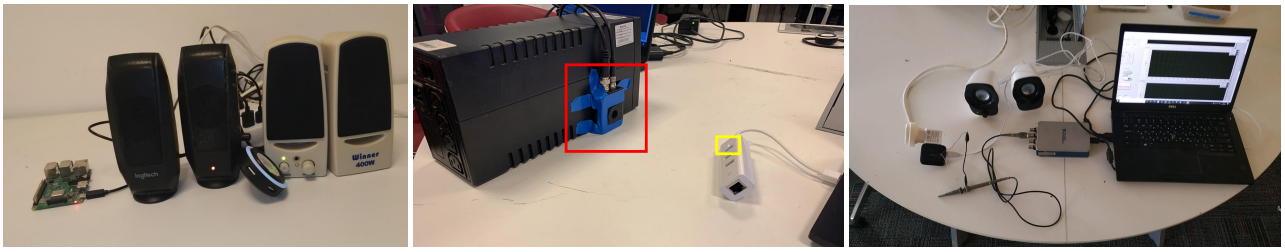

Figure 2: Left: Four of the devices examined in the experiments described in Section 4. Center: A Thorlabs PDA100A2 electro-optical sensor (boxed in red) is directed at the power indicator LED of a USB hub splitter (boxed in yellow). Right: The USB adapter is connected to the (1) speakers, (2) power socket, and (3) ADC via a BNC cable which is used to measure the power consumed by the USB hub splitter.

LED via a telescope. The analog output of the electro-optical sensor is sampled by the ADC to a digital optical signal $opt(t)$. The eavesdropper then processes the optical signal $opt(t)$, using an optical-audio transformation, to an acoustic signal $snd^*(t)$.

In order to keep the digital processing as light as possible in terms of computation, we sample the electro-optical sensor with the ADC set at the minimal sampling frequency allowing comprehensible audio recovery; Glowworm is aimed at recovering speech, and this requires a sufficient sampling frequency rate. The spectrum of speech covers quite a wide portion of the audible frequency spectrum. Speech consists of vowel and consonant sounds; the vowel sounds and the cavities that contribute to the formation of the different vowels range from 85 to 180 Hz for a typical adult male and from 165 to 255 Hz for a typical adult female. In terms of frequency, the consonant sounds are above 500 Hz (more specifically, in the 2-4 KHz frequency range) [2]. As a result, a telephone system samples an audio signal at 8 KHz. However, many studies have shown that an even lower sampling rate is sufficient for recovering comprehensible sound (e.g., 2200 Hz for the visual microphone [22]). In this study, we sample the electro-optical sensor at a sampling rate of 4/8 KHz.

The significance of Glowworm's threat model with respect to related work is that Glowworm is:

(1) Not dependent on the distance between a sound source and a nearby object: Glowworm analyzes the intensity of a device's power indicator LED, which is affected by a device's power consumption. As a result, the attack is not limited based on the required distance between a sound source and a nearby lightweight object (diaphragm) that vibrates in response to sound (as opposed to other sound recovery methods that are limited in that there can be no more than one meter between the sound source and a vibrating object [12, 14, 22, 39, 45, 47, 48, 54, 55, 61, 63, 66]).

(2) External: Glowworm does not rely on compromising a device to obtain the data needed to recover sound (as opposed to other sound recovery methods that require eavesdroppers to compromise a device with malware first [12, 28, 39, 45, 54, 66]).

(3) Passive and relies on a benign sensor: The method relies on a passive electro-optical sensor that is not considered spying equipment and gives no indication regarding its application (as opposed to the laser microphone [47] in which a laser beam is directed at a glass window).

(4) Capable of recovering speech without the need to compile a dictionary: Glowworm can be used to recover any speech (as opposed to other methods that are limited to classifying isolated words contained in a precompiled dictionary [12, 45, 61, 66]).

(5) Not dependent on being withing hearing range: Glowworm can be applied by eavesdroppers that are located beyond hearing range, from a distance of 15-35 meters (as opposed to other methods that require the eavesdropper to be located within 15 meters of the victim [21, 22]).

(6) Capable of recovering speech at a virtual meeting's sound level of 70 dB (in contrast to other methods that can only be used to recover sound at a high volume [22, 39, 48]).

4 Analysis

In this section, we describe the series of experiments performed to evaluate the risk of optical sound recovery posed by the vulnerability of the power indicator LED of various devices. The experiments analyze: (1) the influence of sound played from speakers on the power consumption of various devices, (2) the response of the device's power indicator LED to sound, and (3) the side effects added to the optical signal which are not the result of sound played from the speakers.

The devices used in these experiments are: Logitech S120 speakers [4], Winner speakers [10], a TP-Link UE330 USB hub splitter [9], a MIRACASE MHUB500 USB hub splitter [5], a Raspberry Pi (RP) 4, a Google Nest Mini [11], and Creative Pebble speakers [1]. Four of the devices are presented in Fig. 2.

The experiments were conducted as follows: An electro-optical sensor (the Thorlabs PDA100A2 [8], which is an amplified switchable gain light sensor that consists of a photodiode which is used to convert light/photons to electrical voltage) was directed at the power indicator LED of each device. The voltage was obtained from the electro-optical sensor using a 24-bit ADC NI-9234 card [7] and processed in a LabVIEW script that we wrote. The internal gain of the electro-optical sensor was set at the highest level before reaching saturation. The setup is presented in Fig. 2.

4.1 Understanding How Played Sound Affects the Power Consumption

Here we explore the effect of played sound on a device's power consumption and show that it linearly affects the device's power indicator LED due to the fact that hardware manufacturers do not integrate any voltage stabilizers or filters in some products. We

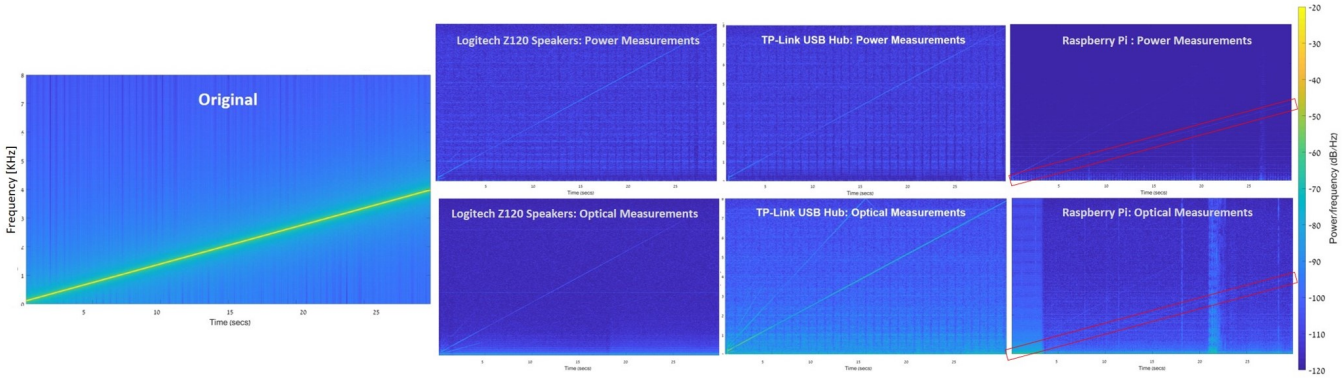

Figure 3: The six spectrograms on the right are obtained from power (upper row) and optical (bottom row) measurements of three devices when the speakers played a frequency scan (0-4 KHz) on the left.

show that optical measurements can be used to recover sound using an electro-optical sensor directed at a device's power indicator LED and eliminate any other reasonable side effects that could explain this phenomenon.

4.1.1 How Played Sound Affects a Device's Power Consumption and the Intensity of Its Power Indicator LED Here, we show that the intensity of a device's power indicator LED is highly correlated with the device's power consumption.

Experimental Setup: We created a USB adapter that allows us to obtain power measurements from any device with a USB input connector (the connector can be seen in Fig. 2). We conducted three experiments. In the first experiment, the Logitech speakers' USB was connected to the adapter which was connected directly to the electricity. In the second experiment, the Logitech speakers' USB was connected to the TP-Link USB hub splitter. The input USB connector of the USB hub splitter was connected to the adapter which was connected to a PC. In the third experiment, the Logitech speakers' USB was connected to the RP. The input micro USB connector of the RP was connected to the adapter which was connected to the electricity. In all of these experiments, the audio was played via the speakers (at 70 dB). The audio played from the speakers is a 30 second audio file that consists of a chirp function (a frequency scan between 0-4 KHz). We obtained optical measurements via the electro-optical sensor which was directed at the power indicator LED of the speakers (in the first experiment), the USB hub splitter (second experiment), and the RP (third experiment). In addition, in each of the experiments, we obtained power measurements from the adapter by connecting it to a BNC cable that was connected to a 24-bit ADC NI-9234 card [7]. The ADC was used to obtain optical and electrical measurements simultaneously from each tested device.

Results: Fig. 3 presents (1) three spectrograms extracted from the optical signal, and (2) three spectrograms extracted from the power signal. As can be seen, the chirp function played by the speakers affected the power consumption of all three devices. In the case of the RP, the frequency scan that was played by the speakers (0-4 KHz) can be spotted in the power consumption signal). In the cases of the TP-Link USB hub splitter and Logitech speaker, a frequency scan between 0-8 KHz can be seen in the power consumption signal (we discuss this phenomenon later in this section). Moreover, as shown in Fig. 3, the intensity of the power indicator LED of the devices is perfectly correlated with the power consumed by the devices (excluding some optical noise which will be discussed later in the paper).

Conclusions: Based on these experiments, we concluded that: (1) The power consumed by the three devices correlates with the sound that the speakers play and the intensity of their power indicator LED. (2) The manufacturers of these devices do not distort/change the known linear response of the intensity of an LED to power consumption [35] by integrating filters and voltage stabilizers into the electrical circuits. (3) The power consumed by the speakers influences the devices providing the power to the speakers (e.g., USB hub splitter, RP). (4) The linear correlation between the power consumed by the device, the audio played, and the intensity of the power indicator LED of the devices shows sound can be recovered by obtaining optical measurements via an electro-optical sensor directed at a device's power indicator LED.

4.1.2 Ruling Out Other Possible Side Effects One might argue that the optical measurements are affected by a phenomenon unrelated to the changes in the intensity of a device's power indicator LED. For example, one reasonable argument is that electromagnetic radiation was emitted from the device and was captured by the electro-optical sensor. Another reasonable argument is that the optical sensor captures minuscule vibrations of the power indicator LED caused by the device's vibrations due to the sound waves produced from the speakers. In order to disprove these claims, we conducted the following set of experiments.

Experimental Setup: We placed an RP on a table and directed the electro-optical sensor at its power indicator LED from a distance of one meter (through a telescope with a 15 cm lens diameter). We connected the USB cable of the Logitech speakers to the RP which was connected to the electricity on the other end. The speakers were placed on a different surface (than the RP) in order to eliminate any vibration resulting from the sound waves produced by the speakers. The speakers played an audio file consisting of a chirp function (a frequency scan between 200-400 Hz).

We conducted the following three experiments: In the first experiment, we obtained optical measurements when the electro-optical

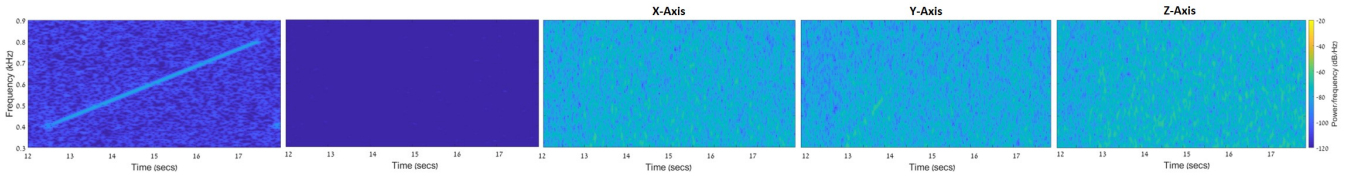

Figure 4: From left to right: Spectrograms obtained from optical measurements when the RP's power indicator LED was visible (first) and covered (second). Spectrograms were obtained from gyroscope measurements from three axes.

sensor was directed at the RP, however we covered the device's power indicator LED with tape (to examine whether the played signal appears in the optical measurements and rule out any effect of EMR). In the second experiment, we attached a gyroscope (MPU-6050 GY-521 [6]) to the RP to measure its vibrations (to examine whether the played signal appears in the gyroscope measurements and rule out any effect of vibration). We obtained measurements from the gyroscope via another RP which was used to sample the gyroscope at 1000 Hz. In the third experiment, we obtained optical measurements when the electro-optical sensor was directed at the RP's power indicator LED (in order to prove that the played signal can be spotted in the optical measurements). The third experiment was done for validation.

Results: The results are presented in Fig. 4. As can be seen, the frequency scan appears in the spectrogram obtained from the optical measurements when a device's power indicator LED is visible. However, the frequency scan cannot be spotted in the spectrograms obtained from (1) the optical measurements when a device's power indicator LED is covered, or (2) the gyroscope measurements in each of the three axes.

Conclusions: Based on these experiments, we concluded that (1) the optical measurements are not affected by electromagnetic radiation (if they were, the frequency scan would have appeared in the spectrogram when the power indicator LED was covered with tape); and (2) the optical measurements are not affected by the vibration caused by the sound waves produced from the speakers (if they were, the frequency scan would have appeared in the spectrograms obtained from the gyroscope on at least one of the three axes). These experiments prove that the frequency scan in the optical measurements obtained from a device's visible power indicator LED is the result of changes in a device's power consumption which linearly affect the intensity of the LED.

4.2 Exploring the Optical Response

Here we explore the recovered optical signal, including the baseline, side effects added, and SNR (signal-to-noise ratio).

4.2.1 Characterizing the Optical Signal When No Sound Is Played Here we examine the characteristics of the optical signal when no sound is played.

Experimental Setup: We obtained five seconds of optical measurements via an electro-optical sensor directed at the power indicator LED of four devices.

Results: The FFT graphs extracted from the optical measurements of the devices when no sound was played are presented in Fig. 5. As can be seen, a peak appears in the FFT at around 100 Hz; this peak is the result of the fixed light frequency of the LED. Since the

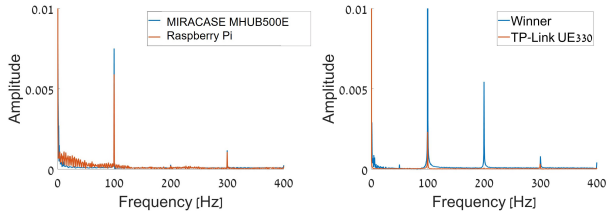

Figure 5: FFT graphs extracted from optical measurements of the power indicator LED of various devices when no sound was played. The frequency of the LED (100 Hz) can be seen in the graph for each device.

optical signal is obtained via an electro-optical sensor directed at a device's power indicator LED, there is a side effect in which the light frequency and its harmonics (200 Hz, 300 Hz, etc.) are added to the raw optical signal. The optical phenomenon that happens at 100 Hz (which was captured by the electro-optical sensor) is the result of power net harmonics. Most electronic devices work with DC voltage that is converted from AC. A diode bridge is integrated into the electrical device, which flips the negative half of the sinus, doubling the base frequency from 50 Hz to 100 Hz. As a result, the LED changes its intensity 100 times a second. These frequencies impact the optical signal and are not the result of the sound we wish to recover.

Conclusions: The light frequency and its harmonics, which are added to the optical signal and are not the result of the sound played, need to be filtered in order to recover the played signal.

4.2.2 Power Indicator LED's Response to Sound at 0-4 KHz In the next experiments, we tested the response of the power indicator LED of various devices to a wide range of frequencies.

Experimental Setup: We conducted the following experiments: In the first experiment, we obtained optical measurements from the power indicator LED of two speakers (Logitech S120 speakers and Winner speakers) that were connected to the electricity. In the second experiment, we obtained optical measurements from the power indicator LED of devices (TP-Link UE330 USB hub splitter [9], MIRACASE MHUB500 USB hub splitter [5], RP) that were used to provide power to the speakers via their USB input ports. In each of the experiments, the audio was played via speakers at a sound level of 70 dB. The audio played from the speakers is a 30 second audio file that consists of a chirp function (a frequency scan between 0-4 KHz).

Results: Fig. 6 presents the spectrograms obtained from the optical measurements. Three observations can be made from the spectrograms: (1) For some devices, the signal that appears in the optical

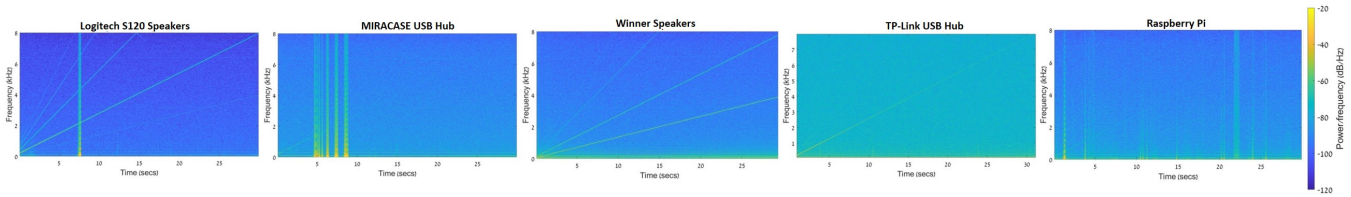

Figure 6: Spectrograms extracted from optical measurements obtained from the power indicator LED of various devices when a chirp function was played (frequency scan between 0-4 KHz). Note that in some devices (e.g., Logitech S120 speakers) the original frequency played by the speakers appears in the spectrum of the optical signal, while in other cases (e.g., TP-Link UE330, MIRACASE MHUB500), only the first harmonic of the frequency appears in the spectrum of the optical signal.

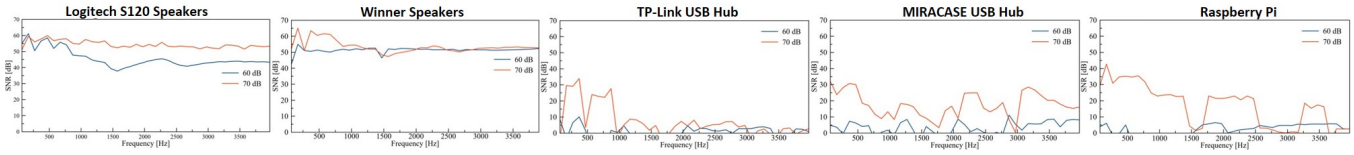

Figure 7: SNR of various devices in the spectrum of 0-4 KHz at three sound levels (60, 70 dB).

Table 1: Summary of Recovered Optical SNR of Various Devices at Sound Level of 70 dB

Device				Recovered Signal									
				0-4000 Hz		0-1000 Hz		1000-2000 Hz		2000-3000 Hz		3000-4000 Hz	
Manufacturer	Model	Type	Tone	Avg. SNR [dB]	STD	Avg. SNR [dB]	STD	Avg. SNR [dB]	STD	Avg. SNR [dB]	STD	Avg. SNR [dB]	STD
Logitech	S120	Speakers	Original	40.75	7.52	45.09	4.46	41.50	3.59	41.65	5.88	34.42	10.73
Winner			Original	58.04	5.67	56.83	7.37	61.34	2.69	58.44	3.30	55.44	6.60
CREATIVE	Pebble Modern 2.0		Original	6.95	9.67	17.46	13.35	4.17	2.39	0.2	0.5	6	5.69
TP-Link	UE330	USB Hub	First harmonic	20.65	14.53	36.85	8.02	21.35	12.37	16.12	10.98	7.61	8.73
MIRACASE	MHUB500		First harmonic	10.72	14.79	31.71	11.41	5.33	9.62	2.74	5.66	2.72	5.55
Raspberry Pi	4	Microcontroller	Original	15.73	9.59	26.85	6.23	18.21	4.79	11.46	6.16	5.93	5.70
Google	Google Nest	Smart Assistant	Original	1.53	4.15	3.81	6.23	2.3	4.68	0	0	0	0

measurements is much stronger (e.g., Logitech S120 speakers) than that of other devices (e.g., the RP). (2) For some devices, the signal obtained matches the original chirp function (e.g., Winner speakers). (3) For some devices, only the first harmonic of the chirp appears in the spectrogram (e.g., TP-Link USB hub splitter).

Conclusions: Based on these experiments, we concluded that: (1) For devices with a weak recovered optical signal, the application of denoising techniques is required to optimize the SNR. (2) For devices where the recovered optical signal appears in the first harmonic, the use of downtuning is required.

Next, we conducted an experiment to calculate the SNR of each of the seven devices (Logitech S120 speakers, Winner speakers, TP-Link UE330 USB hub splitter, MIRACASE MHUB500 USB hub splitter, RP, Google Nest Mini, and Creative Pebble speakers) across the 0-4 kHz spectrum at two levels (60 and 70 dB).

Experimental Setup: We used the same experimental setup as the previous experiment, however this time we played a different audio file which consists of various sine waves (120, 170,....1020 Hz), where each sine wave was played separately for two seconds. We played the audio file via the the speakers at two sound levels (60 and 70 dB) and obtained optical measurements.

Results: The SNR is presented in Fig. 7 and Table 1. The following observations can be made based on the results: (1) The SNR changes depending on the type of device used. This is the result of the differences in their power consumption and the intensity of the light emitted from their power indicator LED. (2) For some devices, the SNR has a low standard deviation (STD) throughout the spectrum examined (e.g., the STD of the SNR of the optical signal obtained from the Logitech S120 speakers is 7.5, and the STD of the SNR of the optical signal obtained from the Winner speakers is 5.6), which indicates a stable response, while for other devices, the SNR has a large STD (e.g., RP), which is usually the result of a decrease in the SNR as a function of the frequency. (3) For some devices, the effective spectrum that can be used to recover sound is narrow. For example, the SNR obtained from the power indicator LED of the MIRACASE MHUB500E is only stable up to 1000 Hz; for this device, the SNR of the spectrum beyond 1000 Hz is extremely unstable. (4) In general, the SNR of the recovered signal improves as the sound level increases. This phenomenon can be explained as follows: When the volume of the sound played by the speakers increases, the power consumption increases. The power is the product of voltage and current. The current consumed from AC-DC converter output stage capacitors (which have a limited amount of energy) increases, and as a result, the voltage level decreases proportionally to the current and volume levels. Since a device's power indicator LED is connected in parallel to the capacitor, it is linearly affected by voltage levels; its intensity also increases, and a greater amount of light is emitted. As a result, more photons are captured by the electro-optical sensor, which yields a better SNR. (5) The improvement in the SNR that results from higher volume levels varies depending on the device;

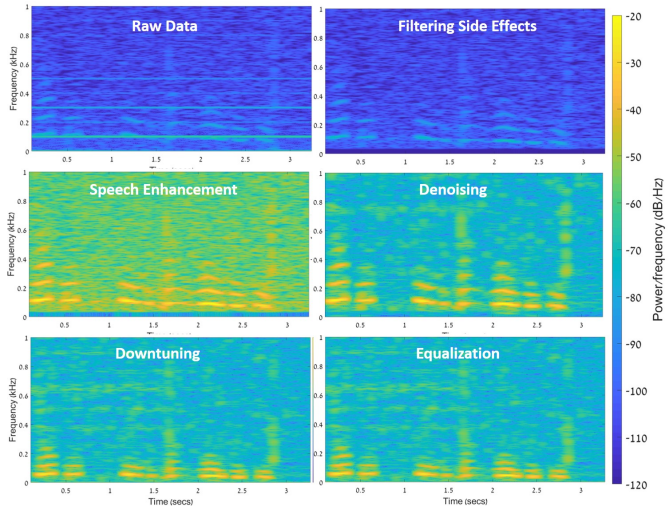

Figure 8: The influence of the five stages of optical-audio transformation (OAT) on the recovered signal.

in some cases, the improvement is significant (e.g., the TP-Link USB hub splitter and RP), while in other cases (e.g., Winner speakers), the improvement is less dramatic. (6) For some vulnerable devices (e.g., Google Nest Mini and Creative Pebble speakers), the SNR is poor due to the weak intensity of their power indicator LED. This fact requires more sensitive electro-optical sensor (with lower noise level) to recover sound from their power indicator LED.

Conclusions: Based on these experiments, we concluded that (1) a sound level of 70 dB (the sound level of virtual meetings) produces a high SNR, and (2) for devices in which the SNR decreases as a function of the frequency, an equalizer needs to be used to amplify the energy of weak frequency ranges.

5 Optical-Acoustic Transformation

In this section, we leverage the findings presented in Section 4 and present optical-acoustic transformation (OAT), which we used to recover audio signals from the optical signals obtained from an electro-optical sensor directed at a device's power indicator LED. Throughout this section, we consider $snd(t)$ as the audio played inside the target's room by the speakers, $opt(t)$ as the optical signal obtained via an electro-optical sensor directed at the power indicator LED of a device, and $snd^*(t)$ as the audio signal recovered from $opt(t)$ using OAT. OAT consists of the following stages:

Filtering Side Effects. As discussed in Section 4 and presented in Fig. 5, there are factors which affect the optical signal $opt(t)$ that are not the result of the sound played $snd(t)$ (e.g., peaks which are added to the spectrum that are the result of the light frequency of the power indicator LED and its harmonics - 100 Hz, 200 Hz, etc.). We filter the light frequency and its harmonics (its first, second, third, ... order harmonics) from $opt(t)$ using notch/bandstop filters.

Downtuning. As discussed in Section 4 and presented in Fig. 6, in some cases only the second order of the frequencies of the played signal $snd(t)$ appears in the optical signal $opt(t)$. As a result, the recovered signal $snd^*(t)$ is uptuned compared to the original signal $snd(t)$ played by the speakers. This case requires the eavesdropper to apply downtuning to the optical signal $opt(t)$ in order to recover sound at the original pitch. Downtuning is a standard procedure in the area of sound processing used to play a song at a lower tone. We implemented this procedure digitally according to [27].

Speech Enhancement. Speech enhancement is performed to maximize the signal's dynamic range before applying additional filters. To do so, we normalize the signal by scaling the values of $opt(t)$ to the range of [-1,1]

Denoising. This is the process of removing noise from a signal to increase the SNR and optimize its quality. Various techniques have been demonstrated to reduce noise, however we reduce noise by applying spectral subtraction, an adaptive technique proposed for denoising single channel speech [59].

Equalizer. As discussed in Section 4 and presented in Fig. 7, the SNR obtained from some devices is unstable and decreases as a function of the frequency. We use an equalizer in order to amplify the response of weak frequencies by adjusting the balance between frequency components within an electronic signal.

The techniques that enable OAT to recover audio signals from the optical signals are extremely popular in the area of speech processing; we used them for the following reasons: (1) the techniques rely on a speech signal that is obtained from a single channel; if eavesdroppers have the capability of sampling a device's power indicator LED using multiple sensors, thereby obtaining several signals via multiple channels, other methods can also be applied to recover an optimized signal, (2) these techniques do not require any prior data collection to create a model; recent methods use learning-based models (e.g., neural networks) to optimize the speech quality in noisy channels, however such methods require a large amount of data for the training phase in order to create robust models, something eavesdroppers would likely prefer to avoid, and (3) the techniques can be applied in real-time applications, so the optical signal obtained can be converted to audio with minimal delay.

The influence of each step of the OAT on the recovered signal when the transformation is used to recover an arbitrary sentence is illustrated in Fig. 8. As can be seen, the raw optical signal is very noisy. However, the application of speech enhancement and denoising techniques significantly improves the SNR. The equalizer is only used for fine-tuning. In the Appendix, we present Algorithm 1, an implementation of OAT's stages to recover audio from optical measurements.

6 Evaluation

In this section, we evaluate the performance of the Glowworm attack in terms of its ability to recover speech from the power indicator LED of various devices. We start by comparing Glowworm's performance to the performance of the visual microphone and Lamphone in a lab setup. Then, we test the influence of distance and the sound volume on Glowworm's performance when recovering speech through an office's transparent glass window/door.

The reader can assess the quality of the recovered sound visually by analyzing the extracted spectrograms, qualitatively by listening to the recovered audio signal online,[1] [2] and quantitatively based on metrics used by the audio processing community to compare a

[1] https://youtu.be/Mi6T2K9zQgE
[2] https://youtu.be/eZD4SdeKe7E

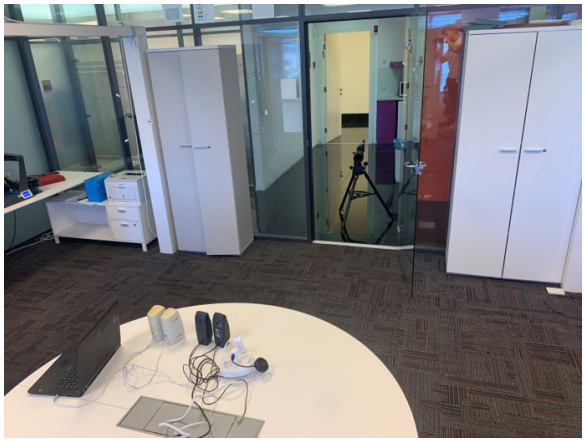

Figure 9: Experimental setup: the telescope and the four devices used in the experiments. A PDA100A2 electro-optical sensor is mounted on the telescope. The electro-optical sensor outputs voltage which is sampled via an ADC (NI-9234) and processed in LabVIEW.

recovered signal to its original signal: (1) Intelligibility - a measure of the comprehensibility of speech in given conditions [3]. To measure intelligibility, we used the metric suggested by [58] which results in values between [0,1]. A higher intelligibility indicates better sound quality. (2) Log-Likelihood Ratio (LLR) - a metric that captures how closely the spectral shape of a recovered signal matches that of the original clean signal [51]. A lower LLR indicates better sound quality. (3) NIST-SNR - the speech-to-noise ratio, which is defined as the logarithmic ratio between the estimated speech power and noise power over 20 consecutive milliseconds. A higher NIST-SNR indicates better sound quality.

We used the following equipment and configurations to recover sound in the experiments conducted and described in this section: a telescope (with a 20 cm lens diameter) was directed at the power indicator LED of the device. We mounted an electro-optical sensor (Thorlabs PDA100A2 [8]) to the telescope. The voltage was obtained from the electro-optical sensor using a 24-bit ADC NI-9234 card [7] and was processed in a LabVIEW script that we wrote. The sampling frequency of the ADC was configured at 2 KHz. In the remainder of this section we refer to this setup as the eavesdropping equipment. The level of the played sound was measured using a professional decibel meter.

6.1 Comparing Glowworm to the Visual Microphone and Lamphone

First, we compare the performance of Glowworm to that of the visual microphone [22] and Lamphone [48] using a similar experimental setup to the one used in the studies presenting those techniques. In those studies, the recovery of six sentences from the TIMIT repository [24] was demonstrated by playing the sentences via speakers and analyzing the recovered speech in a lab setup. We compare Glowworm's performance when recovering the same sentences from the power indicator LED of the four devices that with the highest SNR in Table 1: Two types of speakers (Logitech

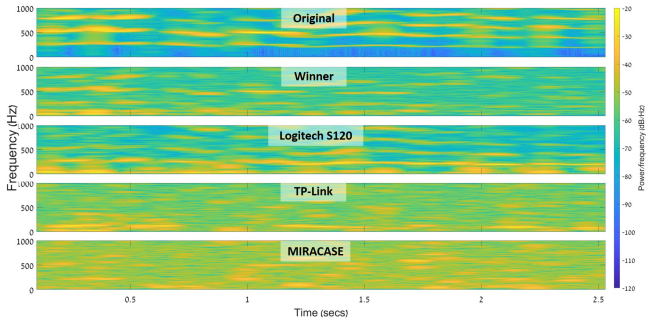

Figure 10: mabw0 sa1:"She had your dark suit in greasy wash water all year" recovered from various devices. The remaining spectrograms from the experiments listed in Table 2 can be seen in Figs. 16-20 in the Appendix.

S120 and Winner speakers) and two types of USB hub splitters (TP-Link UE330 and MIRACASE MHUB500).

Experimental Setup: We replicated the experimental setup used in both the visual microphone [22] and Lamphone [48] studies as follows: We placed the devices on a desktop inside a lab and played the same six sentences from the TIMIT repository [24] that were recovered by the visual microphone and Lamphone via the speakers, at the same volume level used in the visual microphone study (an average sound level of 95 dB). We note that the speakers we used this research are not capable of producing speech at sound levels higher than 85 dB, so we set the maximum sound level. In our experiment, the eavesdropping equipment was placed 2.5 meters from the devices, behind a closed door. Our experimental setup is presented in Fig. 9. In this experiment, the performance of the Glowworm attack was evaluated on the task of recovering speech by applying the attack in a direct manner, obtaining optical measurements from the power indicator LED of two speakers, and in indirect manner, obtaining optical measurements from the power indicator LED of two USB hub splitters.

Results: We recovered speech by applying OAT to the optical measurements. The recovered audio signals are available online[1] where they can be heard. The signals recovered by the power indicator LED of the two speakers and USB hub splitters when we played the sentence "She had your dark suit in greasy wash water all year" are presented in Fig. 10. The spectrograms recovered from the other five sentences can be seen in Figs. 16-20 in the Appendix. The intelligibility, LLR, and NIST-SNR of the recovered signals are reported in Table 2. Comparing these results to the results reported in the original Lamphone [48] and visual microphone [22] studies on the same sentences, we find that: (1) The average intelligibility of the speech recovered from the power indicator LED of the speakers (by applying the attack in a direct manner) is considered good/fair (according to [3]), however the average intelligibility of the speech recovered from the power indicator LED of the USB hub splitters (by applying the attack in an indirect manner) is considered poor. The visual microphone and Lamphone yield the same level of results in terms of intelligibility, as their average intelligibility is also considered good. (2) The average LLR of the speech recovered from the power indicator LED of Winner speakers is 1.74, which is lower (better) than Lamphone's average LLR (1.8) but higher (worse) than

Table 2: Performance of Glowworm on Speech Recovery from Various Devices

		Intelligibility				LLR				NIST-SNR			
		Speakers		USB Hub Splitters		Speakers		USB Hub Splitters		Speakers		USB Hub Splitters	
	Speech	Winner	Logitech S120	TP-Link	MIRACASE	Winner	Logitech S120	TP-Link	MIRACASE	Winner	Logitech S120	TP-Link	MIRACASE
Female speaker - fadg0, sa1	"She had your dark suit in greasy wash water all year"	0.618	0.426	0.378	0.374	1.765	2.238	2.023	2.758	3.3	12.8	5	5
Female speaker - fadg0, sa2	"Don't ask me to carry an oily rag like that"	0.623	0.542	0.341	0.333	1.787	2.39	2.585	2.322	9.5	5	11.5	5
Male speaker - mccs0, sa1	"She had your dark suit in greasy wash water all year"	0.666	0.542	0.366	0.350	2.126	2.134	2.154	2.323	15.5	8.8	10.5	5.5
Male speaker - mccs0, sa2	"Don't ask me to carry an oily rag like that"	0.709	0.539	0.428	0.434	1.663	2.508	2.719	2.581	4	15.8	12.3	3.8
Male speaker - mabw0, sa1	"She had your dark suit in greasy wash water all year"	0.574	0.45	0.368	0.318	1.576	2.029	2.24	2.009	9.8	8.8	8.8	6
Male speaker - mabw0, sa2	"Don't ask me to carry an oily rag like that"	0.697	0.56	0.368	0.347	1.658	2.176	1.774	2.237	13	20.3	9.3	4.8
	Average	0.647	0.509	0.374	0.359	1.763	2.246	2.249	2.372	9.183	11.917	9.567	5.017
	STD	0.051	0.056	0.028	0.041	0.276	0.175	0.317	0.263	4.825	5.539	2.592	0.738

the visual microphone's average LLR (1.53). The average LLR of the other devices was higher (worse) than that of the visual microphone and Lamphone. (3) The average NIST-SNR of the speech recovered from the power indicator LED of the Logitech S120 speakers is 11.9, which is higher (better) than Lamphone's average LLR (9.6) but lower (worse) than the visual microphone's average LLR (24.5). The average NIST-SNR of the other devices was lower (worse) than that of the visual microphone and Lamphone.

We conclude that the quality of the speech recovered by Glowworm is highly dependant on the device that is tested. We note that the Glowworm attack does not rely on the distance between the sound source and a lightweight vibrating object, whereas the results reported by Lamphone and the visual microphone are based on experiments performed when a vibrating object was placed a few centimeters from speakers. As a result, the quality of a signal recovered using the Glowworm attack at a fixed distance is stable and does not vary depending on the distance to nearby objects.

6.2 The Influence of Distance on Glowworm's Performance

Next, we evaluate the influence of distance on Glowworm's performance.

We evaluate Glowworm's performance on the task of recovering sound at the speech level of a typical virtual meeting: 70 dB. In the following set of experiments we attempted to recover sound from the power indicator LED of Winner speakers from various distances. We placed the speakers on a desktop inside an office; the eavesdropping equipment was located outside the office, behind two closed clear glass doors. The setup can be seen in Fig. 11.

First, we start by examining the influence of the sound level on the SNR.

Experimental Setup: We created an audio file that consists of various sine waves (120, 170, 220, ... 1970 Hz) and placed the eavesdropping equipment 15, 25, and 35 meters away from the speakers. We played the audio file via the speakers at 70 dB, obtaining the optical measurements as the sound was played. The electro-optical sensor was configured for the highest gain level before saturation.

Results: Fig. 12 presents the SNR for various distances. As can be seen from the results, the SNR looks very promising and stable through the entire spectrum measured. Unsurprisingly, the SNR

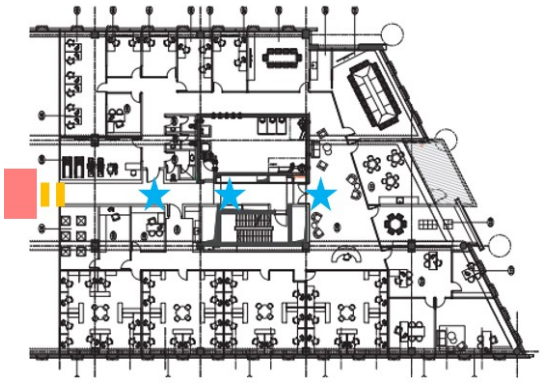

Figure 11: Experimental setup: The eavesdropping equipment, which was placed outside an office (in a location denoted by the red rectangle), was directed at the speakers which were placed in various locations (denoted with blue stars) at a distance of 15, 25, and 35 meters. Two closed glass doors separated the eavesdropping equipment and the speakers (denoted by yellow bars).

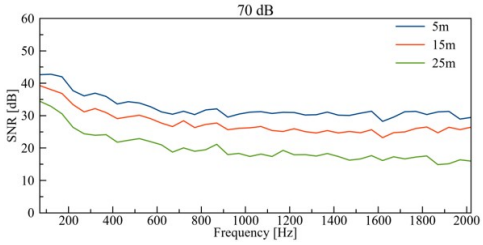

Figure 12: The SNR for various distances at 70 dB.

decreases as a function of the distance, since light deteriorates with distance.

Next, we evaluated Glowworm's performance in terms of its ability to recover speech audio from various distances. In order to do so, we recovered a well-known statement made by Donald Trump: "We will make America great again!"

Experimental Setup: We placed the eavesdropping equipment at three distances (15, 25, and 35 meters) from the Winner speakers' power indicator LED. We played the audio file via the speakers at 70 dB, obtaining the optical measurements as the sound was played.

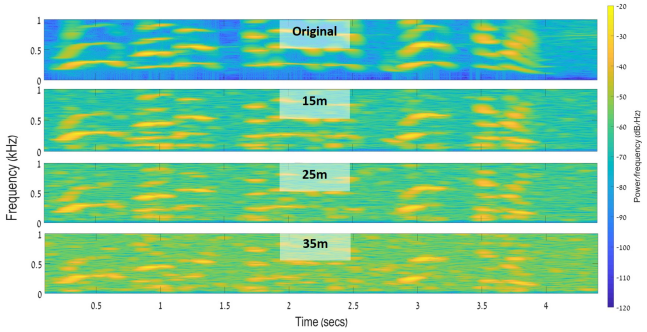

Figure 13: "We will make America great again!" recovered from various distances.

Table 3: "We Will Make America Great Again!" - Results of Recovered Speech from Various Distances

	Intelligibility	LLR	NIST SNR
15m	0.607	1.704	17.3
25m	0.552	3.24	14
35m	0.476	3.359	9.3

The electro-optical sensor was configured for the highest gain level before saturation.

Results: We recovered speech by applying OAT to the optical measurements. The recovered audio signals are available online[2] where they can be heard. The spectrogram of the recovered speech is presented in Fig. 13, and the intelligibility, LLR, and NIST-SNR of the recovered signals are reported in Table 3.

Conclusions: The results demonstrate that the intelligibility of the recovered signals is considered good up to a distance of 15 meters and fair up to a distance of 35 meters.

The results obtained show that Glowworm allows eavesdroppers to recover sound from a distance of 35 meters at a lower sound level than eavesdropping methods proposed in previous studies which require higher sound levels of 85-94 dB [39] and +95 dB [22, 63]. In addition, the results show that by analyzing optical measurements, eavesdroppers can double the range of the previous SOTA method used to recover sound from a device using EMR analysis [21].

7 Potential Improvements

In this section, we suggest methods that eavesdroppers can use to optimize the quality of the recovered audio or increase the range (i.e., distance between the eavesdropper and a device's power indicator LED), without changing the setup of the target location.

The potential improvements suggested below are presented based on the component they optimize.

Telescope. The amount of light that is captured by a telescope with a diameter of $2r$ is determined by the area of its lens (πr^2). As a result, using telescopes with a larger lens diameter enables the sensor to capture more light and optimizes the SNR of the recovered audio signal.

Electro-Optical Sensor. The sensitivity of the system can be enhanced by increasing the sensor's internal gain. Eavesdroppers can use a sensor that supports higher internal gain levels (note that the electro-optical sensor used in this study, PDA100A2 [8], outputs voltage in the range of [-10,10] and supports a maximum internal gain of 70 dB). Alternatively, the sensitivity of the system can be enhanced by using an electro-optical sensor with a lower noise level. Another option for maximizing the SNR is to profile the electro-optical sensor's self-noise (when the light is recorded) in order to filter its self noise.

Sound Recovery System. While many advanced denoising methods have been presented in the audio processing field, a large amount of data is often required to train a model that profiles the noise in order to optimize the output's quality. Such algorithms/-models can be used in place of the simple methods used in this research. In addition, various advanced dedicated algorithms for improving speech quality can also be used to extend the effective band of the recovered signal (e.g., artificial bandwidth extension algorithms [33, 34, 41, 49, 50]). In addition, more sensitive ADC (with lower sound level) can be used to sample the electro-optical sensor.

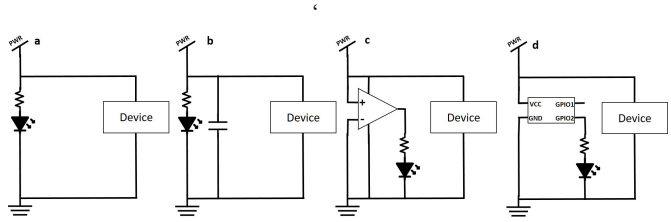

Figure 14: Circuits vulnerable to the Glowworm attack (a) and countermeasures using a capacitor (b), an additional OPAMP amplifier, (c) and the existing OPAMP (d).

8 Countermeasures

In this section, we describe several countermeasure methods that can be used to mitigate or prevent the Glowworm attack.

Manufacturer side. In most devices the power indicator LED is connected directly to the power line (see Fig. 14a). As a result, the device's power indicator LED is highly affected by the power consumption fluctuations that occur when speakers produce sound. To counter this phenomenon, a few approaches should be considered by hardware manufacturers: (1) Using a capacitor: A capacitor can be integrated in parallel to the power LED indicator; in this case, the capacitor behaves as a low-pass filter (see Fig. 14b). This is a straightforward and inexpensive solution for reducing AC fluctuations. However, in devices with high power consumption, the integrated capacitor must be large enough to supply a sufficient amount of current to the speakers. (2) Using an OPAMP: This can be implemented by integrating an additional OPAMP between the power line and the power indicator LED (see Fig. 14c) or by using an existing GPIO port of an integrated microcontroller as a power supply for the power indicator LED (see Fig. 14d). In both cases, this will eliminate power line AC fluctuations by a factor of the OPAMP amplifier's CMRR (common mode rejection ratio).

Consumer side. The attack can also be prevented by placing black tape over a device's power indicator LED. While this solution decreases a device's UX, it prevents the attackers from obtaining optical measurements from vulnerable devices.

9 Responsible Disclosure

We performed the following steps:

(1) We disclosed the details of the attack with the manufacturers of the devices that were analyzed in this research via their bug bounty programs and contact-us email addresses: Google, Logitech, Creative, TP-Link, Raspberry Pi, Winner, and MIRACASE. The email sent to each of the manufacturers contained explanations about the research, the Glowworm attack, proof that their devices are vulnerable to the Glowworm attack (electric and optical spectrograms of chirp functions), and recovered speech signals.
(2) We did not share the paper in order to keep the names of the other manufacturers confidential. In addition, we decided to refrain from informing manufacturers of devices that were not tested in this research about the Glowworm attack. We made this decision in order to prevent the information from spreading before giving the affected device manufacturers time to respond.
(3) We encouraged the manufacturers to meet with us in order to ensure that they understood the problem and assist them in developing a countermeasure.
(4) We explained to the manufacturers that we sent our findings to a conference and our paper may become public around November.
(5) We refrained from: (1) uploading the paper to arXiv, (2) discussing our findings with other researchers, and (3) sending the research to non-academic conferences.

Google, TP-Link, and Creative responded to our disclosure, asked us for more details, sent the findings of this research to their product team, and informed us that they would update us regarding their next steps. AS of this writing, Logitech, Raspberry Pi, Winner, and MIRACASE have not responded to our disclosure.

10 Discussion, Limitations & Future Work

The purpose of this research was to raise awareness regarding the feasibility of recovering sound by analyzing optical measurements obtained from an electro-optical sensor directed at a device's power indicator LED. While we are the first to demonstrate this method in the academic realm, we wonder whether our method is already known within the military and espionage realms. While we can only hypothesize about the answer to this question, for the following reasons we believe that we are not the first to exploit a device's power indicator LED to recover sound: (1) power indicator LEDs have been integrated into devices for many years, (2) power indicator LEDs' linear response to power consumption has been known for many years, (3) sound recovery is of interest to various entities around the world, and (4) virtual meeting platforms have been used for many years, given the fact that their protocols are encrypted. In addition, the case of the "Great Seal Bug" [19] proved that a new technology, the RFID, was used by agencies to eavesdrop three decades before it was scientifically discovered in 1973 [40].

We recommend that other hardware manufacturers empirically test whether their devices are vulnerable to the Glowworm attack. We hope that our findings will encourage hardware manufacturers to take our suggestions to empirically test their devices and redesign their electrical circuits (according to the suggestions provided in Section 8), in order to prevent eavesdroppers from applying the Glowworm attack in the future. However, we are not certain that they will implement our suggestions due to the financial implications of doing so, as some of the solutions may increase the manufacturer's overall cost, decreasing the revenue or requiring the manufacturer to increase the price of the product (which could make the device less attractive to consumers). While the cost of our countermeasures might seem negligible, given the likelihood that the devices are mass produced, the addition of a component to prevent the attack could cost a manufacturer millions of dollars. Given the cost-driven nature of consumers and the profit-driven nature of manufacturers, known vulnerabilities are often ignored as a means of reducing costs. This fact may leave many electrical circuits vulnerable to Glowworm attack for years to come.

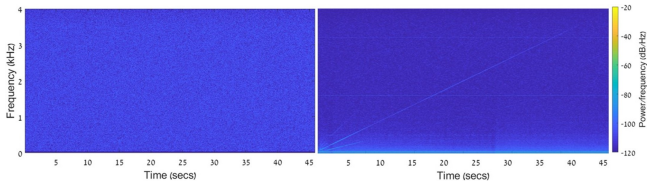

Figure 15: Two spectrograms extracted from Logitech speakers: Z200 (left) and Z120 (right). The results show that not all devices produced by the same manufacturer are vulnerable to optical TEMPEST attacks.

We also note that the area of optical sound eavesdropping has progressed significantly in the past seven years: a few studies have presented innovative methods to recover speech using data acquired from a high frequency video camera [22], LiDAR [55], and an electro-optical sensor [48]. Our attack continues the trend of recovering sound by exploiting optical side effects, and we believe that other studies will address this topic in the next few years.

The Glowworm attack suffers from one main disadvantage: The quality of the sound recovered is proportional to the quality of the equipment used by the eavesdropper. In our study, the cost of our equipment came to $1000 ($250 - telescope, $250 - electro-optical sensor, and $500 - ADC), an investment which allowed us to recover speech from a distance of 35 meters. In order to increase the attack range and recover higher quality sound, more expensive professional equipment is required (e.g., a more sensitive ADC and electro-optical sensor, a professional telescope). Such equipment would enable eavesdroppers to recover sound from vulnerable devices that have very weak LED intensity (e.g., Google Nest Mini, Creative Pebble speakers). In addition, some electrical circuits are not vulnerable to the Glowworm attack because they contain voltage stabilizers and filters that distort/change the known linear response of the intensity of the LED to power consumption. Interestingly, we found that while the power indicator LED of Logitech S120 speakers leaks information regarding the sound that is played from them, other speakers sold by the same manufacturer, Logitech Z200 speakers, do not leak such information, as can be seen in Fig. 15.

For future work, we suggest investigating the possibility of: (1) improving the Glowworm attack without the use of expensive equipment (e.g., improving the recovery model by using advanced

models such as artificial bandwidth extension [33, 34, 41, 49, 50]) and (2) recovering non-acoustic information from devices (e.g, optical cryptanalysis via a device's power indicator LED).

References

[1] [n.d.]. Creative Pebble Speakers. https://us.creative.com/p/speakers/creative-pebble

[2] [n.d.]. Facts About Speech Intelligibility. https://www.dpamicrophones.com/mic-university/facts-about-speech-intelligibility.

[3] [n.d.]. Intelligibility. https://en.wikipedia.org/wiki/Intelligibility_(communication).

[4] [n.d.]. Logitech S-120. https://www.amazon.com/Logitech-S120-2-0-Stereo-Speakers/dp/B000R9AAJA/

[5] [n.d.]. MIRACASE MHUB500 USB Hub Splitter. https://www.amazon.com/TP-Link-Portable-Ethernet-Notebooks-UE330/dp/B01N9M32TA/

[6] [n.d.]. MPU 6050 GY-521 3 Axis Gyro Accelerometer Sensor Module Arduino. http://xtcomp.co.za/image/catalog/GY-521.pdf

[7] [n.d.]. NI 9234 Datasheet. https://www.ni.com/pdf/manuals/374238a_02.pdf

[8] [n.d.]. PDA100A2. https://www.thorlabs.com/thorproduct.cfm?partnumber=PDA100A2

[9] [n.d.]. TP-Link UE330 USB. https://www.amazon.com/TP-Link-Portable-Ethernet-Notebooks-UE330/dp/B01N9M32TA/

[10] [n.d.]. Winner Desktop Speakers. http://www.tonewinner.com/english/

[11] [n.d.]. Winner Speakers. https://store.google.com/us/product/google_nest_mini?hl=en-US

[12] S. A. Anand and N. Saxena. [n.d.]. Speechless: Analyzing the Threat to Speech Privacy from Smartphone Motion Sensors. In *2018 IEEE Symposium on Security and Privacy (SP)*, Vol. 00. 116–133. https://doi.org/10.1109/SP.2018.00004

[13] Dmitri Asonov and Rakesh Agrawal. 2004. Keyboard acoustic emanations. In *IEEE Symposium on Security and Privacy, 2004. Proceedings. 2004.* IEEE, 3–11.

[14] Zhongjie Ba, Tianhang Zheng, Xinyu Zhang, Zhan Qin, Baochun Li, Xue Liu, and Kui Ren. 2020. Learning-based practical smartphone eavesdropping with built-in accelerometer. In *Proceedings of the Network and Distributed Systems Security (NDSS) Symposium*. 23–26.

[15] Michael Backes, Tongbo Chen, Markus Dürmuth, Hendrik PA Lensch, and Martin Welk. 2009. Tempest in a teapot: Compromising reflections revisited. In *2009 30th IEEE Symposium on Security and Privacy*. IEEE, 315–327.

[16] Michael Backes, Markus Dürmuth, Sebastian Gerling, Manfred Pinkal, and Caroline Sporleder. 2010. Acoustic Side-Channel Attacks on Printers.. In *USENIX Security symposium*, Vol. 10. 307–322.

[17] Michael Backes, Markus Dürmuth, and Dominique Unruh. 2008. Compromising reflections-or-how to read LCD monitors around the corner. In *2008 IEEE Symposium on Security and Privacy (sp 2008)*. IEEE, 158–169.

[18] Davide Balzarotti, Marco Cova, and Giovanni Vigna. 2008. Clearshot: Eavesdropping on keyboard input from video. In *2008 IEEE Symposium on Security and Privacy (sp 2008)*. IEEE, 170–183.

[19] Graham Brooker and Jairo Gomez. 2013. Lev Termen's Great Seal bug analyzed. *IEEE Aerospace and Electronic Systems Magazine* 28, 11 (2013), 4–11.

[20] Giovanni Camurati, Sebastian Poeplau, Marius Muench, Tom Hayes, and Aurélien Francillon. 2018. Screaming channels: When electromagnetic side channels meet radio transceivers. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*. 163–177.

[21] Jieun Choi, Hae-Yong Yang, and Dong-Ho Cho. 2020. TEMPEST Comeback: A Realistic Audio Eavesdropping Threat on Mixed-Signal SoCs. In *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security* (Virtual Event, USA) *(CCS '20)*. Association for Computing Machinery, New York, NY, USA, 1085–1101. https://doi.org/10.1145/3372297.3417241

[22] Abe Davis, Michael Rubinstein, Neal Wadhwa, Gautham J Mysore, Frédo Durand, and William T Freeman. 2014. The visual microphone: passive recovery of sound from video. (2014).

[23] Jeffrey Friedman. 1972. Tempest: A signal problem. *NSA Cryptologic Spectrum* 35 (1972), 76.

[24] John S Garofolo, Lori F Lamel, William M Fisher, Jonathan G Fiscus, and David S Pallett. 1993. DARPA TIMIT acoustic-phonetic continous speech corpus CD-ROM. NIST speech disc 1-1.1. *STIN* 93 (1993), 27403.

[25] Daniel Genkin, Adi Shamir, and Eran Tromer. 2014. RSA key extraction via low-bandwidth acoustic cryptanalysis. In *Annual Cryptology Conference*. Springer, 444–461.

[26] Dennis RE Gnad, Jonas Krautter, and Mehdi B Tahoori. 2019. Leaky noise: New side-channel attack vectors in mixed-signal IoT devices. *IACR Transactions on Cryptographic Hardware and Embedded Systems* (2019), 305–339.

[27] François Grondin, A Vakili, and L Demers. 2014. Guitar Pitch Shifter. http://www.guitarpitchshifter.com/about.html.

[28] Mordechai Guri, Yosef Solewicz, Andrey Daidakulov, and Yuval Elovici. 2017. SPEAKE(a)R: Turn Speakers to Microphones for Fun and Profit. In *11th USENIX Workshop on Offensive Technologies (WOOT 17)*. USENIX Association, Vancouver, BC. https://www.usenix.org/conference/woot17/workshop-program/presentation/guri

[29] Mordechai Guri, Boris Zadov, Dima Bykhovsky, and Yuval Elovici. 2019. Ctrl-alt-led: Leaking data from air-gapped computers via keyboard leds. In *2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)*, Vol. 1. IEEE, 801–810.

[30] Mordechai Guri, Boris Zadov, Andrey Daidakulov, and Yuval Elovici. 2018. xled: Covert data exfiltration from air-gapped networks via switch and router leds. In *2018 16th Annual Conference on Privacy, Security and Trust (PST)*. IEEE, 1–12.

[31] Mordechai Guri, Boris Zadov, and Yuval Elovici. 2017. LED-it-GO: Leaking (a lot of) Data from Air-Gapped Computers via the (small) Hard Drive LED. In *International conference on detection of intrusions and malware, and vulnerability assessment*. Springer, 161–184.

[32] Avesta Hojjati, Anku Adhikari, Katarina Struckmann, Edward Chou, Thi Ngoc Tho Nguyen, Kushagra Madan, Marianne S Winslett, Carl A Gunter, and William P King. 2016. Leave your phone at the door: Side channels that reveal factory floor secrets. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. 883–894.

[33] Vasu Iyengar, Rafi Rabipour, Paul Mermelstein, and Brian R Shelton. 1995. Speech bandwidth extension method and apparatus. US Patent 5,455,888.

[34] Peter Jax and Peter Vary. 2003. On artificial bandwidth extension of telephone speech. *Signal Processing* 83, 8 (2003), 1707–1719.

[35] Sean King. 2008. Luminous Intensity of an LED as a Function of Input Power. *ISB J. Phys* 2, 2 (2008), Paper–number.

[36] Paul Kocher, Joshua Jaffe, and Benjamin Jun. 1999. Differential power analysis. In *Annual international cryptology conference*. Springer, 388–397.

[37] Paul Kocher, Joshua Jaffe, Benjamin Jun, and Pankaj Rohatgi. 2011. Introduction to differential power analysis. *Journal of Cryptographic Engineering* 1, 1 (2011), 5–27.

[38] Markus G Kuhn. 2002. Optical time-domain eavesdropping risks of CRT displays. In *Proceedings 2002 IEEE Symposium on Security and Privacy*. IEEE, 3–18.

[39] A. Kwong, W. Xu, and K. Fu. 2019. Hard Drive of Hearing: Disks that Eavesdrop with a Synthesized Microphone. In *2019 2019 IEEE Symposium on Security and Privacy (SP)*. IEEE Computer Society, Los Alamitos, CA, USA. https://doi.org/10.1109/SP.2019.00008

[40] Jeremy Landt. 2005. The history of RFID. *IEEE potentials* 24, 4 (2005), 8–11.

[41] Sen Li, Stéphane Villette, Pravin Ramadas, and Daniel J Sinder. 2018. Speech bandwidth extension using generative adversarial networks. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 5029–5033.

[42] Xiangyu Liu, Zhe Zhou, Wenrui Diao, Zhou Li, and Kehuan Zhang. 2015. When good becomes evil: Keystroke inference with smartwatch. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*. 1273–1285.

[43] Joe Loughry and David A Umphress. 2002. Information leakage from optical emanations. *ACM Transactions on Information and System Security (TISSEC)* 5, 3 (2002), 262–289.

[44] Stefan Mangard, Elisabeth Oswald, and Thomas Popp. 2008. *Power analysis attacks: Revealing the secrets of smart cards*. Vol. 31. Springer Science & Business Media.

[45] Yan Michalevsky, Dan Boneh, and Gabi Nakibly. 2014. Gyrophone: Recognizing Speech from Gyroscope Signals. In *23rd USENIX Security Symposium (USENIX Security 14)*. USENIX Association, San Diego, CA, 1053–1067. https://www.usenix.org/conference/usenixsecurity14/technical-sessions/presentation/michalevsky

[46] Keaton Mowery, Sarah Meiklejohn, and Stefan Savage. 2011. Heat of the moment: Characterizing the efficacy of thermal camera-based attacks. In *Proceedings of the 5th USENIX conference on Offensive technologies*. 6–6.

[47] Ralph P Muscatell. 1983. Laser microphone. US Patent 4,412,105.

[48] Ben Nassi, Yaron Pirutin, Adi Shamir, Yuval Elovici, and Boris Zadov. [n.d.]. *Lamphone: Real-Time Passive Sound Recovery from Light Bulb Vibrations*. Technical Report. Cryptology ePrint Archive.

[49] Hannu Pulakka and Paavo Alku. 2011. Bandwidth extension of telephone speech using a neural network and a filter bank implementation for highband mel spectrum. *IEEE Transactions on Audio, Speech, and Language Processing* 19, 7 (2011), 2170–2183.

[50] Hannu Pulakka, Ulpu Remes, Santeri Yrttiaho, Kalle Palomaki, Mikko Kurimo, and Paavo Alku. 2012. Bandwidth extension of telephone speech to low frequencies using sinusoidal synthesis and a Gaussian mixture model. *IEEE transactions on audio, speech, and language processing* 20, 8 (2012), 2219–2231.

[51] Schuyler R Quackenbush, Thomas Pinkney Barnwell, and Mark A Clements. 1988. *Objective measures of speech quality*. Prentice Hall.

[52] Rahul Raguram, Andrew M White, Dibyendusekhar Goswami, Fabian Monrose, and Jan-Michael Frahm. 2011. iSpy: automatic reconstruction of typed input from compromising reflections. In *Proceedings of the 18th ACM conference on Computer and communications security*. 527–536.

[53] Rahul Raguram, Andrew M White, Yi Xu, Jan-Michael Frahm, Pierre Georgel, and Fabian Monrose. 2013. On the privacy risks of virtual keyboards: automatic reconstruction of typed input from compromising reflections. *IEEE Transactions on Dependable and Secure Computing* 10, 3 (2013), 154–167.

[54] Nirupam Roy and Romit Roy Choudhury. 2016. Listening Through a Vibration Motor. In *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services* (Singapore, Singapore) *(MobiSys '16)*. ACM, New York, NY, USA, 57–69. https://doi.org/10.1145/2906388.2906415

[55] Sriram Sami, Yimin Dai, Sean Rui Xiang Tan, Nirupam Roy, and Jun Han. 2020. Spying with Your Robot Vacuum Cleaner: Eavesdropping via Lidar Sensors. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems* (Virtual Event, Japan) *(SenSys '20)*. Association for Computing Machinery, New York, NY, USA, 354–367. https://doi.org/10.1145/3384419.3430781

[56] Diksha Shukla, Rajesh Kumar, Abdul Serwadda, and Vir V Phoha. 2014. Beware, your hands reveal your secrets!. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*. 904–917.

[57] Chen Song, Feng Lin, Zhongjie Ba, Kui Ren, Chi Zhou, and Wenyao Xu. 2016. My smartphone knows what you print: Exploring smartphone-based side-channel attacks against 3d printers. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. 895–907.

[58] Cees H Taal, Richard C Hendriks, Richard Heusdens, and Jesper Jensen. 2011. An algorithm for intelligibility prediction of time–frequency weighted noisy speech. *IEEE Transactions on Audio, Speech, and Language Processing* 19, 7, 2125–2136.

[59] Navneet Upadhyay and Abhijit Karmakar. 2015. Speech enhancement using spectral subtraction-type algorithms: A comparison and simulation study. *Procedia Computer Science* 54 (2015), 574–584.

[60] Wim Van Eck. 1985. Electromagnetic radiation from video display units: An eavesdropping risk? *Computers & Security* 4, 4 (1985), 269–286.

[61] G. Wang, Y. Zou, Z. Zhou, K. Wu, and L. M. Ni. 2016. We Can Hear You with Wi-Fi! *IEEE Transactions on Mobile Computing* 15, 11 (Nov 2016), 2907–2920. https://doi.org/10.1109/TMC.2016.2517630

[62] He Wang, Ted Tsung-Te Lai, and Romit Roy Choudhury. 2015. Mole: Motion leaks through smartwatch sensors. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*. 155–166.

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

[64] Yi Xu, Jared Heinly, Andrew M White, Fabian Monrose, and Jan-Michael Frahm. 2013. Seeing double: Reconstructing obscured typed input from repeated compromising reflections. In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*. 1063–1074.

[65] Qinggang Yue, Zhen Ling, Xinwen Fu, Benyuan Liu, Kui Ren, and Wei Zhao. 2014. Blind recognition of touched keys on mobile devices. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*. 1403–1414.

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

[67] Li Zhuang, Feng Zhou, and J Doug Tygar. 2009. Keyboard acoustic emanations revisited. *ACM Transactions on Information and System Security (TISSEC)* 13, 1 (2009), 1–26.

11 Sound Recovery Algorithm

The input to the algorithm is (1) $optical - stream$ - a pointer to the optical stream (the output of an ADC that samples the electro-optical sensor), (2) fs - the frequency that the ADC samples, and (3) a $equalizer - function$ - a function, which is used for balancing. The five stages of Algorithm 1 for recovering sound are described below.

Algorithm 1 Recovering Audio from an Optical Signal

```
INPUT: optical-stream, fs, equalizer-function
lightFs = 100
while (!isEmpty(optical-stream) do)
    /*Read from optical-stream to a buffer*/
    opt[] = read(optical-stream,fs)
    snd* = opt
    /*Filtering side effects*/
    for (i = lightFs; i < fs/2; i+=lightFs) do
        snd* = bandstop(i,snd*)
    /*Scaling to [-1,1]*/
    min = min(snd*), max = max(snd*)
    for (i = 0; i < len(snd*); i+=1) do
        snd*[i] = -1 + (snd*[i]-min)*2 / max-min
    /*Noise reduction*/
    snd* = spectral-subtraction(snd*)
    /*Balancing*/
    snd* = equalizer(snd*,equalizer-function)
    play (snd*)
```

12 Appendix - Spectrograms of Recovered Speech

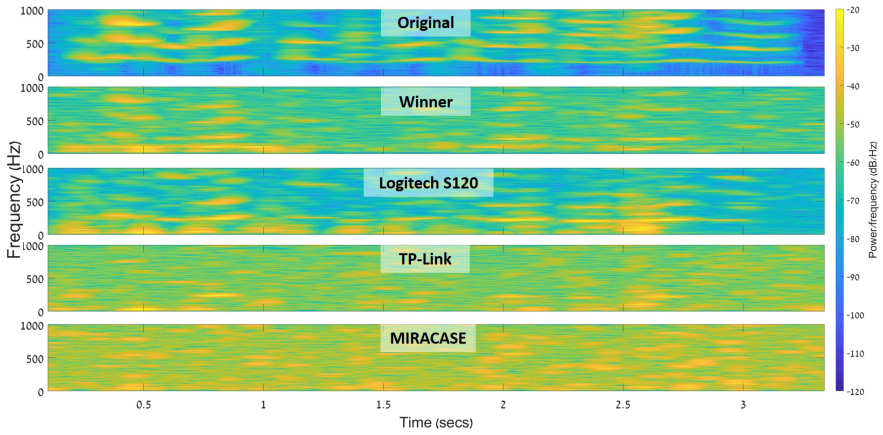

Figure 16: mabw0 sa2: "Don't ask me to carry an oily rag like that" recovered from various devices.

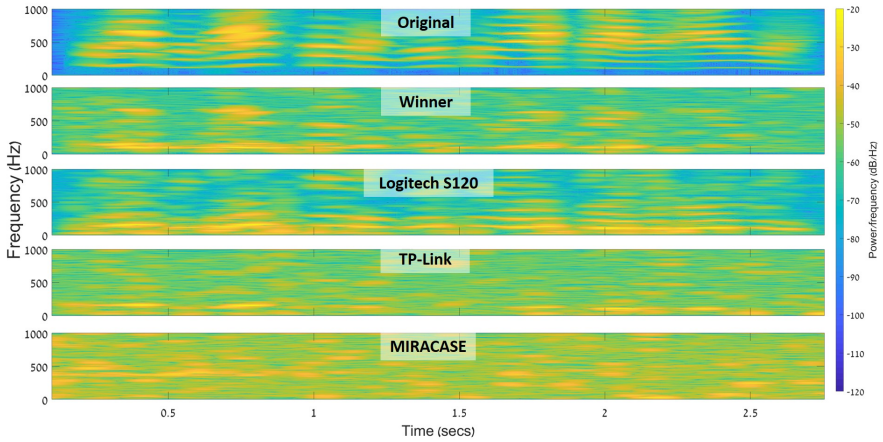

Figure 17: mccs0 sa2: "Don't ask me to carry an oily rag like that" recovered from various devices.

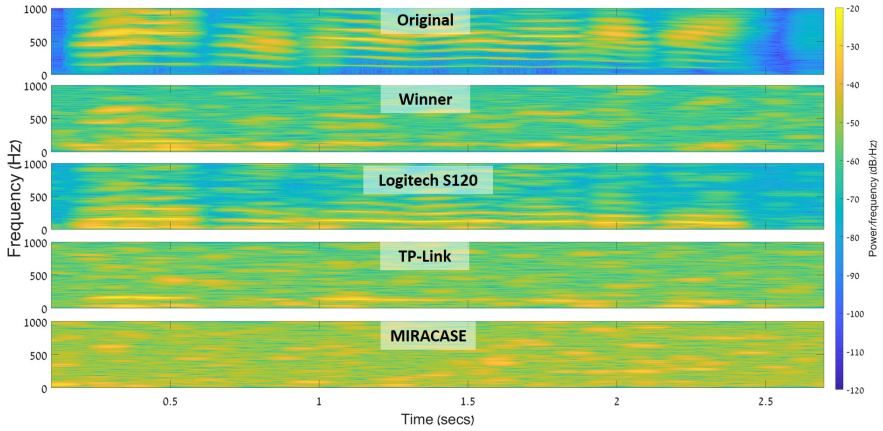

Figure 18: mccs0 sa1: "She had your dark suit in greasy wash water all year" recovered from various devices.

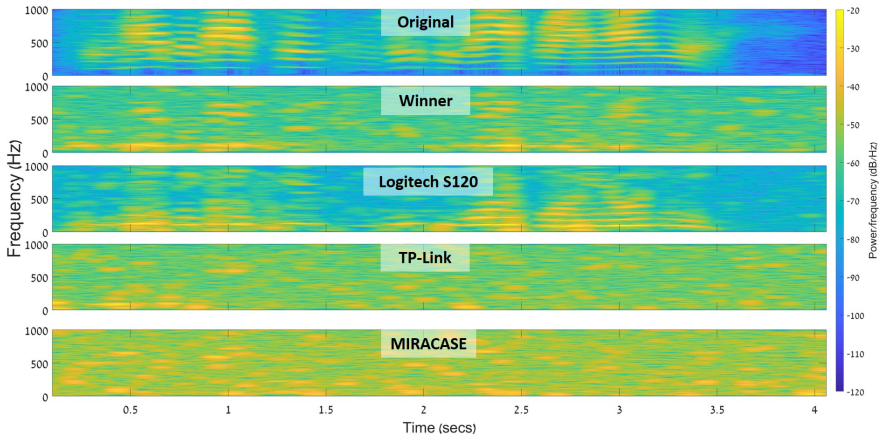

Figure 19: fadg0 sa2:"Don't ask me to carry an oily rag like that" recovered from various devices.

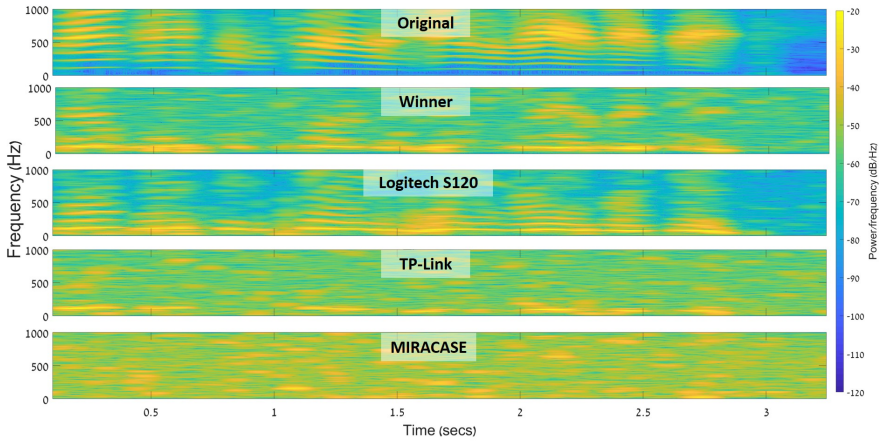

Figure 20: fadg0 sa1: "She had your dark suit in greasy wash water all year" recovered from various devices.