Lamphone: Real-Time Passive Sound Recovery from Light Bulb Vibrations

Ben Nassi,[1] Yaron Pirutin,[1] Adi Shamir,[2] Yuval Elovici,[1] Boris Zadov[1]

[1]Ben-Gurion University of the Negev, [2]Weizmann Institute of Science
{nassib, yaronpir, elovici, zadov}@bgu.ac.il, adi.shamir@weizmann.ac.il
Website - https://www.nassiben.com/lamphone
Video - https://youtu.be/Ob_C3J8zNPA

Abstract

Recent studies have suggested various side-channel attacks for eavesdropping sound by analyzing the side effects of sound waves on nearby objects (e.g., a bag of chips and window) and devices (e.g., motion sensors). These methods pose a great threat to privacy, however they are limited in one of the following ways: they (1) cannot be applied in real time (e.g., visual microphone), (2) are not external, requiring the eavesdropper to compromise a device with malware (e.g., Gyrophone), or (3) are not passive, requiring the eavesdropper to direct a laser beam at an object (e.g., laser microphone).

In this paper, we introduce "Lamphone," a novel side-channel attack for eavesdropping sound; this attack is performed by using a remote electro-optical sensor to analyze a light bulb's frequency response to sound. We show how fluctuations in the air pressure on the surface of a light bulb (in response to sound), which cause the bulb to vibrate very slightly (a millidegree vibration), can be exploited by eavesdroppers to recover speech and singing, passively, externally, and in real time. We analyze a light bulb's response to sound via an electro-optical sensor and learn how to isolate the audio signal from the optical signal. We develop an algorithm to recover sound from the optical measurements obtained from the vibrations of a light bulb and captured by the electro-optical sensor. Finally, we show that Lamphone is capable of recovering speech audio with good/fair intelligibility from 45 meters at a lower sound level than previous studies.

1 Introduction

Eavesdropping, the act of secretly or stealthily listening to a target/victim without his/her consent [1], by analyzing the side effects of sound waves on nearby objects (e.g., a bag of chips) and devices (e.g., motion sensors) is considered a great threat to privacy. In the past five years, various studies have demonstrated novel side-channel attacks that can be applied to eavesdrop via compromised devices placed in physical proximity of a target/victim [10, 11, 16, 17, 20, 23, 29, 36]. In these studies, data from devices that are not intended to serve as microphones (e.g., motion sensors [10, 11, 17, 23, 36], speakers [16], vibration devices [29], and magnetic hard disk drives [20]) are used by eavesdropper to recover sound. Sound eavesdropping based on the methods suggested in the abovementioned studies is very hard to detect, because applications/programs that implement such methods do not require any risky permissions (such as obtaining data from a video camera or microphone). As a result, such applications do not raise any suspicion from the user/operating system regarding their real use (i.e., eavesdropping). However, such methods require the eavesdropper to compromise a device located in proximity of a target/victim in order to: (1) obtain data that can be used to recover sound, and (2) exifltrate the raw/processed data.

To prevent eavesdroppers from implementing the abovementioned methods which rely on compromised devices, organizations deploy various mechanisms to secure their networks (e.g., air-gapping the networks, prohibiting the use of vulnerable devices, using firewalls and intrusion detection systems). As a result, eavesdroppers typically utilize three well-known methods that don't rely on a compromised device. The first method exploits radio signals sent from a victim's room to recover sound. This is done using a network interface card that captures Wi-Fi packets [32, 33] sent from a router placed in physical proximity of a target/victim. While routers exist in most organizations today, the primary disadvantages of these methods are that they cannot be used to recover speech [33] or they rely on a precollected dictionary to achieve their goal [32] (i.e., only words from the dictionary can be classified).

The second method, the laser microphone [24, 25], relies on a laser transceiver that is used to direct a laser beam into the victim's room through a window; the beam is reflected off of an object and returned to the laser transceiver which converts the beam to an audio signal. In contrast to [32, 33], laser microphones can be used in real time to recover speech, however the laser beam can be detected using a dedicated optical sensor. The third method, the visual microphone [13], exploits vibrations caused by sound from various materials (e.g., a bag of chips, glass of water, etc.) in order to recover speech, by using a video camera that supports a very high

frame per second (FPS) rate (over 2200 Hz). In contrast to the laser microphone, the visual microphone is totally passive, so its implementation is much more difficult for organizations/victims to detect. However, the main disadvantage of this method, according to the authors, is that the visual microphone cannot be applied in real time, because it takes a few hours to recover a few seconds of speech, since processing high resolution and high frequency (2200 frames per second) video requires significant computational resources.

In this paper, we introduce "Lamphone," a novel side-channel attack that can be applied by eavesdroppers to recover sound from a room that contains a floor/ceiling/desk light bulb. Lamphone recovers sound optically via an electro-optical sensor which is directed at a floor/ceiling/desk bulb; such bulbs vibrate due to air pressure fluctuations which occur naturally when sound waves hit the light bulb's surface. We explain how a bulb's response to sound (a millidegree vibration) can be exploited to recover sound, and we establish a criterion for the sensitivity specifications of a system capable of recovering sound from such small vibrations. Then, we evaluate a bulb's response to sound, identify factors that influence the recovered signal, and characterize the recovered signal's behavior. Based on our findings, we present an algorithm we developed in order to isolate the audio signal from the optical signal obtained by directing an electro-optical sensor at a light bulb. We evaluate Lamphone's performance on the task of recovering sound and show that when eavesdroppers have a clear line of sight to a target light bulb, that may contain transparent objects (e.g., a glass window/door) between the light bulb and the eavesdroppers, Lamphone is capable of recovering speech audio (1) at 80 dB (the sound level of a Zoom conversation) with excellent intelligibility from a distance of 25 meters and with good intelligibility from 45 meters, and (2) at 70 dB with fair intelligibility from a distance of 45 meters. In addition, we also evaluate Lamphone's performance on the task of recovering sound from a light bulb located in an an office building, that is covered in curtain walls. We show that eavesdroppers can exploit light emitted through curtain walls and recover sound from 25 meters with fair intelligibility.

The rest of the paper is structured as follows: In Section 2, we review existing methods for eavesdropping. In Section 3, we present the threat model. In Section 4, we analyze the response of a light bulb to sound. We present an algorithm for recovering sound in Section 5, and in Section 6, we evaluate Lamphone's performance on the task of recovering sound. In Section 7, we describe potential improvements that can be made to optimize the quality of the recovered sound, and we present countermeasure methods against the Lamphone attack in Section 8. We discuss the limitations of the attack and suggest future work directions in Section 9.

2 Background & Related Work

In this section, we explain how microphones work and describe two categories of eavesdropping methods (external and internal) and two sound recovery techniques. Then, we review and categorize related research focused on eavesdropping methods and discuss the significance of Lamphone with respect to those methods.

Table 1: Summary of Related Work (NM - not mentioned in the paper).

		Exploited Device	Sampling Rate	Sound Level	Technique
Internal	Motion Sensors	Gyroscope [23]	200 Hz	75 dB	Classification
		Accelerometer [10, 11, 36]	200 Hz	75 dB	
		Fusion of motion sensors [17]	2 KHz	85 dB	
	Misc.	Vibration motor [29]	16 KHz	80 dB	Recovery
		Speakers [16]	48 KHz	NM	
		Magnetic hard drive [20]	17 KHz	90 dB	
External	Radio Receiver	Network interface card [32]	300 Hz	NM	Classification Recovery
		Software-defined radio [33]	5 MHz	95 dB	
	Optical Sensor	High speed video camera [13]	2200 FPS	95 dB	Recovery
		Laser transceiver [24, 25]	40 KHz		

Microphones are devices that convert acoustic energy (sound waves) into electrical energy (the audio signal) [3]. Most microphones create electrical signals from sound waves using a three-step process involving the following components [5]. (1) Diaphragm: In the first step, sound waves (fluctuations in air pressure) are converted to mechanical motion by means of a diaphragm, a thin piece of material (e.g., plastic, aluminum), which vibrates when it is struck by sound waves. (2) Transducer: In the second step, when the diaphragm vibrates, the coil (attached to the diaphragm) moves in the magnetic field, producing a varying current in the coil through electromagnetic induction. (3) ADC (analog-to-digital converter): In the third step, the analog electric signal is sampled to a digital signal at standard audio sample rates (e.g., 44.1, 88.2, 96 kHz) using ADC.

There are two categories of eavesdropping methods which differ in terms of the location of the three components. *Internal methods* for eavesdropping are methods used to convert sound to electrical signals that rely on a single device. This device consists of the abovementioned components (i.e., the three components are co-located) and is placed near the source of the sound (the victim/target). Internal methods rely on a compromised device/sensor (e.g., a smartphone's gyroscope [23], magnetic hard drive [20], or speaker [16]) that is located in physical proximity to a victim/target and require the eavesdropper to exifltrate the output (electrical signal) from the device (e.g., via the Internet).

External methods are methods where the three components are not co-located. As with internal methods, the diaphragm is located in proximity of the source of the sound (the victim/target, however the diaphragm is based on objects (rather than devices), such as a glass window (in the case of the laser microphone), a bag of chips (in the visual microphone [13]), and a light bulb (in Lamphone). However, the other two components are part of another device (or devices) that can be located far from the victim/target, such as a laser transceiver (in the case of the laser microphone), a video camera (in the visual microphone), or an electro-optical sensor (in Lamphone).

There are two types of techniques used for eavesdropping: classification and audio/sound recovery.

Classification techniques can classify signals as isolated words. The signals obtained are uniquely correlated with sound, however they are not comprehensible (i.e., the signals cannot be recognized by the human ear) due to their poor quality (various factors can affect the quality, e.g., a low sampling rate). These methods require a dedicated classification model that relies on comparing a given signal to a dictionary compiled prior to eavesdropping (e.g., Gyrophone [23], AccelWord [36]). The biggest disadvantages of such methods are that words that do not exist in the dictionary cannot be classified and word separation techniques required to remove the silence.

Audio recovery consists of techniques in which the recovered signal can be played and recognized by the human ear (e.g., laser microphone, visual microphone [13], Hard Drive of Hearing [20], SPEAKE(a)R [16], etc.). They do not compare the obtained signal to a collection of signals gathered in advance or require a dedicated dictionary.

Several studies [10, 11, 17, 23, 36] have shown that measurements obtained from motion sensors that are located in proximity of a victim can be used for classification. They variously demonstrated that the response of MEMS gyroscopes [23], accelerometers [10, 11, 36], and geophones [17] to sound at 75-85 dB can be used to classify words and identify speakers and their genders, even when the sensors are located within a smartphone and the sampling rate is limited to 200 Hz. Two other studies [16, 29] showed that the process of output devices can be inverted to recover speech. In [29], the authors established a microphone by recovering sound at 80 dB from a vibration motor, and in [16], the audio from speakers was recovered. A recent study [20] exploited magnetic hard disks to recover audio, showing that measurements of the offset of the read/write head from the center of the track of the disk can be used to recover songs and speech at 90 dB.

The main disadvantages of the internal eavesdropping methods mentioned above ([10, 11, 16, 17, 20, 23, 29, 36]) are that (1) they require the eavesdropper to compromise a device located near the victim, and (2) security aware organizations implement security policies and mechanisms aimed at preventing the creation of microphones using such devices.

Two studies [32, 33] used the physical layer of Wi-Fi packets to eavesdrop sound at 95 dB. In [33], the authors suggested a method that analyzes the received signal strength (RSS) indication of Wi-Fi packets sent from a router to recover sound by using a device with an integrated network interface card. They showed that this method can be used to recover the sound from a piano located two meters away, however the authors did not demonstrate their method on the task of recovering speech. In [32], the authors suggested a method that analyzes the channel state information (CSI) of Wi-Fi packets sent from a router to classify words. The main disadvantage of this method is that it relies on a precollected dictionary.

The laser microphone [24, 25] is a well-known method that uses an external device. In this case, a laser beam is directed by the eavesdropper through a window into the victim's room; the laser beam is reflected off an object and returned to the eavesdropper who converts the beam to an audio signal. For decades, this method has been extremely popular in the area of espionage; its main disadvantage is that it can be detected using a dedicated optical sensor that analyzes the directed laser beams.

The most famous method related to our research is the visual microphone [13]. In this method, the eavesdropper analyzes the response of material inside the victim's room (e.g., a bag of chips, water, etc.) to sound waves at 95 dB, using video obtained from a high speed video camera (2200 FPS), and recovers speech. However, as was indicated by the authors, it takes a few hours to recover sound from a few seconds of video, because thousands of frames must be processed. In addition, this method relies on a high speed camera (at least 2200 FPS), which is an expensive piece of equipment.

Two studies were able to recover speech from encrypted VoIP by exploiting side effects of the compression's process (variable bitrate) [34, 35]. However, this paper focuses on sound eavesdropping techniques than turn a physical object into a diaphragm. Works that exploit a vulnerability in a digital protocol are not in the scope of the paper. Table 1 presents a summary of related work in the area of sound eavesdropping.

3 Threat Model

In this section, we describe the threat model and compare it to methods presented in other studies. We assume a victim located inside a room/office that contains a hanging/desk/floor light bulb. We consider an eavesdropper that is a malicious entity interested in spying on the victim in order to capture the victim's conversations and make use of the information provided in the conversation (e.g., perform extortion based on private information revealed by the victim). In order to recover the sound in this scenario, the eavesdropper performs the Lamphone attack.

Lamphone consists of the following primary components: (1) Telescope - This piece of equipment is used to focus the field of view on the light bulb from a distance. (2) Electro-optical sensor - This sensor is mounted on the telescope and consists of a photodiode (a semiconductor device) that converts light into an electrical current. The current is generated when photons are absorbed in the photodiode. Photodiodes are used in many consumer electronic devices (e.g., smoke detectors, medical devices). (3) 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 dedicated hardware (e.g., using capacitors, resistors, etc.). Alternatively, the eavesdropper can use an ADC to sample the electro-optical sensor and process the

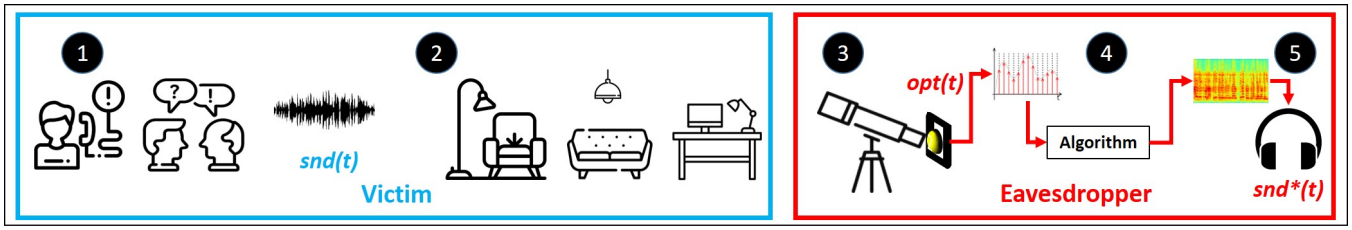

Figure 1: Lamphone's threat model: The sound $snd(t)$ from the victim's room (1) creates fluctuations on the surface of the desk/floor/hanging light bulb (the diaphragm) (2). The eavesdropper directs an electro-optical sensor (the transducer) at the light bulb via a telescope (3). The optical signal $opt(t)$ is sampled from the electro-optical sensor via an ADC (4) and processed, using an algorithm to recover the acoustic signal $snd^*(t)$ (5).

data using a sound recovery algorithm running on a laptop. In this study, we use the latter digital approach.

The conversation held in the victim's room creates sound $snd(t)$ that results in fluctuations in the air pressure on the surface of the light bulb. These fluctuations cause the bulb to vibrate, resulting in a pattern of displacement over time that the eavesdropper measures with an optical sensor that is directed at the bulb 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 audio recovery algorithm, to an acoustic signal $snd^*(t)$. Fig. 1 outlines the threat model.

As discussed in Section 2, microphones rely on three components (a diaphragm, transducer, and ADC). In Lamphone, the light bulb is used as a diaphragm which captures the sound. The transducer, in which the vibrations are converted to electricity, consists of the light that is emitted from the bulb (located in the victim's room) and the electro-optical sensor that creates the associated electricity (located outside the room at the eavesdropper's location). An ADC is used to convert the electrical signal to a digital signal in a standard microphone and in Lamphone. As a result, the Lamphone method is entirely passive and external.

The significance of Lamphone's threat model with respect to related work is that Lamphone: (1) is an external method that relies on a line of sight between the optical sensor and the bulb (as opposed to other methods that require eavesdroppers to compromise a device located in physical proximity of the victim [10, 16, 17, 20, 23, 29, 32, 33, 36]), (2) relies on an electro-optical sensor that is passive (as opposed to the laser microphone [24, 25] which utilizes an active laser beam), (3) can be performed in real time (as opposed to the visual microphone [13]), (4) is a technique for sound recovery and not for classification, so it does not rely on a pretrained dictionary or additional techniques for word separation (as opposed to [10, 17, 23, 32, 36]).

In order to keep the digital processing as light as possible in terms of computation, we want to sample the electro-optical sensor with the ADC at the minimal sampling frequency that allows comprehensible audio recovery. Lamphone is aimed at recovering sound (e.g., speech, singing), and a sufficient sampling frequency is required. 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 to recover comprehensible sound (e.g., 2200 Hz for the visual microphone [13]). In this study, we sample the electro-optical sensor at a sampling rate of 2-4 kHz.

4 Bulbs as Microphones

In this section, we perform a series of experiments aimed at explaining why light bulb vibrations can be used to recover sound and evaluate a bulb's response to sound empirically.

4.1 The Physical Phenomenon

First, we measure the vibration of a light bulb when sound waves hit the light bulb's surface, and we establish a criterion for the sensitivity specifications of a system capable of recovering sound from these vibrations

4.1.1 Measuring a Light Bulb's Vibration

To measure the response of a light bulb to sound, we examine how sound produced in proximity to the light bulb affects a bulb's three-dimensional vibration (as presented in Fig. 2).

Experimental Setup: We attached a gyroscope (MPU-6050 GY-521 [6]) to the bottom of an E27 LED hanging light bulb (12 watts); the bulb was not illuminated during this experiment. A Raspberry Pi 3 was used to sample the gyroscope at 800 Hz. We placed Logitech Z533 speakers very close to the bulb (one centimeter away) and played various sine waves (100, 150, 200, 250, 300, 350, 400 Hz) from the speakers at three volume levels (60, 70, 80 dB). We obtained measurements from the gyroscope while the sine waves were played. We repeated this experiment again for an E14 LED light bulb.

Results: Based on the measurements obtained from the gyroscope, we calculated the average peak-to-peak difference (in degrees) for θ and ϕ (which are presented in Fig. 3). The

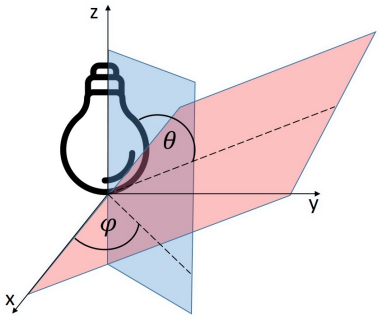

Figure 2: A 3D scheme of a light bulb's axes.

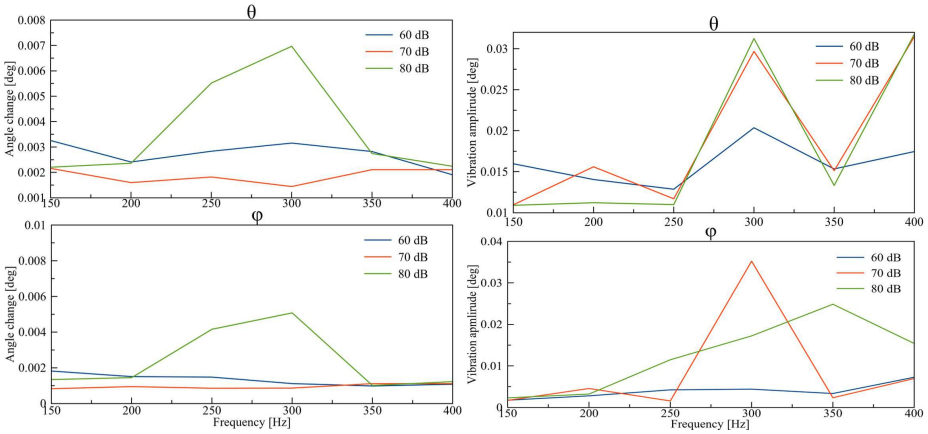

Figure 3: Peak-to-peak difference of angles ϕ and θ for E27 (left) and E14 (right) light bulbs at the 100-400 Hz spectrum.

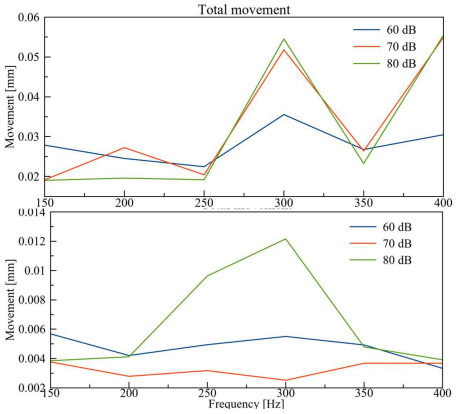

Figure 4: The peak-to-peak movement for E14 (top) and E27 (bottom) bulbs at the range of 100-400 Hz.

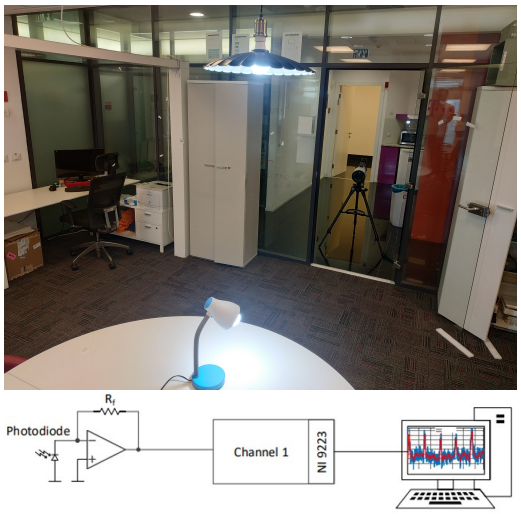

Figure 5: Experimental setup: the telescope and the two light bulbs that were used in the experiments. A PDA100A2 electro-optical sensor [8] is mounted on the telescope. The electro-optical sensor outputs voltage that is sampled via an ADC (NI-9234) [7] and processed in LabVIEW.

average peak-to-peak difference was computed by calculating the peak-to-peak difference between every 800 consecutive measurements (that were collected from one second of sampling) and averaging the results. The frequency response as a function of the average peak-to-peak difference is presented in Fig. 3. The results presented in Fig. 3 reveal three interesting insights: the average peak-to-peak difference for the angle of the bulb is: (1) very small (1-7 millidegrees for an E27 light bulb and 2-35 millidegrees for an E14 light bulb), (2) increases as the volume increases, and (3) changes as a function of the frequency.

Based on the known formula of the spherical coordinate system [9], we calculated the 3D vector (x,y,z) that represents the peak-to-peak vibration on each of the axes. We calculated the Euclidean distance between this vector and the vector of the initial position. Fig. 3 presents the results which show that sound caused caused a movement of 3.5-12 microns of the E27 light bulb and a vibration of 17-55 microns of the E14 light bulb.

4.1.2 Capturing the Optical Changes

We now explain how eavesdroppers can determine the sensitivity of the equipment (an electro-optical sensor, telescope, and ADC) needed to recover sound based on a bulb's vibration. The graphs presented in Fig. 3 establish a criterion for recovering sound: the eavesdropping system (consisting of an electro-optical sensor, telescope, and ADC) must be sensitive enough to capture the small optical differences that are the result of a hanging bulb's vibrations of 3.5-55 microns.

In order to demonstrate how eavesdroppers can determine the sensitivity of the equipment they will need to satisfy the abovementioned criterion, we conduct another experiment.

Experimental Setup: We directed a telescope (with a lens diameter of 25 cm) at a 1050 lumens E27 LED bulb (as can be seen in Fig. 5). We mounted an electro-optical sensor (the

Table 2: Expected Voltage for Each Frequency (based on linear equations calculated from Fig. 6 and expected movement from Fig. 4). Green cells can be detected by a 24 bit ADC and yellow cells can be detected by a 32 bit ADC.

		Expected voltage change for E14 light bulb						Expected voltage change for E27 light bulb					
Distance	Linear equation	150 Hz	200 Hz	250 Hz	300 Hz	350 Hz	400 Hz	150 Hz	200 Hz	250 Hz	300 Hz	350 Hz	400 Hz
1m - 2m	y = -0.59x + 2.56	10.2 µV	10.8 µV	10.2 µV	33 µV	13.2 µV	31.8 µV	2.22 µV	2.4 µV	5.4 µV	7.2 µV	2.52 µV	2.4 µV
2m - 3m	y = -0.52x + 2.41	8.9 µV	9.42 µV	8.9 µV	28.8 µV	11.5 µV	27.74 µV	1.94 µV	2.09 µV	4.71 µV	6.28 µV	2.2 µV	2.9 µV
3m - 4m	y = -0.14x + 1.27	2.47 µV	2.62 µV	2.47 µV	7.98 µV	3.2 µV	7.7 µV	0.54 µV	0.58 µV	1.31 µV	1.74 µV	0.61 µV	0.58 µV
4m - 6m	y = -0.136+ 1.24	1.27 µV	1.34 µV	1.27 µV	4.11 µV	1.64 µV	3.95 µV	0.51 µV	0.55 µV	1.23 µV	1.64 µV	0.57 µV	0.55 µV
6m - 7m	y = -0.12x + 1.14	2.1 µV	2.22 µV	2.1 µV	6.64 µV	2.71 µV	6.54 µV	0.43 µV	0.49 µV	1.11 µV	1.48 µV	0.51 µV	0.49 µV
7m - 9m	y = -0.1x + 1.02	1.7 µV	1.8 µV	1.7 µV	5.5 µV	2.2 µV	5.3 µV	0.37 µV	0.4 µV	0.9 µV	1.2 µV	0.42 µV	0.4 µV
9m - 10m	y = -0.005x + 0.16	0.09 µV	0.09 µV	0.09 µV	0.28 µV	0.11 µV	0.28 µV	19 nV	21 nV	47 nV	63 nV	22 nV	21 nV

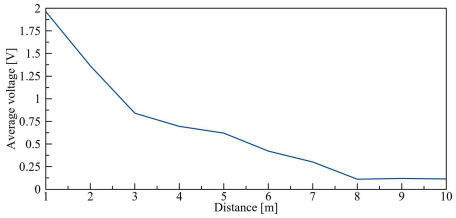

Figure 6: Output obtained from the electro-optical sensor from various ranges.

Thorlabs PDA100A2 [8], which is an amplified switchable gain light sensor that consists of a photodiode, used to convert light to electrical voltage) 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 internal gain of the electro-optical sensor was set at 50 dB. We placed the telescope at various distances (1, 2, 3, 4, 6, 7, 9, 10 meters) from the light bulb and measured the voltage that was obtained from the electro-optical sensor at each distance.

Results: The results of this experiment are presented in Fig. 6. These results were used to compute the linear equation between each two consecutive points. Based on the linear equations, we calculated the expected voltage for each expected movement in the 100-400 Hz spectrum for E27 and E14 light bulbs for a sound level of 80 dB (based on the results from Fig. 4). The linear equations and the expected voltage for each movement are presented in Table 2.

We now explain how to use the data in Table 2 in order to determine which frequencies can be recovered from the obtained optical measurements for a sound level of 80 dB. A 24-bit ADC with an input range of [-5,5] voltage (e.g., like the card used in our experiments) provides a sensitivity of:

$$\frac{10}{2^{24}-1} \approx 0.6\ \mu\text{V} \tag{1}$$

Analyzing Table 2, we find that a sensitivity of 0.6 µV (which is provided by a 24-bit ADC) is sufficient for recovering the entire spectrum (100-400 Hz) from a maximum range of nine meters for an E14 light bulb, because the smallest vibration of the bulb (17 microns) from this range is expected to yield a difference of 1.7 µV (for a frequency of 150 Hz and a range of nine meters). In the case of an E27 light bulb, the sensitivity provided by a 24-bit ADC is sufficient to recover the entire spectrum from a shorter range of up to three meters (because the E27 light bulb is heavier than the E24 light bulb, and its vibrations are smaller). The green cells in Table 2 indicate frequencies that can be recovered by a 24-bit ADC (i.e., their value is greater than 0.6 µV). As can be seen from the table, the setup we used is not sensitive enough to recover the entire measured spectrum of: (1) an E14 light bulb from a range beyond nine meters, and (2) an E27 light bulb from a range beyond three meters. In order to recover frequencies from a greater distance, an ADC that provides a higher sensitivity is required. A 32-bit ADC with an input range of [-5,5] voltage provides a sensitivity of:

$$\frac{10}{2^{32}-1} \approx 2.3\ \text{nV} \tag{2}$$

A sensitivity of 2.3 nV, which is provided by a 32-bit ADC, is sufficient for recovering the entire spectrum (100-400 Hz) in the ranges that were measured, because the smallest vibration of the bulb (3.5 microns) is expected to yield a difference of 19 nV (for a frequency of 150 Hz and a range of 10 meters). The yellow cells in Table 2 indicate the frequencies that can be recovered by a 32-bit ADC (i.e., their value is greater than 2.3 nV and lower than 0.6 µ).

In order to optimize the setup so it can be used to detect frequencies that cannot be recovered, eavesdroppers can: (1) increase the internal gain of the electro-optical sensor, (2) use a telescope with a lens capable of capturing more light (we demonstrate this later in the paper), or (3) use an ADC that provides greater resolution and sensitivity.

4.2 Exploring the Optical Response to Sound

The experiments presented in this section were performed to evaluate bulbs' response to sound. The experimental setup described in the previous subsection (presented in Fig. 5) was also used throughout these set of experiments.

4.2.1 Characterizing Optical Signal in Silence

First, we learn the characteristics of the optical signal when no sound is played.

Experimental setup: We obtained five seconds of optical measurements from the electro-optical sensor when no sound was played in the lab.

Results: The FFT graph extracted from the optical measurements when no sound was played is presented in Fig. 7. Each bulb works at a fixed light frequency (e.g., 100 Hz). Since $opt(t)$ is obtained via an electro-optical sensor directed at a bulb, the light frequency and its harmonics are added to

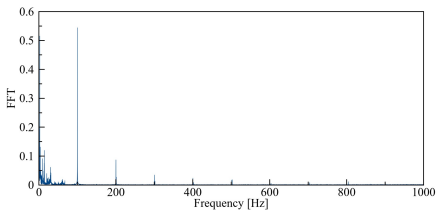

Figure 7: Baseline - FFT of the optical signal in silence (no sound is played).

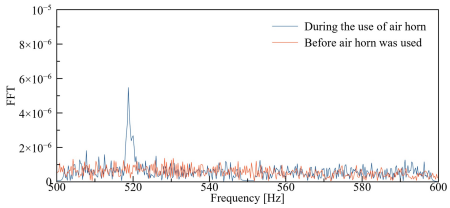

Figure 8: Difference in FFT before and when an air horn was used.

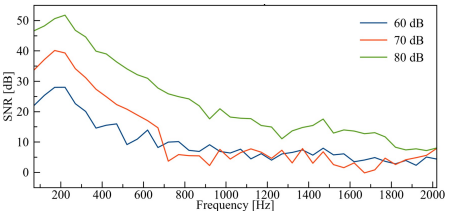

Figure 9: SNR for a desk lamp at 100-2000 Hz.

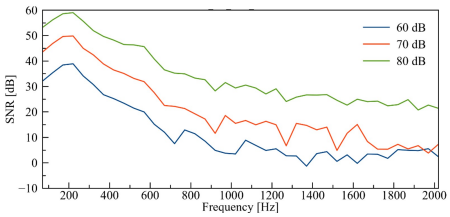

Figure 10: SNR for a hanging light bulb at 100-2000 Hz.

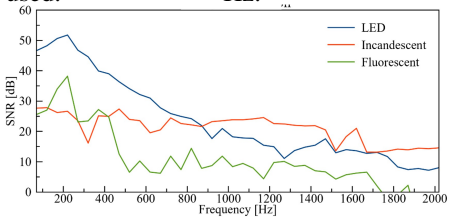

Figure 11: A comparison of the SNR obtained from various bulbs.

the raw signal $opt(t)$. These frequencies strongly impact the optical signal and are not the result of the sound that we want to recover. From this experiment we concluded that filtering will be required.

4.2.2 Bulb's Response to a Single Sine Wave

Next, we show that the effect of sound on a nearby bulb can be exploited to recover sound by analyzing the light emitted from the bulb via an electro-optical sensor in the frequency domain.

Experimental Setup: In this experiment, we used an air horn that plays a sine wave at a frequency of 518 Hz. We pointed the electro-optical sensor at the bulb and obtained optical measurements. Then we placed the air horn five centimeters away from the bulb and operated the horn, obtaining sensor measurements as we did so.

Results: Fig. 8 presents two FFT graphs created from two seconds of optical measurements obtained before and while the air horn was used. As can be seen from the results, the peak that was added to the frequency domain at around 518 Hz shows that the sound the air horn produced affects the optical measurements obtained via the electro-optical sensor. In this experiment, we specifically used a device (air horn) that does not create an electro-magnetic side effect (in addition to the sound), in order to prove that the results obtained are caused by fluctuations in the air pressure on the surface of the bulb (and not by anything else).

4.2.3 Bulb's Response to Sound at 100-2000 Hz

In the next experiment, we tested the response of a hanging light bulb and the light bulb in a desk lamp to a wide spectrum of frequencies. These experiments were conducted using speakers that were placed five centimeters away from the light bulb on a dedicated stand.

Experimental Setup: We created an audio file that consists of various sine waves (120, 170, 220, ... 1020 Hz) where each sine wave was played for two seconds. We played the audio file via the speakers near the bulb at three volume levels (60, 70, 80 dB) and obtained the optical signal via the electro-optical sensor.

Results: Figs. 9 and 10 present the SNR obtained from the desk lamp light bulb and the hanging light bulb. Analyzing the signal with respect to the original signal reveals two insights: (1) The response of the recovered signal decreases as the frequency increases until its power reaches same level as the noise. (2) The SNR improves as the volume increases. From this experiment we concluded that we would have to increase the SNR using speech enhancement and denoising techniques, and strengthen the response of higher frequencies in order to recover them using an equalizer.

4.2.4 Various Bulbs' Responses to Sound

Next, we compare the response of various bulbs to sound.

Experimental Setup: We repeated the previous experiment for three different types of 12 watt E27 bulbs: LED, florescent, and incandescent. In each experiment, a different bulb was used, along with the same audio file; we obtained the optical measurements via the electo-optical sensor, resulting in an optical signal for each of the bulbs.

Results: We calculate the SNR obtained from the three optical signals. Fig. 11 presents the results. As can be seen, sound can be recovered from the three bulbs that were tested. However, the SNR of the LED and incandescent bulbs is much higher than the SNR obtained from the fluorescent bulb.

5 Sound Recovery Model

In this section, we leverage the findings presented in Section 4 and present Algorithm 1 for recovering audio from measurements obtained from an electro-optical sensor directed at a light bulb. We assume that $snd(t)$ is the audio that is played inside the victim's room. The input to the algorithm is (1) $optical - stream$, a pointer to the optical stream

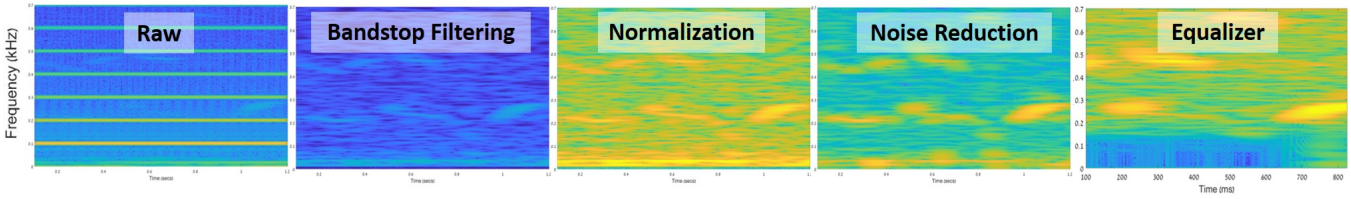

Figure 12: The effect of each stage of Algorithm 1 in recovering the word "lamb" from an optical signal.

(the output of an ADC that samples the electro-optical sensor), (2) fs the frequency that the ADC samples, and (3) a $equalizer-function$ that is used for balancing. The stages of Algorithm 1 for recovering sound are described below and presented in Fig. 12.

Algorithm 1 Recovering Audio from Optical Signal

```
1:  INPUT: optical-stream, fs, equalizer-function
2:  bulbFs = 100
3:  while (!isEmpty(optical-stream) do)
4:      /*Read from optical-stream to a buffer*/
5:      opt[] = read(optical-stream,fs)
6:      snd* = opt
7:      /*Filtering side effects*/
8:      for (i = bulbFs; i < fs/2; i+=bulbFs) do
9:          snd* = bandstop(i,snd*)
10:     /*Scaling to [-1,1]*/
11:     min = min(snd*), max = max(snd*)
12:     for (i = 0; i < len(snd*); i+=1) do
13:         snd*[i] = -1 + (snd*[i]-min)*2 / max-min
14:     /*Noise reduction*/
15:     snd* = spectral-subtraction(snd*)
16:     /*Balancing*/
17:     snd* = equalizer(snd*,equalizer-function)
18:     play (snd*)
```

1) Filtering Side Effects: As discussed in Section 4 and presented in Fig. 7, there are factors which affect the optical signal $opt(t)$ that are not the result of the sound played (e.g., peaks which are added to the spectrum that are the result of the lighting frequency of the light bulb and its harmonics - 100 Hz, 200 Hz, etc.). We filter these frequencies using bandstop filters (lines 7-8 in Algorithm 1). The effect of the filters applied to the optical signal is illustrated in Fig. 12.

2) Speech Enhancement: Speech enhancement (using audio signal processing techniques) is performed to optimize the speech quality by improving the intelligibility and overall perceptual quality of the speech signal. We enhance the speech by normalizing the values of $opt(t)$ to the range of [-1,1] (lines 10-12 in Algorithm 1). The impact of this stage is enhancement of the signal (as can be seen in Fig. 12).

3) Noise Reduction: Noise reduction is the process of removing noise from a signal in order to optimize its quality. We reduce the noise by applying spectral subtraction, one of the first techniques proposed for denoising single channel speech [31] (line 14 in Algorithm 1).

4) Equalizer: Equalization is the process of adjusting the balance between frequency components within an electronic signal. We use an equalizer in order to amplify the response of weak frequencies. The equalizer is provided as input to Algorithm 1 and applied in its last stage (line 16).

The techniques used in this study to recover speech 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 the light bulb using other 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; novel speech processing methods use neural networks to optimize the speech quality in noisy channels, however such neural networks require a large amount of data for the training phase in order to create robust models, a requirement that 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.

6 Evaluation

In this section, we evaluate the performance of the Lamphone attack in terms of its ability to recover sound from the light bulb of a desk lamp and a hanging light bulb. We start by comparing the performance of Lamphone to the visual microphone in a lab setup. We continue by testing the influence of distance and sound volume on the performance of Lamphone when there are no obstacles or only transparent obstacles exist between the light bulb and the telescope (e.g., transparent glass window/door). Finally, we evaluate the Lamphone's performance for recovering sound using light emitted through the curtain walls on an office building at our university.

The reader can assess the the quality of the recovered sound visually by analyzing the extracted spectrograms, qualitatively by listening to the recovered audio signal online,[1] and quantitatively based on metrics used by the audio processing community to compare a recovered signal to its original signal: (1) Intelligibility - a measure of how comprehensible speech is in given conditions. Intelligibility is affected by the level and quality of the speech signal, and the type and level of

[1] https://youtu.be/0eaFXXS7eU4

Table 3: Comparison between the results of Visual Microphone (VM) and Lamphone for Sound Recovery of Speech.

		Intelligibility			LLR			WSS			NIST STNR		
	Speech	VM	HL	DL	VM	HL	DL	VM	HL	DL	VM	HL	DL
Female speaker - fadg0, sa1	"She had your dark suit in greasy wash water all year"	0.72	0.74	0.72	1.47	1.93	1.79	120.29	88.2	75.55	26.8	16.8	16
Female speaker - fadg0, sa2	"Don't ask me to carry an oily rag like that"	0.65	0.69	0.67	1.37	2.44	2.1	197.83	68.82	71.76	43.3	3.8	4.5
Male speaker - mccs0, sa1	"She had your dark suit in greasy wash water all year"	0.59	0.75	0.7	1.31	2.03	1.72	149.55	72.81	63.1	27.3	14	10.3
Male speaker - mccs0, sa2	"Don't ask me to carry an oily rag like that"	0.67	0.76	0.71	1.55	2.09	1.86	137.04	72.92	59.23	18	3	2.8
Male speaker - mabw0, sa1	"She had your dark suit in greasy wash water all year"	0.77	0.69	0.67	1.68	1.71	1.48	211.11	72.71	54.97	6	16	5.5
Male speaker - mabw0, sa2	"Don't ask me to carry an oily rag like that"	0.72	0.73	0.69	1.81	2.09	1.89	162.11	74.35	73.77	25.8	4.3	5.3
	Average	0.68	0.72	0.69	1.53	2.04	1.8	162.98	74.96	66.39	24.53	9.65	7.4

background noise and reverberation [4]. To measure intelligibility we used the metric suggested by [30] which results in values between [0,1]. The results are classified as follows: bad [0,0.3], poor [0.3,0.45], fair [0.45,0.6], good [0.6,0.75], and excellent [0.75,1] [4]. (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 [28]. This metric has been used in speech research for many years to compare speech signals [12] and is also used to evaluate the quality of non-speech audio [15]. A lower LLR indicates better sound quality. (3) Weighted Spectral Slope (WSS) - a distance measure that computes the weighted difference between the spectral slopes in each frequency band [22]. The spectral slope is the difference between adjacent spectral magnitudes in decibels. A lower WSS indicates better speech quality. (4) NIST Speech SNR (NIST-SNR) - the speech to noise ratio defined as the logarithmic ratio between the speech power and noise power estimated over 20 consecutive milliseconds. A higher NIST-SNR indicates better sound quality.

We used the following equipment and configurations to recover sound in all of the experiments conducted and described in this section: a telescope (SkyWatcher with a 35cm lens diameter) was directed at the light bulbs. 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 rest of this section we refer to this setup as the eavesdropping equipment. The level of the played sound was measured with a professional decibel meter.

6.1 Comparing Lamphone to the Visual Microphone

First, we compare the performance of Lamphone to that of the visual microphone [13]. The authors proposing the visual microphone demonstrated the recovery of six sentences from the TIMIT repository [14] by playing the sentences via speakers and analyzing the resulting vibrations of a bag of chips via a high frequency video camera (2200 FPS) from a distance of two meters. We compare Lamphone's performance when recovering the same sentences by analyzing the vibrations of two 12 Watt E14 light bulbs: a (1) hanging light bulb and (2) a light bulb in a desk lamp.

Experimental Setup: We duplicated the experimental setup used in the visual microphone study [13] as follows: We placed speakers on a dedicated stand so their vibrations won't affect the bulbs. We played the same six sentences from the TIMIT repository that were recovered by the visual microphone via the speakers at the same volume level used in the visual microphone study (95 dB). In our experiment, the eavesdropping equipment was placed 2.5 meters from the light bulbs, behind a closed door. Our experimental setup is presented in Fig. 5 .

Results: We applied Algorithm 1 on the optical measurements and recovered speech. The recovered audio signals are available online[1] where they can be heard. The spectrograms of the six recovered sentences can be seen in Figs. 20 and 21 in the appendix. The intelligibility, LLR, WSS, and NIST-SNR of the recovered signals is reported in Table 3 which also contains the results reported in the visual microphone study for each of the six recovered sentences. The results presented in Table 3 reveal four interesting insights: (1) The average intelligibility of the speech recovered by Lamphone from a hanging bulb is 0.04 higher (better) than the average intelligibility of the speech recovered when using the visual microphone. (2) The average LLR of the speech recovered by the visual microphone is lower (better) when using Lamphone with both a desk lamp light bulb (average LLR of .27) and a hanging bulb (average LLR of .51). (3) The average WSS of the speech recovered using Lamphone with both a desk lamp light bulb (average LLR of .27) and a hanging bulb (average LLR of .51) is lower (better) at 96.59 and 88.02 than the speech recovered by the visual microphone. (4) The average NIST-SNR of the speech recovered by the visual microphone is higher (better) than the average NIST-SNR of the recovered speech from the recovered speech when using Lamphone with a desk lamp light bulb (average NIST-SNR of 17.1) and a hanging bulb (average NIST-SNR of 14.8).

Analyzing the results of this set of experiments, we conclude that the quality of the recovered speech by Lamphone and the visual microphone is at the same level. The answer to the question of which method is better depends on the metric used to evaluate the methods. For some metrics, Lamphone method yields better results while in other cases, visual microphone yields better results.

6.2 The Influence of Sound Level and Distance on Lamphone's Performance

Next, we evaluate the influence of distance and sound volume on Lamphone's performance. In this case, we assume that there are no obstacles between the light bulb and the

Table 4: "We Will Make America Great Again!" - Results of Recovered Speech from Various Distances and Sound Levels.

	Intelligibility			LLR			WSS			NIST-SNR		
	60 dB	70 dB	80 dB	60 dB	70 dB	80 dB	60 dB	70 dB	80 dB	60 dB	70 dB	80 dB
5m	0.39	0.57	0.77	2.79	1.55	1.52	290.19	146.99	77.21	6.8	13	24.3
15m	0.34	0.56	0.78	2.2	2.09	1.74	255.12	197.96	126.44	5	24.8	40
25m	0.34	0.55	0.78	3.27	2.3	1.59	255.17	207.01	73.36	1.8	24.8	22.3
35m	0.35	0.51	0.68	2.95	1.92	1.84	280.2	228.11	85.4	5.5	3	10.8
45m	0.31	0.54	0.66	2.42	1.87	1.84	275.03	196.42	91.86	5.3	19.8	10

Figure 13: The light bulb of a desk lamp is located at the end of the hallway at a distance of 45 meters away.

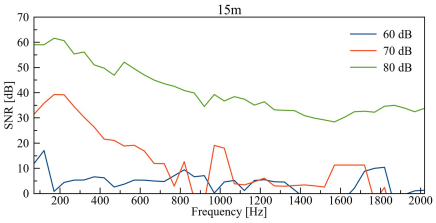

Figure 14: The SNR from a distance of 15 meters away for 60, 70, 80 dB volume levels.

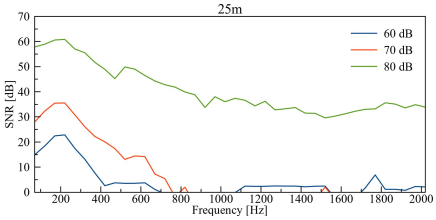

Figure 15: The SNR from a distance of 25 meters away for 60, 70, 80 dB volume levels.

eavesdropping equipment or only transparent obstacles exist between the light bulb and the eavesdropping equipment (e.g., a clear glass door/window).

We evaluate the performance of Lamphone for recovering sound at normal speech levels (60, 70, 80 dB). A conversation at 60 db is a normal level for a conversation between two people located near one another. A conversation at 80 dB is the average sound level for a conversation held via Zoom.

In the following set of experiments we tried to recover sound from a 12 watt E14 desk lamp (placed on a desk) light bulb from various distances. We placed speakers on a dedicated stand so their vibrations would not affect the bulb; the eavesdropping equipment was located behind two closed clear glass doors. The setup can be seen in Fig. 13.

First, we start by testing 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 and 25 meters away from the light bulb. We played the audio file via speakers at 60, 70, and 80 dB while obtaining the optical measurements. The electro-optical sensor was configured for the highest gain level before saturation.

Results: Figs. 14 and 15 present the SNR for distances of 15 and 25 meters for each of the three sound levels measured. As can be seen from the results, the SNR for 80 dB looks very promising through the entire spectrum measured. The SNR for 70 db reaches a noise level around 800 Hz, so effectively there is a narrower bandwidth that allows sound recovery compared to 80 dB. The SNR for 60 dB is very low, and only sound at low frequencies can be recovered.

Next, we evaluated Lamphone's performance in terms of its ability to recover speech audio from various distances. In order to do so, we decided to recover a famous statement made by President Donald Trump: "We will make America great again."

Experimental Setup: We placed the eavesdropping equipment at five distances (5, 15, 25, 35, 45 meters) from the light bulb. We played the audio file via speakers at 60, 70, and 80 dB while obtaining the optical measurements. The electro-optical sensor was configured for the highest gain level before saturation.

Results: We applied Algorithm 1 on the optical measurements and recovered speech. The recovered audio signals are available online[1] where they can be heard. The spectrograms of the the recovered speech can be seen in Figs. 22 and 24 in the appendix. The intelligibility, LLR, WSS, and NIST-SNR of the recovered signals are reported in Table 4. The results reveal three interesting insights: (1) For a sound level of 80 dB, the intelligibility of the recovered signals is considered excellent up to a distance of 25 meters (0.77) and good from 45 meters away (0.66). (2) For a sound level of 70 dB, the intelligibility of the recovered signals is fair for a distance of 5-45 meters. (3) For a sound level of 60 dB, the intelligibility is considered poor for a distance of 5-45 meters away.

The results obtained showed that Lamphone allows eavesdroppers to recover sound in real time at 70 dB from a distance of 45 meters at a lower sound level than eavesdropping methods proposed in previous studies which require higher sound levels of 75-84 dB (e.g., [10, 23, 29, 36]), 85-94 dB (e.g., [17, 20]), +95 dB (e.g., [13, 33]). However, to improve the intelligibility and increase the effective range of the attack, eavesdroppers can increase the system's sensitivity (see Section 7 for suggested improvements).

6.3 Recovering Sound Using Light Emitted Through Curtain Walls

Next, we evaluate the performance of Lamphone for recovering sound at 80 dB from a desk lamp light bulb located in an office building at our university that is covered by cur-

Figure 16: Experimental setup: From left to right (1) The distance between the eavesdropper (located on a pedestrian bridge) to the LED bulb of a desk lamp (in an office on the third floor of a building covered by curtain walls) is 25 meters, (2) the location of the eavesdropper, and (3) the office with the desk lamp with an LED bulb.

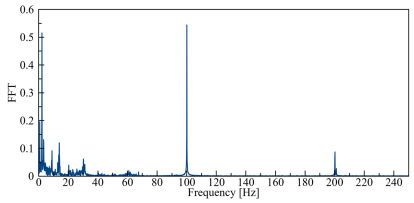

Figure 17: FFT graph of optical measurements when no sound is played.

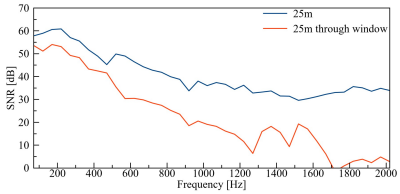

Figure 18: SNR obtained from a distance of 25 meters away w/o curtain walls.

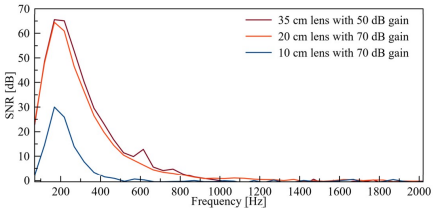

Figure 19: SNR obtained from telescopes with lens diameters of 10, 20, 35 cm

tain walls. The use of curtain walls is common in modern office buildings, since they are designed to allow the offices to benefit from natural light. In the previous set of experiments we showed that eavesdroppers can recover speech at 80 dB from 25 meters away with excellent intelligibility (0.77) when a direct line of sight to a light bulb exists, even if a clear glass window/door exists between the light bulb and the eavesdropping equipment. This time we assess Lamphone's performance for recovering sound from a light bulb, using light emitted through a curtain wall, in order to evaluate the quality of recovered audio when there is no clear/transparent line of sight.

Fig. 16 presents the experimental setup. A desk lamp with a 12 watt E14 LED light bulb was placed in an office at our university located on the third floor of a building covered by curtain walls, which reduce the amount of light emitted from the offices (as can be seen in Fig. 16). We placed the eavesdropping equipment on a pedestrian bridge, positioned an aerial distance of 25 meters away from the target office.

We start by examining the effect of the setup on the optical measurements obtained. We note that the setup is very challenging since: (1) there is no clear line of sight between the eavesdropping equipment and the light bulbs (i.e., curtain walls are placed in between and reduce the amount of emitted light), and (2) the pedestrian bridge, on which the eavesdropping equipment is placed, is located above a train station and railroad tracks which have a natural vibration of their own.

Experimental Setup: We directed the telescope at the desk lamp light bulb in the office (as can be seen in Fig. 16). We obtained measurements for three seconds via the electro-optical sensor when no sound is played in the office.

Results: As can be seen from the FFT graph presented in Fig. 16, the peaks of 100 Hz and 200 Hz, which are the result of the lighting frequency of the light bulb, are part of the signal (as discussed in Section 4). However, we observed a very interesting phenomenon in which noise is added to the low frequencies (< 40 Hz) of the optical signal. This phenomenon is the result of the natural vibration of the bridge. Since this phenomenon adds substantial noise to the signal obtained, we used a high-pass filter (> 40 Hz) to optimize the results.

Experimental Setup: We created an audio file that consists of various sine waves (120, 170, 220, ... 1970 Hz) where each sine wave was played for two seconds. We played the audio file via speakers at 80 dB while obtaining the optical measurements. We directed the telescope at the desk lamp light bulb in the office. The electro-optical sensor was configured for the highest gain level before saturation.

Results: Fig. 18 shows the SNR obtained from this experiment (where curtain wall exists between the light bulb and the eavesdropping equipment). Fig. 18 also shows the SNR obtained from a distance of 25 meters when there is a clear line of sight between the eavesdropping equipment and the light bulb (i.e., when there is no curtain wall in between). Fig. 18 reveals an interesting insight: The loss of light (a result of the curtain wall) decreases the SNR significantly, especially for the band beyond 600 Hz. As a result, there is a narrower band that can be used to recover sound with a high SNR.

Next, we evaluated Lamphone's performance in terms of its ability to recover speech and non-speech audio.

Experimental Setup: We decided to recover two well-known songs: "Let it Be" by The Beatles and "Clocks" by Coldplay and the following two sentences: "We will make America great again" spoken by Donald Trump and "Mary had a little lamb whose fleece was white as snow and every

Table 5: Results of Recovered Speech and Songs Through Curtain Walls From 25 Meters.

	Intelligibility	LLR	WSS	NIST SNR
"We will make America great again!"	0.59	3.38	175.69	8.5
"Mary had a little lamb whose fleece was white as snow and every where that Mary went the lamb was sure to go"	0.54	3.22	147.18	10.5
Clocks	0.41	1.66	99	3.8
Let it Be	0.34	6.19	154.27	12

where that Mary went the lamb was sure to go" from the TIMIT repository [14]. We played the four audio files in the office via speakers at 80 dB. We directed the telescope at the desk lamp light bulb in the office (as can be seen in Fig. 16). The electro-optical sensor was configured for the highest gain level before saturation.

Results: We applied Algorithm 1 on the optical measurements and recovered speech. The recovered audio signals are available online[1] where they can be heard. The spectrograms of the the recovered speech are presented in Figs. 25 - 28 in the appendix. The intelligibility, LLR, WSS, and NIST-SNR of the recovered signals are reported in Table 4. Interestingly, the results show that eavesdroppers can recover speech at 80 dB from 25 meters away with fair intelligibility by exploiting light emitted through curtain walls. We note that both songs are recognized by Shazam.

7 Potential Improvements

In this section, we suggest methods that eavesdroppers can use to optimize the quality of the recovered audio or increase the distance between the eavesdropper and the light bulb, without changing the setup of the target location. The potential improvements suggested below are presented based on the component they are aimed at optimizing.

Telescope: The amount of light that is captured by a telescope with diameter r 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. In order to prove this claim, we compared the SNR obtained by directing an electro optical sensor through three telescopes with lens diameters of 10, 20, 35 cm from a distance of 25 meters. The results can be seen in Fig. 19 which presents the SNR obtained from three telescopes (with lens diameters of 10, 20, 35 cm). The SNR of the recovered audio signal obtained by the telescope with a lens diameter of 35 cm and an electro-optical sensor gain of 50 dB is identical to the SNR of the recovered audio signal obtained by the telescope with a lens diameter of 20 cm and an electro-optical sensor gain of 70 dB. Eavesdroppers can exploit this fact and use a telescope with a larger lens diameter in order to optimize the quality of the signal captured.

Electro-Optical Sensor: The sensitivity of the system can be enhanced by increasing the internal gain of the sensor. 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 [-5,5] and supports a maximum internal gain of 70 dB). However, any amplification that increases the signal obtained beyond this range results in saturation that prevents the SNR from reaching its full potential. This claim is demonstrated in Fig. 19 which presents the SNR obtained from three telescopes (with lens diameters of 10, 20, 35 cm). Since the signal that was captured by the telescope with a lens diameter of 35 cm was very strong (due to the fact that a lot of light was captured by the large lens), we could not increase the internal gain to a level beyond 50 dB from a distance of 25 meters. As a result, the SNR obtained by the telescope with a lens diameter of 35 cm did not reach its full potential and yielded the same SNR as a telescope with a lens diameter of 20 cm and an electro-optical sensor gain of 70 dB. With that in mind, eavesdroppers can optimize the SNR of the optical measurements by using a sensor that supports a wider range of output. Another option is to sample the signal from multiple sensors. Given N sensors that sample a signal, the SNR increases by $\sqrt{N}$. Thus, eavesdroppers can optimize the SNR of the optical signal by obtaining measurements using several electro-optical sensors directed at the light bulb and sample the bulb's vibrations simultaneously from several channels.

Sound Recovery System: The sound recovery system implemented in this paper uses the digital approach and consists of two components: an ADC and a sound recovery algorithm. As discussed in Section 4, a 24-bit ADC with an input range of [-5,5] voltage provides a sensitivity of 0.6 uV (see Equation 1). Only bulb vibrations that are expected to yield a greater voltage change (i.e., >0.6 uV) can be recovered by Lamphone. A 32-bit ADC provides a higher level of sensitivity of 2.3 nV and optimizes the system's sensitivity significantly (see Equation 2). In addition, many advanced denoising methods have been suggested by experts in the field of speech enhancement. Advanced algorithms (e.g., neural networks) provide excellent results for filtering the noise from an audio signal, however often a large amount of data is 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 (e.g., normalization, spectral subtraction, noise gating, etc.). Another option for maximizing the SNR is to profile the electro-optical sensor's thermal noise (when the light is recorded) in order to filter the noise that is added to the analog output of the sensor.

8 Countermeasures

In this section, we describe several countermeasure methods that can be used to mitigate or prevent the Lamphone attack. There are several factors that influence the SNR of the

recovered audio signal which are in the victim's control and can be used to prevent/mitigate the attack.

One approach is aimed at reducing the amount of light captured by the electro-optical sensor. As shown in Fig. 18, the SNR decreases as the amount of light captured by the electro-optical sensor decreases. Several techniques can be used to limit the amount of light emitted. For example, weaker bulbs can be used; the difference between a 12 watt E27 bulb and a 9 watt E27 bulb is negligible for lighting an average room. However, since a 9 watt E27 bulb emits less light than a 12 watt E27 bulb, less light is captured by the electro-optical sensor, and the quality of the recovered audio signal decreases. Another technique is to use curtain walls. As was shown in Section 18, curtain walls decrease the quality of the recovered audio significantly. Another approach is to limit the light bulb's vibrations. Lamphone relies on the fluctuations in air pressure on the surface of a light bulb which result from sound and cause the bulb to vibrate. A light bulb's vibration can be reduced by using a heavier bulb. There is less vibration from a heavier bulb in response to air pressure on the bulb's surface (as was shown in Section 4, the SNR of the E27 light bulb is lower than the SNR of the E14 light bulb).

9 Limitations, Discussion & Future Work

Lamphone suffers from a few disadvantages. (1) Inexpensive hardware and equipment (like the telescope, electro-optical sensor, and ADC used in our experiments) can be used to recover sound from 45 meters away with fair intelligibility for speech at 70/80 dB. However, in order to increase the attack range and recover high quality sound, more expensive and professional equipment is required (e.g., a more sensitive ADC and electro-optical sensor, a professional telescope). (2) The equipment used by the eavesdropper must be positioned near the target room, but in a location that will not raise suspicion or lead to detection; it may be difficult or costly for the eavesdropper to identify a location that meets these requirements (e.g., a room in a nearby building or a parked van). As a result, we consider Lamphone an attack that can be applied by parties with great financial resources (e.g., armies and police departments) and not by average civilians.

We believe that in the next few years, new studies will improve the proposed method of recovering sound from light, so it will pose a greater threat to individuals' privacy; future research may improve the method such that it could even be applied by eavesdroppers with less resources and from a greater distance between the speaker/eavesdropping equipment and the light bulb. A pressing question that arises is: how long will it take the scientific community to improve this method sufficiently so it will pose a greater threat to individuals' privacy. An analysis of the scientific progress of another eavesdropping method might help answer this question. The Gyrophone method of recovering sound from a smartphone's motion sensors [23] was first introduced at USENIX 2014. The main disadvantage of Gyrophone at that time was the low accuracy of the model that was used to classify isolated words (the accuracy was only slightly better than a random guess). However, greater understanding regarding this eavesdropping technique and the threat model was obtained over the years by other studies [10, 17, 36], and a recent study presented at NDSS 2020 was able to improve this method such that it could be used to classify isolated words from a smartphone's accelerometer with 99% accuracy [11]. Based on the progress made in the last six years since Gyrophone was first introduced, we believe that future studies will improve understanding on Lamphone and suggest new ways to overcome the abovementioned limitations.

For future work, we suggest investigating whether sound can be recovered via other lightweight sources (e.g., decorative LED flowers), how inexpensive equipment could be used to improve the range of the attack (e.g., by improving the recovery model), how to extend the distance between the light bulb and the speaker (e.g., by using artificial bandwidth extension [18, 19, 21, 26, 27]), and how to apply the attack by using compact equipment.

References

[1] Eavesdropping. https://en.wikipedia.org/wiki/Eavesdropping.

[2] Facts about speech intelligibility. https://www.dpamicrophones.com/mic-university/facts-about-speech-intelligibility.

[3] How microphones work. https://www.mediacollege.com/audio/microphones/how-microphones-work.html.

[4] Intelligibility. https://en.wikipedia.org/wiki/Intelligibility_(communication).

[5] Microphones. https://www.explainthatstuff.com/microphones.html.

[6] Mpu 6050 gy-521 3 axis gyro accelerometer sensor module arduino.

[7] Ni 9234 datasheet.

[8] Pda100a2.

[9] Spherical coordinate system. https://en.wikipedia.org/wiki/Spherical_coordinate_system.

[10] S. A. Anand and N. Saxena. Speechless: Analyzing the threat to speech privacy from smartphone motion sensors. In *2018 IEEE Symposium on Security and Privacy (SP)*, volume 00, pages 116–133.

[11] 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 *Proceedings of the Network and Distributed Systems Security (NDSS) Symposium*, pages 23–26, 2020.

[12] R Crochiere, J Tribolet, and L Rabiner. An interpretation of the log likelihood ratio as a measure of waveform coder performance. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 28(3):318–323, 1980.

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

[14] John S Garofolo, Lori F Lamel, William M Fisher, Jonathan G Fiscus, and David S Pallett. Darpa timit acoustic-phonetic continous speech corpus cd-rom. nist speech disc 1-1.1. *STIN*, 93:27403, 1993.

[15] Wang Guang-Yan, Zhao Xiao-qun, and Wang Xia. Musical noise reduction based on spectral subtraction combined with wiener filtering for speech communication. 2009.

[16] 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)*, Vancouver, BC, 2017. USENIX Association.

[17] 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*, IPSN '17, pages 181–192, New York, NY, USA, 2017. ACM.

[18] Vasu Iyengar, Rafi Rabipour, Paul Mermelstein, and Brian R Shelton. Speech bandwidth extension method and apparatus, October 3 1995. US Patent 5,455,888.

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

[20] A. Kwong, W. Xu, and K. Fu. Hard drive of hearing: Disks that eavesdrop with a synthesized microphone. In *2019 2019 IEEE Symposium on Security and Privacy (SP)*, Los Alamitos, CA, USA, may 2019. IEEE Computer Society.

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

[22] Jianfen Ma, Yi Hu, and Philipos C Loizou. Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions. *The Journal of the Acoustical Society of America*, 125(5):3387–3405, 2009.

[23] Yan Michalevsky, Dan Boneh, and Gabi Nakibly. Gyrophone: Recognizing speech from gyroscope signals. In *23rd USENIX Security Symposium (USENIX Security 14)*, pages 1053–1067, San Diego, CA, 2014. USENIX Association.

[24] Ralph P Muscatell. Laser microphone, October 25 1983. US Patent 4,412,105.

[25] Ralph P Muscatell. Laser microphone, October 23 1984. US Patent 4,479,265.

[26] Hannu Pulakka and Paavo Alku. 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):2170–2183, 2011.

[27] Hannu Pulakka, Ulpu Remes, Santeri Yrttiaho, Kalle Palomaki, Mikko Kurimo, and Paavo Alku. 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):2219–2231, 2012.

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

[29] 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*, MobiSys '16, pages 57–69, New York, NY, USA, 2016. ACM.

[30] Cees H Taal, Richard C Hendriks, Richard Heusdens, and Jesper Jensen. An algorithm for intelligibility prediction of time–frequency weighted noisy speech. volume 19, pages 2125–2136. IEEE, 2011.

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

[32] G. Wang, Y. Zou, Z. Zhou, K. Wu, and L. M. Ni. We can hear you with wi-fi! *IEEE Transactions on Mobile Computing*, 15(11):2907–2920, Nov 2016.

[33] 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, MobiCom '15, pages 130–141, New York, NY, USA, 2015. ACM.

[34] Andrew M White, Austin R Matthews, Kevin Z Snow, and Fabian Monrose. Phonotactic reconstruction of encrypted voip conversations: Hookt on fon-iks. In *Security and Privacy (SP), 2011 IEEE Symposium on*, pages 3–18. IEEE, 2011.

[35] Charles V Wright, Lucas Ballard, Scott E Coull, Fabian Monrose, and Gerald M Masson. Spot me if you can: Uncovering spoken phrases in encrypted voip conversations. In *Security and Privacy, 2008. SP 2008. IEEE Symposium on*, pages 35–49. IEEE, 2008.

[36] 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. ACM, 2015.

10 Appendix - Spectrograms of Recovered Speech

The spectrograms of the recovered speech for the six sentences that were used to compare the results to the results obtained by the visual microphone are presented in Figs. 20 and 21.

The spectrograms of the recovered speech ("We Will Make America Great Again!") from various distances (5, 15, 25, 35, 45 meters) and with sound levels (60, 70, 80 dB) are presented in Figs. 24 - 22.

The spectrograms of the recovered speech and songs from the bridge are presented in Figs. 25 - 28.

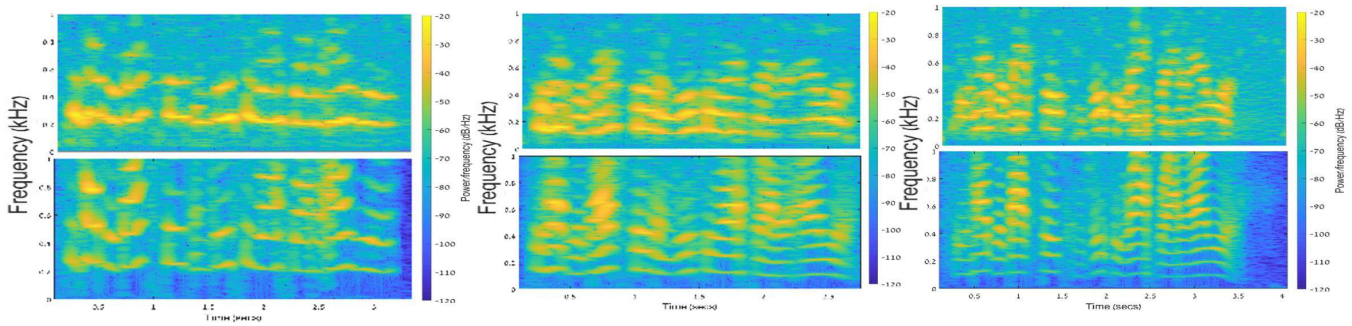

Figure 20: fadg0,mccs0,mabw0 sa1: "She had your dark suit in greasy wash water all year." Recovered (top) and original (bottom) speech.

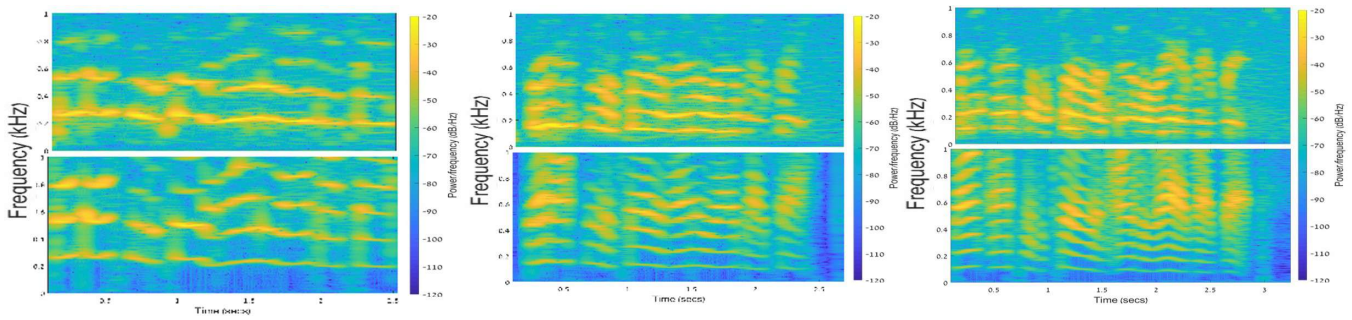

Figure 21: fadg0,mccs0,mabw0 sa2: "Don't ask me to carry an oily rag like that." Recovered (top) and original (bottom) speech.

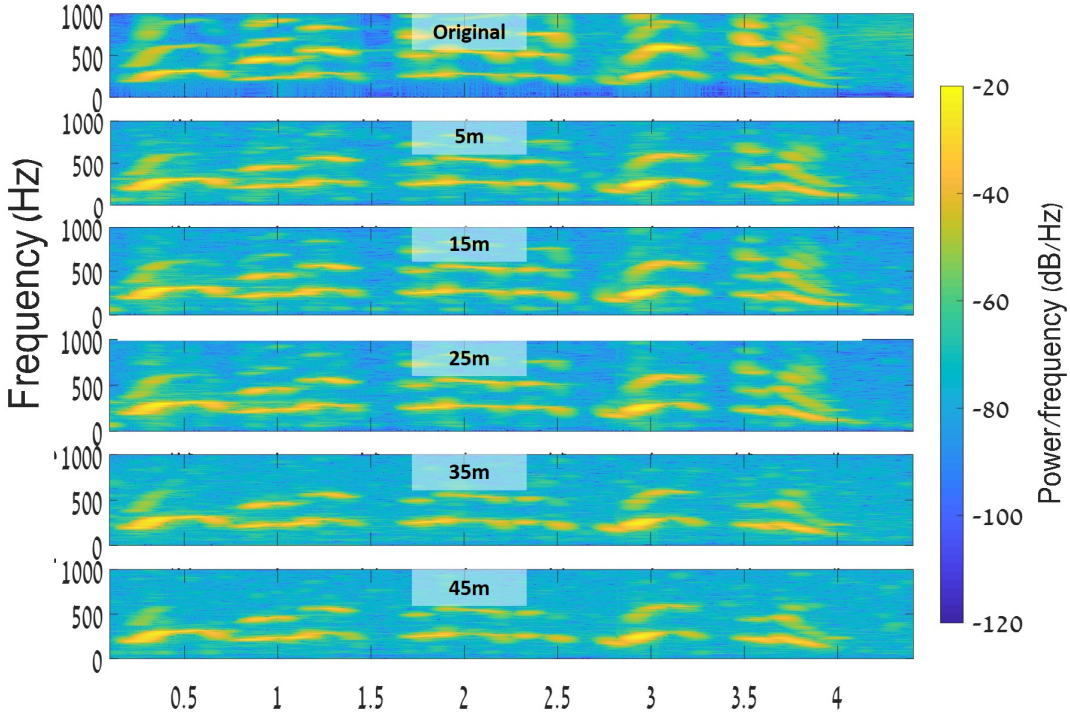

Figure 22: "We will make America great again" played at 80 dB and recovered from various distances

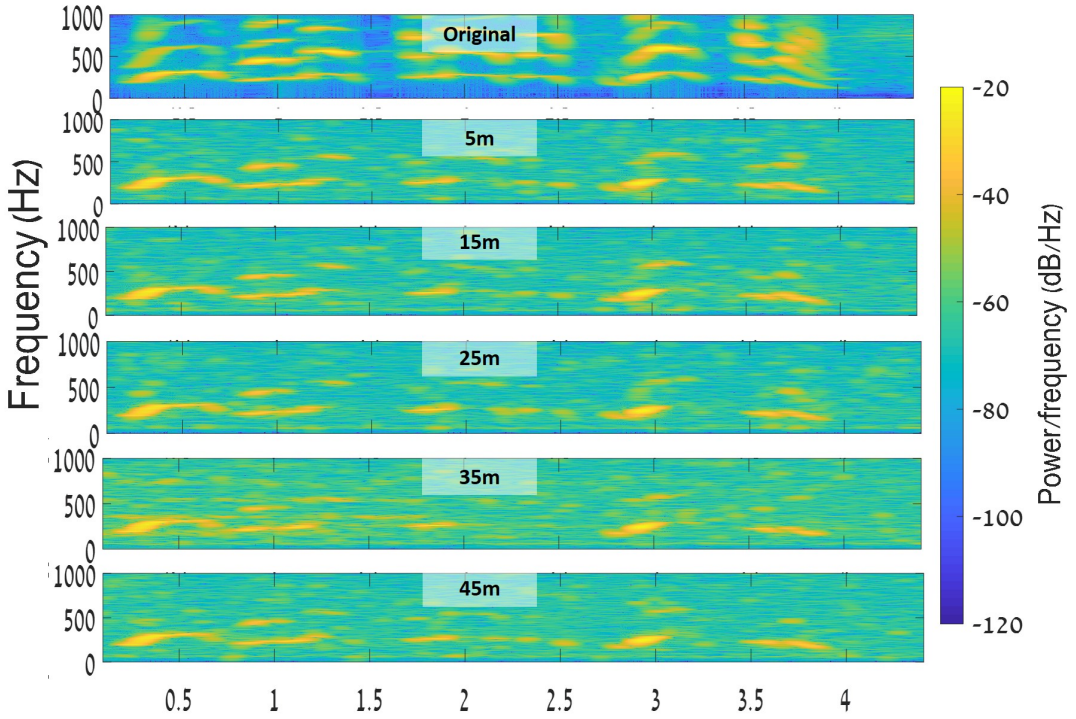

Figure 23: "We will make America great again" played at 70 dB and recovered from various distances

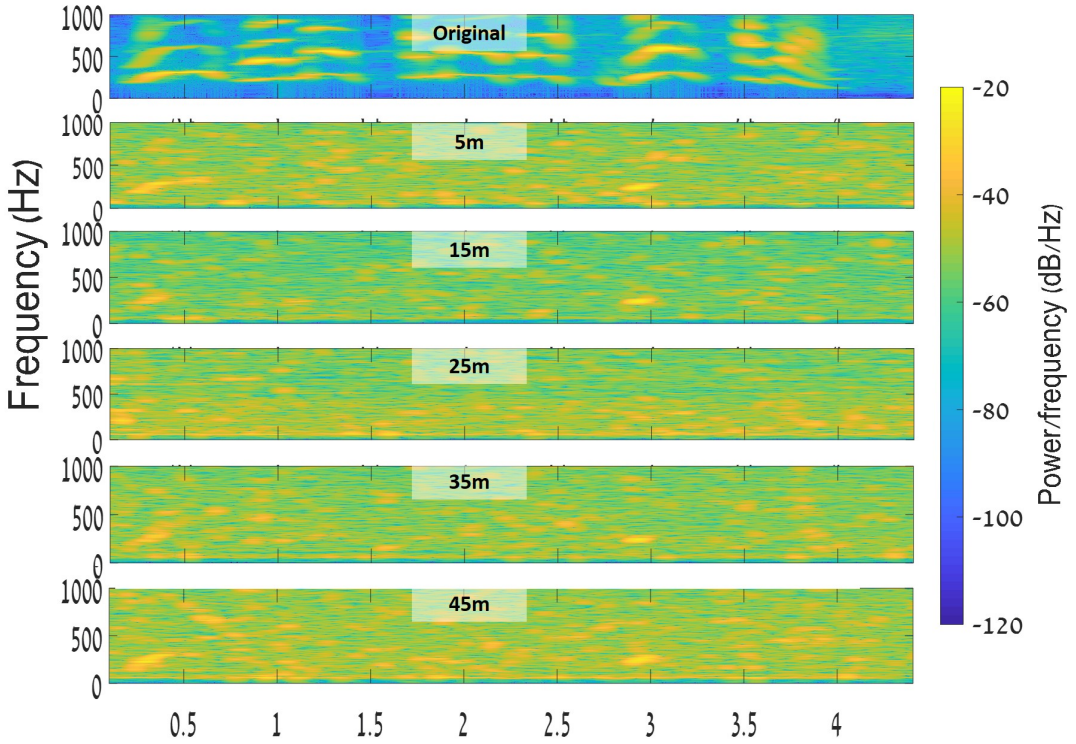

Figure 24: "We will make America great again" played at 60 dB and recovered from various distances

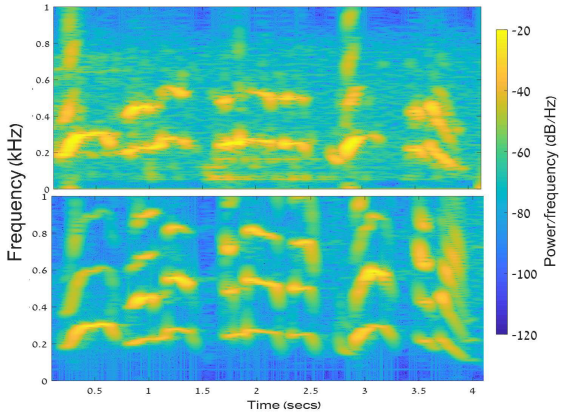

Figure 25: "We will make America great again" recovered from a bridge 25 meters away from the target office. Recovered (top) and original (bottom) speech.

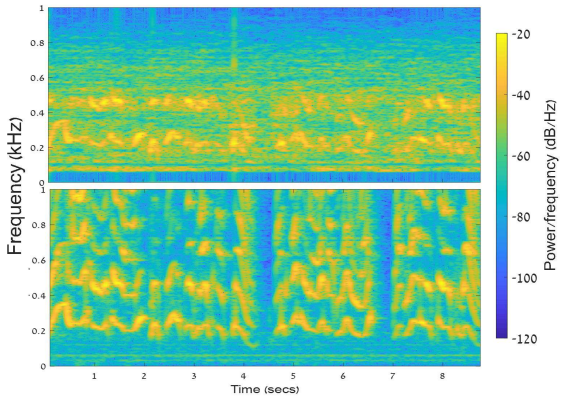

Figure 26: "Mary had a little lamb" recovered from a bridge 25 meters away from the target office. Recovered (top) and original (bottom) speech.

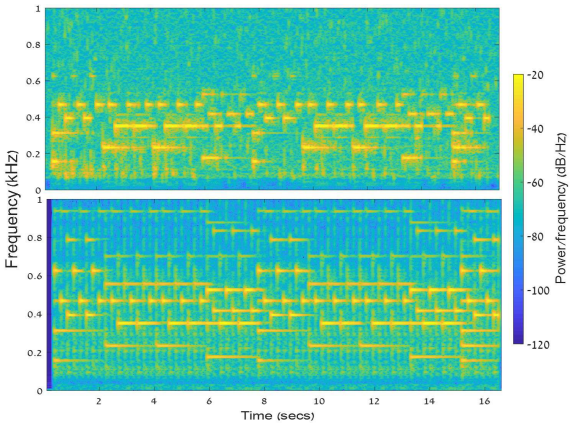

Figure 27: Clocks by Coldplay recovered from a bridge 25 meters away from the target office. Recovered (top) and original (bottom) song.

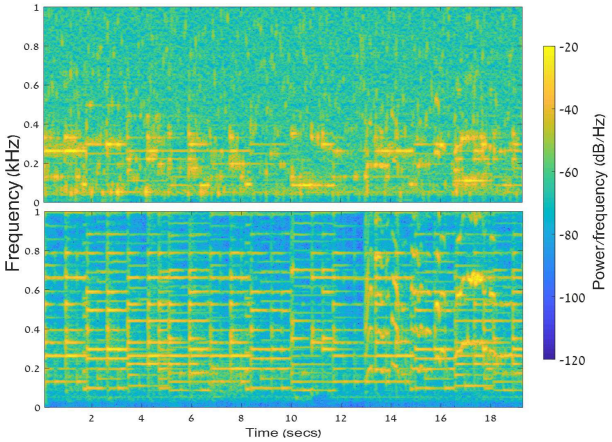

Figure 28: Let it be by The Beatles recovered from a bridge 25 meters away from the target office. Recovered (top) and original (bottom) song.