Lamphone: Passive Sound Recovery from a Desk Lamp's Light Bulb Vibrations

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

[1] Ben-Gurion University of the Negev, [2] Weizmann Institute of Science
{nassib, yaronpir, razsw, elovici, zadov}@post.bgu.ac.il, adi.shamir@weizmann.ac.il

Abstract

In this paper, we introduce "Lamphone," an optical side-channel attack used to recover sound from desk lamp light bulbs; such lamps are commonly used in home offices, which became a primary work setting during the COVID-19 pandemic. We show how fluctuations in the air pressure on the surface of a light bulb, which occur in response to sound and cause the bulb to vibrate very slightly (a millidegree vibration), can be exploited by eavesdroppers to recover speech passively, externally, and using equipment that provides no indication regarding its application. 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 compare Lamphone to related methods presented in other studies and show that Lamphone can recover sound at high quality and lower volume levels that those methods. Finally, we show that eavesdroppers can apply Lamphone in order to recover speech at the sound level of a virtual meeting with fair intelligibility when the victim is sitting/working at a desk that contains a desk lamp with a light bulb from a distance of 35 meters.

1 Introduction

The COVID-19 era, with its lockdowns, social distancing, isolation, and quarantining, forced many people to stay at home, work from home, and refrain from attending in-person meetings. COVID-19 restrictions created a risky reality where meetings that were usually held in person in secured buildings (e.g., offices, clinics, banks) were replaced by phone calls and virtual meetings in which the participants were based in unsecured home environments (e.g., home offices). As a result, personal, confidential, and business information is frequently being exchanged from unsecured home environments in countries that are still under severe COVID-19 restrictions and in countries that have already overcome COVID-19 where companies have adopted a hybrid working approach.

The sensitive information shared by individuals in home environments may attract eavesdroppers seeking to obtain valuable information revealed in these settings for various malicious purposes (e.g., spying, shaming, blackmailing). Various side-channel attacks that can be applied by eavesdroppers to recover sound from unsecured environments have been demonstrated [7, 8, 12, 14, 15, 18, 22, 23, 27, 28, 31, 32, 35]. These studies have contributed to improved understanding regarding the risks of sound eavesdropping via non-acoustic data. These methods suffer from at least one of the following disadvantages: (1) Some methods rely on obtaining and exfiltrating data from a compromised device located in proximity of a target/victim [7, 8, 14, 15, 18, 22, 27, 28, 35], a fact that requires the eavesdropper to compromise a target device with malware in advance. (2) Some methods are limited to the classification of isolated words that appear in a precompiled dictionary [7, 8, 15, 22, 27, 28, 35], a fact that requires the eavesdropper to invest additional time in compiling words for a dictionary. (3) Other methods are limited to recovering sound at a high volume (+85 dB) [12, 18, 31, 32], a fact that limits their effectiveness at recovering speech from virtual meetings (the sound level of such meetings is typically 75 dB). (4) Some methods rely on spying equipment, a fact that limits their availability in some countries due to regulations restricting the sale of this equipment [23]. (5) Other methods require that an active laser beam be directed at objects located near the target [23], a fact that increases the likelihood of detection.

In this paper, we identify a new risk to personal and business meetings held in unsecured home environments. We show that the presence of desk lamps (and more specifically the light bulbs used in desk lamps), which are commonly used in home offices and bedrooms, can be exploited by eavesdroppers to recover intelligible speech at levels of 75 dB externally, passively, and using equipment that is not associated with spying and provides no indication regarding its application. We introduce "Lamphone," a novel side-channel attack capable of recovering speech optically via an electro-optical sensor directed at a desk lamp's 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 developed an optical-acoustic transformation (OAT) to isolate the audio signal from the optical signal obtained by directing an electro-optical sensor at a desk lamp's bulb. Finally, we evaluate Lamphone's performance on the task of recovering sound and show that Lamphone is capable of recovering speech audio at the sound level of a virtual meeting with fair intelligibility from a distance of 35 meters.

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 our optical-acoustical transformation in Section 5, and in Section 6, we evaluate Lamphone's performance on the task of sound recovery. 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 Related Work

In this section, we review related research focused on eavesdropping methods. Several studies [7, 8, 15, 22, 28, 35] have shown that measurements obtained from motion sensors located in proximity of a victim can be used for the classification of words. They variously demonstrated that the response of MEMS gyroscopes [22], accelerometers [7, 8, 35], and geophones [15] to sound at 75-85 dB can be used to classify words and identify speakers and their genders, even when the sensors are located in a smartphone and the sampling rate is limited to 200 Hz. Two other studies [14, 27] showed that the process of output devices can be inverted to recover speech. In [27], the authors established a microphone by recovering sound at 80 dB from a vibration motor, and in [14], the audio from speakers was recovered. A recent study [18] 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.

Two studies [31, 32] used the physical layer of Wi-Fi packets to eavesdrop sound at 95 dB. In [32], the authors suggested a method that analyzes the received signal strength 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. In [31], the authors suggested a method that analyzes the channel state information of Wi-Fi packets sent from a router to classify words.

Two other studies were able to recover speech from encrypted VoIP data by exploiting side effects of the compression process (variable bitrate) [33, 34]. A recent study suggested a new TEMPEST attack against devices to recover sound from speakers by analyzing the emitted EMR (electromagnetic radiation) [10].

Other methods utilized optical sensors to recover sound [12, 23]. The laser microphone [23] is a well-known method that uses an external device. In this case, the eavesdropper directs a laser beam 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. The method most related to our research is the visual microphone [12]. In this method, the eavesdropper analyzes the response of material inside the victim's room (e.g., a bag of chips) to sound waves at 95 dB, using video obtained from a high speed video camera (2200 FPS), and recovers speech.

3 Threat Model

In this section, we describe the threat model and compare its significance to methods presented in other studies.

Assumptions. We assume that the target of this attack is a person located inside his/her house, exchanging/sharing sensitive information in a phone call or virtual meeting. We assume that the victim makes the call/attends the meeting from an office/room that contains a lamp with a light bulb (e.g., a home office or bedroom) and is positioned up to 50 cm away from the light bulb, a reasonable distance for an individual seated at a standard desk with a desk lamp in a home office or seated on a bed, next to a nightstand with a tabletop lamp, in a bedroom. We consider the eavesdropper to be a malicious entity interested in recovering speech from the victim's conversation by performing the Lamphone attack. The eavesdropper could use the recovered information for various malicious purposes, including spying, shaming, blackmailing, business intelligence gathering, etc. We assume that the eavesdropper is located within 35 meters of the target room. The eavesdropper could be: (1) a person located in a room in an adjacent building (e.g., a nosey neighbor), or (2) a person in a nearby car (e.g., a private detective). We consider this threat to be highly likely in the COVID-19 era due to the increased number of personal and business meetings being held in unsecured home environments.

Components. The Lamphone attack 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. (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 data using a sound recovery algorithm running on a laptop. In this study,

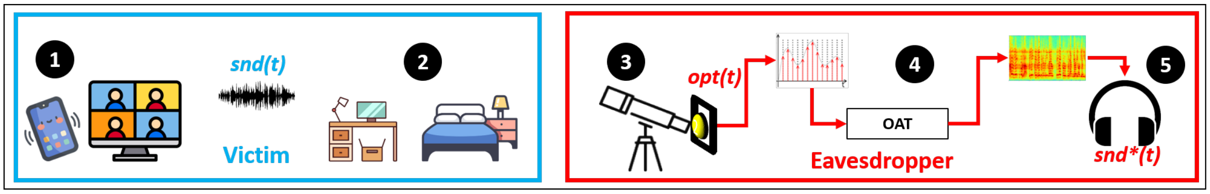

Figure 1: Lamphone's threat model: The sound $snd(t)$ from the victim's room (1) creates fluctuations on the surface of the desk lamp's light bulb (the diaphragm) (2). The eavesdropper directs an electro-optical sensor (the transducer) at the light bulb via a telescope (3). The optical signal $opt(t)$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

we use the latter digital approach.

The physical phenomenon. The conversation held in the target room creates sound $snd(t)$ that results in fluctuations in the air pressure on the surface of the light bulb, which causes the bulb to vibrate. As a result, a static electro-optical sensor that is directed by the eavesdropper via a telescope towards the vibrating light bulb captures the changes in the light intensity over time that result from the minuscule bulb vibrations. The time series of light intensities, which are correlative to the air pressure that hits the surface of the bulb and produced by nearby speech/sound, represent a modulation of the speech/sound inside the victim's room. The changes in the analog output of the electro-optical sensor are sampled by the ADC to a digital optical signal $opt(t)$. The eavesdropper then transforms the optical signal $opt(t)$ to an acoustic signal $snd^*(t)$ using an optical-acoustical transformation (OAT). Fig.1 outlines the threat model.

In general, microphones rely on three components (a diaphragm, transducer, and ADC). In Lamphone, the light bulb is used as a diaphragm, which vibrates when sound waves hit its surface. The transducer, which is used to convert the diaphragm's vibrations to electricity, consists of the light emitted from the light bulb (in the target room) and the electro-optical sensor (used by the eavesdropper), which creates the associated electricity. An ADC is used to convert the electrical signal to a digital signal (as in standard microphones).

Significance. The significance of Lamphone with respect to related methods presented in other studies is that Lamphone: (1) is an external method that relies on a line of sight between the electro-optical sensor and the light bulb (as opposed to other methods that require eavesdroppers to compromise a device located in physical proximity of the victim in order to obtain data and exfiltrate it [7,14,15,18,22,27,31,32,35]), (2) relies on an electro-optical sensor that is passive and does not provide any indication regarding its use (as opposed to the laser microphone [23]), (3) is composed of hardware (ADC, photodiode) that is not associated with spying (as opposed to the laser microphone [23]), (4) recovers intelligible audio signals, so it is not limited to classifying isolated words that appear in a precompiled dictionary (as opposed to [7,15,22,31,35]), and (5) capable of recovering speech at a virtual meeting's sound level of 75 dB (as opposed to [12,18,31,32]).

The two methods most related to ours are the visual microphone [12] and the laser microphone [23]. Both methods recover sound using optical sensors: the laser microphone recovers speech at standard sound levels, however it utilizes an active optical transceiver for this task (which is indicative of its use and considered spying equipment, which limits its availability). The visual microphone recovers speech by utilizing a passive high-frequency video camera (equipment which is not associated with spying), however it is limited to recovering speech at a high volume level (an average volume level of 95 dB), which is beyond the sound level of a virtual meeting (such meetings have an average volume level of 75 dB). Lamphone combines the advantages of both methods: Lamphone utilizes a passive photodiode (which is not considered spying equipment or indicative regarding its use) and is effective at recovering speech at the sound level of a typical virtual meeting. In addition, Lamphone is computationally lighter than the visual microphone. In Lamphone, an ADC is used to create the optical signal (time series) by sampling a photodiode that is used to convert light to electricity which is associated with the amount of light captured by the sensor (and changes due to the bulb's minuscule vibrations). In contrast, the visual microphone extracts the optical signal by analyzing a video stream and extracting a time series that is associated with an object's vibration over time. The optical signal (time series) is created by converting each frame obtained by the high-frequency video camera (an operation that consists of computations made on three HD resolution matrices) to a single value (scalar) in time. This requires an additional computation stage in order to create the optical signal due to the fact that the desired frequency of 2200 Hz is obtained by converting 2200 frames to 2200 scalars each second. As a result, the visual microphone requires greater computational resources than Lamphone.

4 Bulbs as Microphones

In this section, we describe 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 Physical Analysis

First, we measure the vibration of a light bulb when sound waves hit the light bulb's surface and establish a criterion for

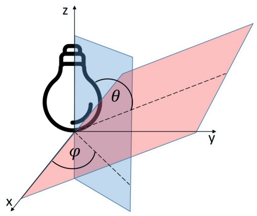

Figure 2: A 3D scheme of a light bulb's axes.

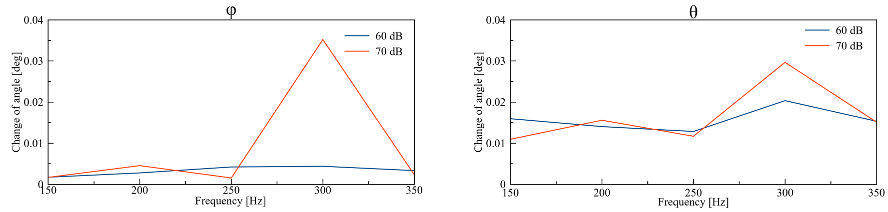

Figure 3: Peak-to-peak difference of angles ϕ (left) and θ (right) for an E14 light bulb in the 150-350 Hz spectrum.

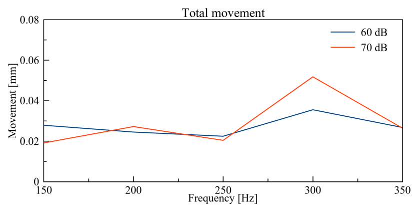

Figure 4: The peak-to-peak movement for an E14 bulb in the range of 150-350 Hz.

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 [2]) to the bottom of an E14 LED light bulb (10 watts); the bulb was not illuminated during this experiment (see Fig. 21 in the appendix). A Raspberry Pi 3 was used to sample the gyroscope at 700 Hz. We placed Logitech Z533 speakers in front of the bulb (a few centimeters away) and played various sine waves (150, 200, 250, 300, 350 Hz) from the speakers at two volume levels (60 dB and 70 dB). We obtained measurements from the gyroscope while the sine waves were played.

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).We calculated the peak-to-peak values, since they reflect the distance between the farthest and closest points that the light bulb reaches when it vibrates. The average peak-to-peak difference was computed by calculating the peak-to-peak difference between every 700 consecutive measurements (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 the figure reveal three interesting insights: the average peak-to-peak difference for the angle of the bulb is: (1) very small 2-35 millidegrees, (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 [6], 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. As seen in the figure, the sound caused movements of 17-55 microns.

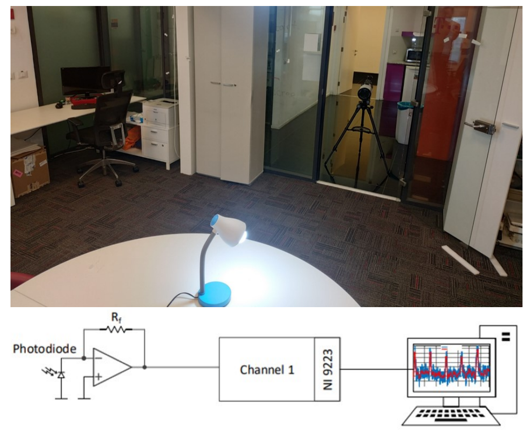

Figure 5: Experimental setup: the telescope and the E14 bulb used in the experiments. A PDA100A2 electro-optical sensor [5] is mounted on the telescope. The electro-optical sensor outputs voltage that is sampled via an ADC (NI-9234) [3] and processed in LabVIEW.

4.1.2 Capturing the Optical Changes

We now explain how eavesdroppers can determine the sensitivity of the equipment needed to recover sound based on a bulb's vibration. The graphs presented in Fig. 3 establish the 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 resulting from a bulb's vibrations of 17-55 microns.

In order to demonstrate how eavesdroppers can determine the sensitivity of the equipment needed to satisfy the above-mentioned criterion, we conduct another experiment.

Experimental Setup: We directed a telescope (with a lens diameter of 25 cm) at a 12W E14 LED bulb (as can be seen in Fig. 5). We mounted an electro-optical sensor (the Thorlabs PDA100A2 [5], which is an amplified switchable gain light sensor that consists of a photodiode) to the telescope. The voltage was obtained from the electro-optical sensor using a

Table 1: Expected voltage for each frequency (based on the linear equations calculated from Fig. 6 and the expected movement from Fig. 4). Green cells can be detected by the sensitivity provided by the ADC while yellow cells cannot.

		Expected voltage change for E14 light bulb									
		ADC with sensitivity of 0.6 µV					ADC with sensitivity of 4 µV				
Distance	Linear equation	150 Hz	200 Hz	250 Hz	300 Hz	350 Hz	150 Hz	200 Hz	250 Hz	300 Hz	350 Hz
1m - 2m	y = -0.59x + 2.56	15.9 µV	15.3 µV	14.8 µV	32.5 µV	17.7 µV	15.9 µV	15.3 µV	14.8 µV	32.5 µV	2.95 µV
2m - 3m	y = -0.52x + 2.41	14 µV	13.5 µV	13 µV	28.6 µV	15.6 µV	14 µV	13.5 µV	13 µV	28.6 µµV	15.6 µV
3m - 4m	y = -0.14x + 1.27	3.78 µV	3.64 µV	3.5 µV	7.7 µV	4.2 µV	3.78 µV	3.64 µV	3.5 µV	7.7 µV	4.2 µV
4m - 6m	y = -0.136+ 1.24	3.67 µV	3.54 µV	3.4 µV	7.48 µV	4.08 µV	3.67 µV	3.54 µV	3.4 µV	7.48 µV	4.08 µV
6m - 7m	y = -0.12x + 1.14	3.24 µV	3.12 µV	3 µV	6.6 µV	3.6 µV	3.24 µV	3.12 µV	3 µV	6.6 µV	3.6 µV
7m - 9m	y = -0.1x + 1.02	2.7 µV	2.6 µV	2.5 µV	5.5 µV	3 µV	2.7 µV	2.6 µV	2.5 µV	5.5 µV	3 µV

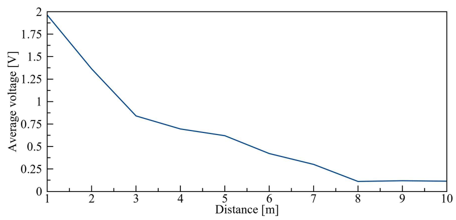

Figure 6: Output obtained from the electro-optical sensor from various distances.

24-bit ADC NI-9234 card [3] and 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 meters) from the light bulb and measured the voltage obtained from the electro-optical sensor for each distance.

Results: The results of this experiment are presented in Fig. 6. These results were used to compute the linear equation between every two consecutive points. Based on the linear equations, we calculated the expected voltage for each expected movement in the 150-350 Hz spectrum for the E14 light bulb for a sound level of 70 dB (based on the results in Fig. 4). The linear equations and the expected voltage for each movement are presented in Table 1.

We now explain how to use the data in Table 1 in order to determine which frequencies can be recovered from the optical measurements obtained for a sound level of 70 dB. The sensitivity of the ADC can be calculated using the equation: $\frac{R}{2^{B}-1}$, where R denotes the dynamic range of the output of the ADC, and B denotes the resolution of the output in bits. For example, 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: $\frac{10}{2^{24}-1} \approx 0.6$ µV.

Analyzing Table 1, we find that an ADC with a sensitivity of 0.6 µV is sufficient for recovering the entire spectrum (150-350 Hz) from a distance of nine meters, because the smallest vibration of the bulb (17 microns) from this distance is expected to yield a difference of 2.7 µV (for a frequency of 150 Hz and a distance of nine meters). Such an ADC can provide the sensitivity required to recover the entire spectrum from any distance. However, given an ADC with a lower sensitivity of 4 µV, only part of the spectrum can be recovered beyond some distance (e.g., beyond three meters), because in the frequency range of 150-250 Hz the vibrations of the bulb are expected to yield values below 4 µV. The green cells in Table 1 indicate frequencies that can be recovered by the sensitivity provided by two ADCs (with sensitivities of 0.6 µV and 4 µV). The yellow cells in Table 1 indicate the frequencies that cannot be recovered by the ADCs in use. As can be seen in the table, the entire measured spectrum can be recovered with: (1) an ADC which provides 0.6 µV sensitivity from all distances, and (2) an ADC which provides 4 µV sensitivity from a maximal distance of three meters. This experiment and these calculations can be used to determine the required ADC given the electro-optical sensor used to capture light from a light bulb.

4.2 Exploring the Optical Response to Sound

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

4.2.1 Characterizing the 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. As can be seen, peaks appears in the FFT at 100 Hz and its harmonics (200 Hz, 300 Hz, etc.). 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 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.

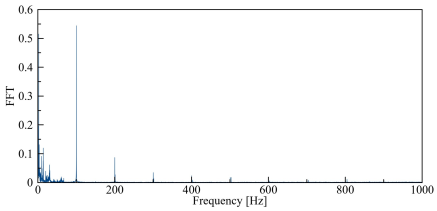

Figure 7: Baseline - FFT of the optical signal in silence (no sound is played).

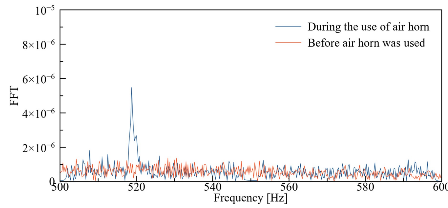

Figure 8: Difference in FFT before and when an air horn was used.

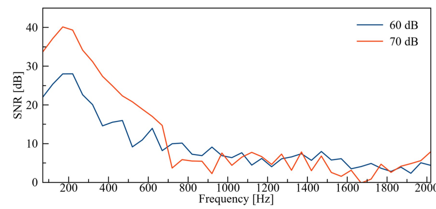

Figure 9: SNR for a desk lamp at 100-2000 Hz.

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. 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 electromagnetic 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 the light bulb in a desk lamp to a wide spectrum of frequencies. These experiments were conducted using speakers that were placed in front of 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 two volume levels (60 dB and 70 dB) and obtained the optical signal via the electro-optical sensor.

Results: Fig. 9 presents the SNR obtained from the desk lamp 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 the 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 by using an equalizer.

4.3 Explaining the Physical Phenomenon

The experiments conducted in this section show that the intensity of the light captured by an electro-optical sensor changes as a function of the distance between the light bulb and the electro-optical sensor (Fig. 6). The changes in the distance between the light bulb and a static electro-optical sensor are caused by sound waves that hit the surface of the bulb and cause the bulb to vibrate (Fig. 4). As a result, the electro-optical sensor outputs voltage levels (which serve as optical measurements) that modulate the sound near the light bulb (Fig. 9) with some additional side effects (Fig. 7).

5 Optical Acoustical Transformation

In this section, we leverage the findings presented in Section 4 and present an optical-acoustic transformation (OAT), which we use to recover audio signals from the optical signals obtained from an electro-optical sensor directed at a light bulb. Throughout this section, we consider $snd(t)$ as the audio played inside the victim's room by the speakers, $opt(t)$ as the optical signal obtained via an electro-optical sensor directed at a tabletop/desk lamp, and $snd^*(t)$ as the audio signal recovered from $opt(t)$ using the OAT. The OAT consists of the following steps:

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. In addition, since the lower range of the speech spectrum is around 100 Hz, we use a high-pass filter (>100 Hz) to remove any phenomenon that is added to the optical signal that is not the result of the sound.

2) Speech Enhancement: Speech enhancement is performed to optimize the speech quality by improving the intelligibility and overall perceptual quality of the speech signal. We enhance the speech signal by normalizing the values of $opt(t)$ to the range of [-1,1].

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 [30]. Spectral subtraction is considered an adaptive technique, i.e., it characterizes the noise from the signal. Adap-

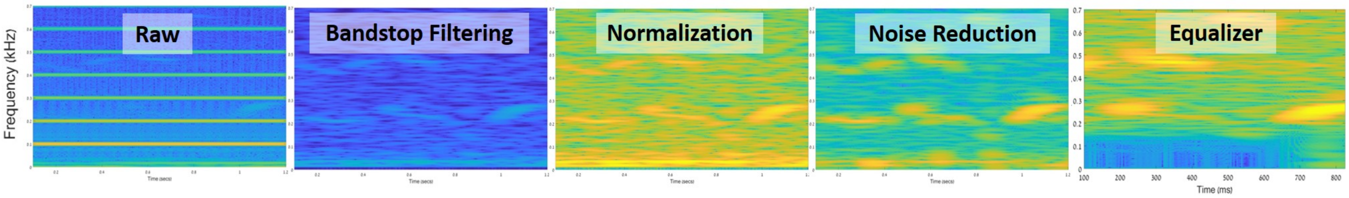

Figure 10: The effect of each step of the OAT in recovering the word "lamb" from an optical signal.

tive techniques are highly effective at removing noise from signals when there is no prior knowledge regarding the distribution of the noise or when the distribution of the noise changes between different setups.

4) Equalizer: Equalization is the process of adjusting the balance between frequency components within an electronic signal. We use an equalizer to amplify the response of weak frequencies.

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. 10. As can be seen, the raw optical signal is very noisy, however the application of each step significantly improves the SNR. In the appendix, we present Algorithm 1, an implementation of the OAT's steps to recover audio from optical measurements.

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; (3) the techniques can be applied in real-time applications, so the optical signal obtained can be converted to audio with minimal delay; and (4) these techniques (e.g., spectral subtraction and equalization) can overcome changes in the optical signal's SNR level which can be caused by different noise levels.

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. We start by examining the influence of environmental factors and various types of bulbs and lamps on the SNR of the recovered sound. We continue by comparing Lamphone's performance to related work in a lab setup. Finally, we examine the influence of the distance between the bulb and the victim and the electro-optical sensor and the bulb on Lamphone's performance.

The reader can assess the quality of the recovered sound visually by analyzing the extracted graphs (spectrograms and SNR), 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 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 the background noise and reverberation [1]. To measure intelligibility, we used the metric suggested by [29], 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 [26]. This metric has been used in speech research for many years to compare speech signals [11]. 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 [21]. 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, which is defined as the logarithmic ratio between the speech power and the noise power estimated over 20 consecutive ms. 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 (25 cm lens diameter) was directed at the light bulbs. We mounted an electro-optical sensor (Thorlabs PDA100A2 [5]) to the telescope. The electro-optical sensor was configured for the highest gain level before saturation. The output of the electro-optical sensor (voltage associated with light intensity) was sampled with two different ADCs. A 24-bit ADC NI-9234 card was used for the experiments presented in Sections 6.1-6.3. A PXI-1082 measurement system (PXI Sound and Vibration Module PXI-4498 with a 24-bit ADC) was used for the experiments presented in Section 6.4, in order to optimize the SNR (its noise level is lower by almost two orders of magnitude compared to the NI-9234). We used Logitech Z200 speakers, which were placed on a dedicated stand, to produce the sound. The data 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

[1] https://youtu.be/kfdXhX8hWok
[2] https://youtu.be/86CDP9QP1Bw

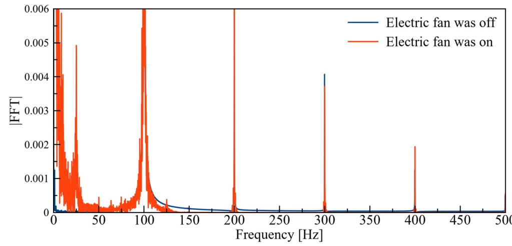

Figure 11: Influence of an electric fan on the spectrum.

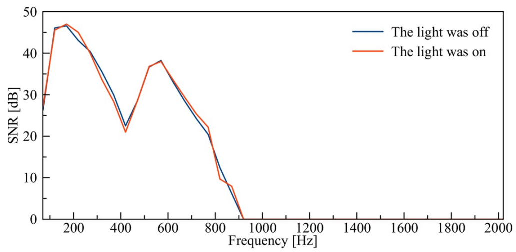

Figure 12: The effect of the light in a room on the SNR.

level of the sound played was measured with a professional decibel meter.

6.1 Influence of Environmental Conditions

6.1.1 Effect of Fan/Air Conditioner

First, we explore the effect of air produced by an external device (e.g., a fan, air conditioner) in proximity to the desk lamp. Such devices may cause unwanted vibrations which may affect the quality of the recovered signals.

Experimental Setup: We directed an electric fan at the desk lamp light bulb from a distance of one meter. We obtained two optical signals via the electro-optical sensor: (1) a baseline signal when the electric fan was off, and (2) an additional signal when the electrical fan was on.

Results: Fig. 11 presents the SNR calculated from the two optical signals. As can be seen, air produced by the electric fan affects the spectrum below 120 Hz.

Conclusions: The range below 120 Hz is filtered in the first step of the OAT using a high-pass filter. Therefore, despite the effect of the electric fan on the spectrum, it has a negligible effect on the quality of the recovered speech, because most of the speech energy is present above 100 Hz.

6.1.2 Effect of Ambient Light

Next, we explore the effect of ambient light on the SNR.

Experimental Setup: We played a frequency scan via speakers that were placed 10 centimeters from the desk lamp light bulb. We obtained two optical signals via the electro-optical sensor from a distance of five meters: when the light in the room was off (darkness) and on.

Results: Fig. 12 presents the SNR calculated from the two optical signals. As can be seen, the SNR graphs are almost identical in their behavior and quality.

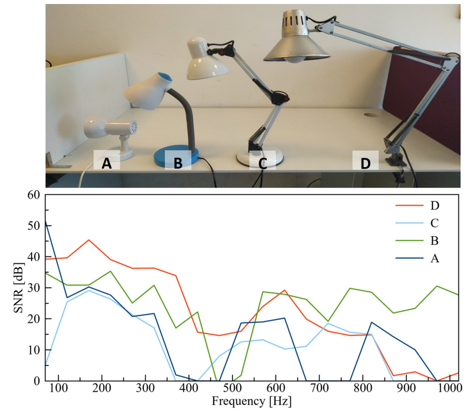

Figure 13: SNR of different types of desk lamps.

Conclusions: Based on this experiment, we concluded that ambient light does not affect the SNR of the recovered sound.

6.2 Influence of Lamps and Bulbs

6.2.1 Effect of the Lamp

Here we explore the effect of the type of lamp used on the SNR.

Experimental Setup: We directed an electro-optical sensor at a desk lamp through a telescope located five meters away. We placed the speakers 10 centimeters away from the lamp and played a frequency scan from the speakers (70 Hz-1000 Hz). We obtained optical signals during the frequency scan. We repeated this experiment four times, each time using a different type of desk lamp (presented in Fig. 13). The same bulb was used in each of the lamps: a 12W E27 light bulb.

Results: Fig. 13 presents the SNR calculated from the optical signals. As can be seen from the results, the behavior of the SNR graphs is similar for the four desk lamps examined: the SNR decreases as a function of the frequency. However, the SNR level of each of the lamps differs (as does the effective bandwidth that can be used to recover sound). For example, a lamp with a long swing arm (lamp D) produces higher SNR values than lamps with short swing arms (lamps B and C). In addition, the fixed desk lamp (lamp A) produces significantly lower SNR values than the adjustable table lamps. Based on this experiment, we concluded that all of the lamps can be used to recover sound, however the type of lamp used affects the quality of the recovered sound.

6.2.2 Effect of the Light Bulb

Here we explore the effect of the type of light bulb used on the SNR.

Experimental Setup: We repeated the previous experiment, this time using a fixed desk lamp, playing a frequency scan and obtaining optical measurements from four different types of E27 light bulbs: Incandescent 40W (31 grams), Leelite 15W LED (67 grams), Leelite 19W LED (86 grams), and S-10A60 15W LED (39 grams). In addition, we obtained optical

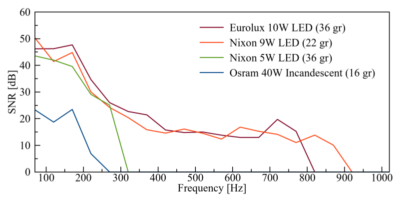

Figure 14: Comparison of SNR between different E14 bulbs.

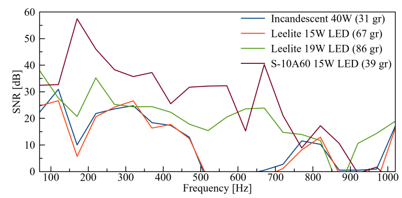

Figure 15: Comparison of SNR between different E27 bulbs.

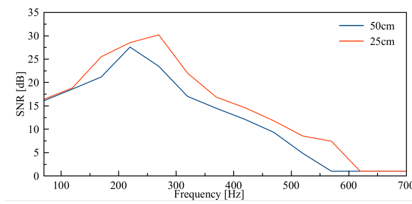

Figure 16: SNR for various distances between the light bulb and the speakers.

Table 2: Comparison of the results of the visual microphone (VM) [12] and Lamphone for the sound recovery of speech. A green cell indicates a better result. The spectrograms associated with these results can be seen in Figs. 22 and 23 in the appendix.

		Intelligibility		LLR		WSS		NIST-STNR	
	Speech	VM	Lamphone	VM	Lamphone	VM	Lamphone	VM	Lamphone
Female speaker - fadg0, sa1	"She had your dark suit in greasy wash water all year"	0.72	0.72	1.47	1.79	120.29	75.55	26.8	16
Female speaker - fadg0, sa2	"Don't ask me to carry an oily rag like that"	0.65	0.67	1.37	2.1	197.83	71.76	43.3	4.5
Male speaker - mccs0, sa1	"She had your dark suit in greasy wash water all year"	0.59	0.7	1.31	1.72	149.55	63.1	27.3	10.3
Male speaker - mccs0, sa2	"Don't ask me to carry an oily rag like that"	0.67	0.71	1.55	1.86	137.04	59.23	18	2.8
Male speaker - mabw0, sa1	"She had your dark suit in greasy wash water all year"	0.77	0.67	1.68	1.48	211.11	54.97	6	5.5
Male speaker - mabw0, sa2	"Don't ask me to carry an oily rag like that"	0.72	0.69	1.81	1.89	162.11	73.77	25.8	5.3
	Average	0.68	0.69	1.53	1.8	162.98	66.39	24.53	7.4
	STD	0.06	0.02	0.19	0.2	35.23	8.48	12.27	4.89

Table 3: Comparison of the NIST-SNR results of the Hard Drive of Hearing [18] and Lamphone for the sound recovery of speech. A green cell indicates a better result.

	Hard Drive of Hearing	Lamphone
Male (List 57)	11.2	22.65
Female (List 1)	7.8	23.35
Average	9.5	23
STD	2.4	0.49

measurements from four different types of E14 light bulbs: Eurolux 10W LED (36 grams), Nixon 9W LED (22 grams), Nixon 5W LED (36 grams), and Osram 40W Incandescent (16 grams).

Results: Figs. 14 and 15 present the SNR calculated from the eight optical signals. As can be seen from the results, the bulbs that produced higher SNR values were the more powerful bulbs (higher wattage), and LED bulbs produced higher SNR values than other types of bulbs. In addition, lighter LED bulbs produced higher SNR values than heavier ones with the same power level.

6.3 Comparing Lamphone to Related Work

6.3.1 Comparing Lamphone to the Visual Microphone

Here, we compare the performance of Lamphone to that of the visual microphone [12]. The authors proposing the visual microphone demonstrated the recovery of six sentences from the TIMIT repository [13] 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 a 12W E14 desk lamp light bulb.

Experimental Setup: We followed the experimental setup used in the visual microphone study [12] as follows: We placed speakers on a dedicated stand (so their vibrations would not affect the bulb) at the same distance that the bag of chips was placed in the visual microphone study (5 cm). We played the same six sentences from the TIMIT repository recovered by the visual microphone via the speakers at the same volume level used in the visual microphone study (95 dB). We placed the eavesdropping equipment 2.5 meters from the light bulb, behind a closed door (the same distance that the video camera was placed in the visual microphone study). Our experimental setup is presented in Fig. 5.

Results: We used the OAT to recover speech from the optical measurements. 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. 22 and 23 in the appendix. We evaluated the intelligibility, LLR, WSS, and NIST-SNR of the recovered signals and reported the results in Table 2. We also downloaded the same six audio signals that were recovered and published in the study presenting the visual microphone, and evaluated their performance according to the same metrics. The results presented in Table 2 reveal four interesting insights: (1) The range of the intelligibility of the speech recovered by Lamphone is 0.67-0.72, which is considered good/fair according to [1]. The average intelligibility of the speech recovered by Lamphone is 0.1 higher (better) than the average intelligibility of the speech recovered when using the visual microphone. In addition, the standard deviation (STD) of the intelligibility obtained by Lamphone is 0.04 lower (better) than the STD of the intelligibility obtained by the visual microphone. This indicates that the results obtained by Lamphone are of a higher quality and more stable in terms of intelligibility. (2) The average LLR of the speech recovered by the visual microphone is 0.27 lower (better) than that obtained when using Lamphone. In addition, the STD of the LLR obtained by the visual microphone is 0.01 lower (better) than the STD of the LLR obtained by Lamphone. This indicates that the results obtained by the visual microphone are of a higher quality and more stable in terms of the LLR. (3) The average WSS of the speech recovered using Lamphone is 96.59 lower (better) than the speech recovered by the visual microphone. In addition, the STD of the WSS obtained by Lamphone is 26.75 lower (better) than the STD of the WSS obtained by the visual microphone. This indicates that the results obtained by Lamphone are of a higher quality and more stable in terms of the WSS. (4) The average NIST-SNR of the speech recovered by the visual microphone is 17.1 higher (better) than the average NIST-SNR of the recovered speech when using Lamphone. However, the STD of the NIST-SNR obtained by Lamphone is 7.3 lower (better) than the STD of the NIST-SNR obtained by the visual microphone. This indicates that the results obtained by Lamphone are more stable but of a lower quality in terms of the NIST-SNR.

6.3.2 Comparing Lamphone to the Hard Drive of Hearing

Here, we compare the performance of Lamphone to that of the Hard Drive of Hearing [18]. The authors proposing the Hard Drive of Hearing demonstrated the recovery of two recordings from the Harvard sentences database: a female sample (list 1) and a male sample (list 57). The specific audio samples are taken from the Open Speech Repository [4]. We compare Lamphone's performance when recovering the same sentences by analyzing the vibrations of a 12W E14 desk lamp light bulb.

Experimental Setup: We followed the experimental setup used in the Hard Drive of Hearing [18] study as follows: We placed speakers on a dedicated stand (so their vibrations would not affect the bulb) the same distance as in the Hard Drive of Hearing study (25 cm). We played the two audio samples from the Open Speech Repository recovered by the Hard Drive of Hearing via the speakers at the same volume level used in the Hard Drive of Hearing study (85 dB). In our experiment, the eavesdropping equipment was placed 2.5 meters from the light bulb, behind a closed door. Our experimental setup is presented in Fig. 5.

Results: We used the OAT to recover speech from the optical measurements. Since we were unable to obtain the recovered audio samples from the Hard Drive of Hearing study, we compared Lamphone's performance to the results reported in their paper. The authors of the Hard Drive of Hearing study evaluated their recovered signals using the NIST-SNR, so we compare the sentences recovered by Lamphone and the Hard Drive of Hearing based on the NIST-SNR. The results of our comparison are presented in Table 3. The average NIST-SNR of the speech recovered by the Lamphone is 13.5 higher (better) than the average NIST-SNR of the recovered speech

Figure 17: Experimental setup: An electro-optical sensor is mounted to a telescope (left picture) that is directed at a desk lamp light bulb (placed 50 cm from the speakers) from 35 meters away (right picture).

reported in the Hard Drive of Hearing paper. Moreover, the STD of the NIST-SNR obtained by Lamphone is 1.91 lower (better) than the STD of the NIST-SNR obtained by the visual microphone.

6.3.3 Conclusions

Analyzing the results of the experiments conducted in this section, we concluded that: (1) that the quality of the speech recovered 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. (2) The quality of the signals recovered by Lamphone is better than the quality of the signals recovered by the Hard Drive of Hearing.

6.4 The Influence of Distance on Lamphone's Performance

Here, we evaluate the influence of the distance between the light bulb and: (1) the eavesdropping equipment (as the distance increases, the sensor captures less light), and (2) the victim (as the distance increases, the magnitude of the light bulb vibrations decreases) on the SNR. The setup can be seen in Fig. 17.

First, we tried to assess the influence of the distance between the victim and the bulb on the spectrum of the recovered signal.

Experimental Setup: The eavesdropping equipment was located 10 meters away from the light bulb. We placed the speakers at two distances (25 cm and 50 cm) away from the light bulb. We played a frequency scan via the speakers at the volume level of a virtual meeting (such meetings have an average volume level of 75 dB) while obtaining the optical measurements.

Results: Fig. 16 presents the SNR obtained by placing the speakers at distances of 25 cm and 50 cm from the light bulb. Based on the results, we concluded that the effective bandwidth containing a strong signal that can be used to recover sound (>15dB) is narrow: it ends at around 340 Hz for a distance of 50 cm and at around 400 Hz for a distance of 25 cm.

Table 4: "We Will Make America Great Again!" - Results of recovered speech from various distances.

Distance between speakers and light bulb	Distance between electro-optical sensor and light bulb											
	15m				25m				35m			
	Intelligibility	WSS	LLR	NIST-SNR	Intelligibility	WSS	LLR	NIST-SNR	Intelligibility	WSS	LLR	NIST-SNR
25 cm	0.522	660.47	6.72	21	0.487	643.1	5.11	21.8	0.454	686.65	4.22	17.5
50 cm	0.457	669.37	3.5	12.3	0.4	651.16	4.34	21	0.362	635.9	3.75	11.5

Next, we tried to assess the influence of distance on the speech recovered by Lamphone.

Experimental Setup: We placed the eavesdropping equipment at three distances (15, 25, 35 meters) away from the light bulb, and we placed the speakers at two distances (25 cm and 50 cm) away from the light bulb Then we played a famous statement made by former President Donald Trump: "We will make America great again!" via speakers at the volume level of a virtual meeting while obtaining the optical measurements.

Results: We used the OAT to recover speech from the optical measurements. The recovered audio signals are available online[2]. The spectrograms of the recovered speech can be seen in Figs. 24 and 25 in the appendix. The intelligibility, LLR, WSS, and NIST-SNR of the recovered signals are reported in Table 4. As can be seen from the results, the quality of Lamphone is considered fair when it is used to recover sound from a distance of: (1) up to 35 meters when the lamp is located 25 cm from the victim, and (2) up to 15 meters when the lamp is located 50 cm from the victim.

7 Potential Improvements

In this section, we suggest methods that eavesdroppers can use to optimize the quality of the recovered audio 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 captured by an electro-optical sensor mounted to a telescope with diameter r is determined by the area of the telescope's lens (πr^2). As a result, there is a quadratic difference in the amount of light captured by two telescopes with lens diameters of x and y (where $x = y + z$) = (πx^2) - (πy^2) = $2\pi yz$ + πz^2. This is empirically demonstrated in Fig. 18, which presents three SNR graphs extracted from optical measurements obtained by three telescopes with different lens diameters (7.5 cm, 10 cm, 20 cm) using the same experimental setup. As can be seen from the results, as the lens diameter increases, the amount of light captured by the mounted electro-optical sensor also increases. As a result, the optical signal yields a higher SNR and has a wider effective spectrum.

Lens: A close-up lens can be used as an optical amplifier for the optical signal by increasing the ratio of the area of the lamp's light (and thus the amount of light captured by the telescope) to the total area captured by the telescope. This is empirically demonstrated in Figure 19, which presents two SNR graphs extracted from optical measurements obtained from a telescope in two setups: (1) with a close-up lens placed between the telescope and the electro-optical sensor, and (2) without a close-up lens. As can be seen from the results, the increased ratio of the area of the lamp (which resulted in an increased light ratio) to the total area captured by the telescope yields a higher SNR and a wider effective spectrum.

Electro-Optical Sensor: The sensitivity of the system can be enhanced by using an improved electro-optical sensor to obtain the optical measurements. This is empirically demonstrated in Figure 20, which presents two SNR graphs extracted from optical measurements that were obtained from two different optical sensors (PDA100A2 and APD410A). As can be seen from the results, the sensitivity of the electro-optical sensor affects the SNR level. Another option is to sample the signal from multiple optical 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 directing several electro-optical sensors at the light bulb in order to obtain multiple measurements and sample the bulb's vibrations simultaneously from several channels.

ADC: As discussed in Section 4, a 24-bit ADC with an input range of [-5,5] voltage provides a sensitivity of 0.6 uV. 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 significantly optimizes the system's sensitivity. In addition, in this research we experimented with two types of 24-bit ADCs (PXI-4498 and NI-9234). The ADC with the lower self-noise level provided a higher SNR which yielded signals of a higher quality.

OAT: Many advanced denoising methods in the field of speech enhancement can be used in addition to or as an alternative to the OAT steps. 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 such models. The effective bandwidth of the recovered speech signal can be extended by using artificial bandwidth extension algorithms for speech [16,17,20,24,25]). Such algorithms use a dedicated model for speech and artificially add information to the high frequencies of the audio signal according to the information that appears in the lower frequencies. By doing so, bandwidth extension algorithms extend the effective bandwidth that can be heard by the human ear, which improves the quality of the audio.

Conclusions: The experiments conducted in this section

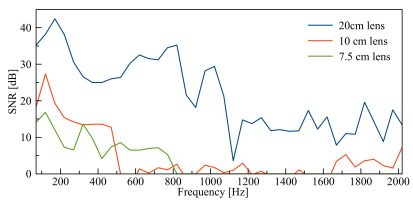

Figure 18: Impact of a telescope's lens diameter on the SNR.

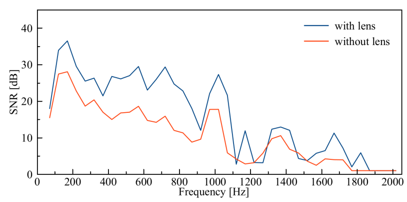

Figure 19: Impact of a close-up lens on the SNR.

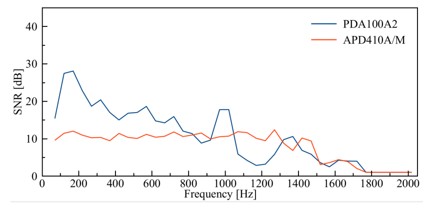

Figure 20: SNR obtained by two different electro-optical sensors.

reveal than the use of improved equipment yields a higher SNR. We concluded that the quality of the sound recovered by Lamphone is proportional to the quality of the equipment used. Improved equipment can be used to extend the possible distance between both the lamp and the victim and the lamp and the electro-optical sensor.

8 Countermeasures

In this section, we analyze the effectiveness of known countermeasures against sound recovery and compare their effectiveness against Lamphone and methods proposed in other studies. The countermeasures are analyzed according to their effectiveness: high (denoted as ●), medium (denoted as ◐), and low (denoted as ○). A summary of this comparison is presented in Table 5.

8.1 Prevention

Countermeasures used to eliminate the vulnerability. The countermeasure: (●) - is effective at preventing objects from turning into diaphragms, (○) - is ineffective against the methods.

Removing potential diaphragms from offices - Banning any lightweight object/device (e.g., a bag of chips, smartphones) that vibrates when it is hit by sound waves from the environment (●- against the vast majority of the methods [7, 8, 12, 14, 15, 18, 22, 27, 28, 31, 32, 35], including Lamphone), however laser microphones can recover sound from the vibration of window panes which cannot be removed from most environments (○- against the laser microphone [23]).

8.2 Mitigation

Countermeasures used to prevent the exploitation of the vulnerability. The countermeasure: (●) - prevents eavesdroppers from exploiting the vulnerability (i.e., recovering sound), (◐) - reduces the likelihood of exploiting the vulnerability (e.g., lowers the quality of the recovered sound or requires eavesdroppers to use more expensive equipment), (○) - is ineffective against the methods.

Limiting the sale of spying equipment - Limiting the availability of equipment associated with spying (e.g., laser transceiver) to specific entities (e.g., police departments). This method prevents eavesdroppers from acquiring the equipment needed to obtain data that can be used to recover sound (●- against laser microphone [23]). However, many sensors/devices that are not associated with spying were also found effective at obtaining data that can be used to recover sound (○- [7, 8, 12, 14, 15, 18, 22, 27, 28, 31, 32, 35], including Lamphone).

Preventing leakage - Containing the data or physical side effect inside the room and preventing the leakage of information from the room using software (e.g., a firewall used to prevent the exfiltration of data) or a physical apparatus (e.g., a curtain used to eliminate the line of sight to a vibrating object), or by changing the location of the targeted room (e.g., targeting an inner room with no windows). Such countermeasures are highly effective against methods that recover sound by obtaining data over the Internet and via optical sensors (●- against [7, 8, 12, 14, 15, 18, 22, 23, 27, 28, 35], including Lamphone). However, such countermeasures are ineffective against methods that recover sound from RF signals, because it is difficult to prevent the leakage of RF signals (○- against [31, 32]).

Creating a safety perimeter - Limiting/decreasing the ability of eavesdroppers to recover sound by forcing the eavesdropper to apply the attack from a distant location (e.g., by installing a fence around a house or by moving the surface on which the desk lamp is placed away from windows). Such countermeasures can limit eavesdroppers' ability to obtain the RF and optical data required to recover sound. However, eavesdroppers can use improved equipment (e.g., antennas, telescopes) to obtain radio signals and create a line of sight (◐- [12, 23, 31, 32], including Lamphone). Such countermeasures are ineffective against methods that recover sound by obtaining data over the Internet (○- against [7, 8, 14, 15, 18, 22, 27, 28, 35]).

8.3 Detection

Countermeasures used to identify the application of methods used to exploit the vulnerability. The countermeasure: (●) - can detect the application of a method to recover sound, (○) - is ineffective against the methods.

Detecting active attacks - Deploying software used to detect attempts to compromise a network and exfiltrate data (e.g., by using an IDS) and deploying a device used to detect the use of a laser transceiver (e.g., by using an optical sensor). Such methods are effective at detecting the preliminary stage of compromising a device in order to obtain data from a

Table 5: Effectiveness (high (●), medium (◐), low (○)) of countermeasures against sound eavesdropping methods.

	Prevention	Mitigation			Detection
	Removing potential diaphragms from offices	Limiting the sale of spying equipment	Preventing leakage	Creating a safety perimeter	Detecting the use of active attacks
Methods relying on a compromised device [7, 8, 14, 15, 18, 22, 27, 28, 35]	●	○	●	○	●
RF-based methods [31, 32]	●	○	○	◐	○
Visual microphone [12]	●	○	●	◐	○
Laser microphone [23]	○	●	●	◐	●
Lamphone	●	○	●	◐	○

target device (●- [7, 8, 14, 15, 18, 22, 27, 28, 35]) and the use of the laser microphone (●- [23]). However, this method is ineffective against passive external methods (○- [12, 31, 32], including Lamphone).

8.4 Conclusions

By analyzing the effectiveness of the countermeasure methods described above and presented in Table 5, we concluded that victims cannot rely on regulations limiting the sale of equipment needed for the Lamphone attack to prevent eavesdroppers from recovering sound, since Lamphone relies on a photodiode (a commonly used sensor that is not associated with spying), unlike the laser microphone which relies on a laser transceiver. In addition, victims cannot install a dedicated mechanism to detect the use of Lamphone, because its implementation does not provide any indication regarding its application, unlike the laser microphone where a dedicated optical sensor can be used to detect its use. Given this, potential victims should take care to protect themselves from the threat posed by the Lamphone attack. For example, simple but effective adjustments to a home office could be made, including installing curtains to eliminate the line of sight or replacing desk lamps with other lighting solutions to eliminate the vulnerability. Alternatively, an inner room without windows could be used for sensitive conversations, or a homeowner could install a fence to increase his/her distance from potential eavesdroppers and mitigate the attack.

9 Limitations, Discussion, and Future Work

The purpose of this research was to raise awareness of the feasibility of recovering sound by analyzing the vibrations of a desk lamp light bulb. 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 light bulb to recover sound: (1) desk lamps and tabletop lamps have existed for many years, and (2) sound recovery is of interest to various entities (NSA, FBI, etc.) around the world. In addition, the case of the "Great Seal Bug" [9] showed that a new technology, the RFID, was used secretly by Soviet agencies to eavesdrop sound three decades before it was scientifically discovered and published in 1973 [19].

As was indicated in Section 7, Lamphone's primary disadvantage is the fact that the quality of the sound recovered is proportional to the quality of the equipment used by the eavesdropper. As a result, better equipment (e.g., an ADC with a lower self-noise level, a more sensitive electro-optical sensor, a telescope with a wider lens, and dedicated optical lenses) is required to recover sound from longer distances between the eavesdropper and the light bulb, and the victim and the light bulb.

We believe that over the next few years, new studies will improve Lamphone, so it will pose a greater threat to individuals' privacy, improving the method so that it could be applied by eavesdroppers with less resources and from a greater distance. An important question to consider is how long will it take the scientific community to improve this method sufficiently so it poses a major 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 [22] was first introduced at USENIX 2014. The main disadvantage of Gyrophone at that time was the low accuracy of the model used to classify isolated words (the accuracy was only slightly better than a random guess). However, over the years, additional research led to improved understanding of this eavesdropping technique and the threat model [7, 15, 35], 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 [8]. Based on the progress made in the seven years since Gyrophone was first introduced, we believe that future studies will improve our understanding of Lamphone and suggest new ways of recovering sound from light bulbs. Improved understanding regarding optical sound recovery could extend Lamphone's scope to hanging light bulbs located farther away from the victim (as opposed to a desk lamp light bulb).

For future work, we suggest investigating how the OAT can be improved by integrating advanced algorithms for speech processing (e.g., [16, 17, 20, 24, 25]) and denoising (e.g., the use of autoencoders), and how to apply the attack by using more compact equipment. We also suggest investigating the accuracy of a light to text model by training a neural network that receives optical signals and outputs transcription/text.

References

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

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

[3] Ni 9234 datasheet.

[4] The open speech repository. http://www.voiptroubleshooter.com/open_speech/american.html.

[5] Pda100a2.

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

[7] 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.

[8] 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.

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

[10] Jieun Choi, Hae-Yong Yang, and Dong-Ho Cho. Tempest comeback: A realistic audio eavesdropping threat on mixed-signal socs. In *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security*, CCS '20, page 1085–1101, New York, NY, USA, 2020. Association for Computing Machinery.

[11] 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.

[12] Abe Davis, Michael Rubinstein, Neal Wadhwa, Gautham J Mysore, Fredo Durand, and William T Freeman. The visual microphone: Passive recovery of sound from video. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 33(4):79:1–79:10, 2014.

[13] 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.

[14] 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.

[15] 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.

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

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

[18] 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.

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

[20] 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.

[21] 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.

[22] 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.

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

[24] 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.

[25] 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.

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

[27] 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.

[28] Sriram Sami, Yimin Dai, Sean Rui Xiang Tan, Nirupam Roy, and Jun Han. Spying with your robot vacuum cleaner: Eavesdropping via lidar sensors. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*, SenSys '20, page 354–367, New York, NY, USA, 2020. Association for Computing Machinery.

[29] 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.

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

[31] 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.

[32] 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.

[33] 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.

[34] 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.

[35] 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

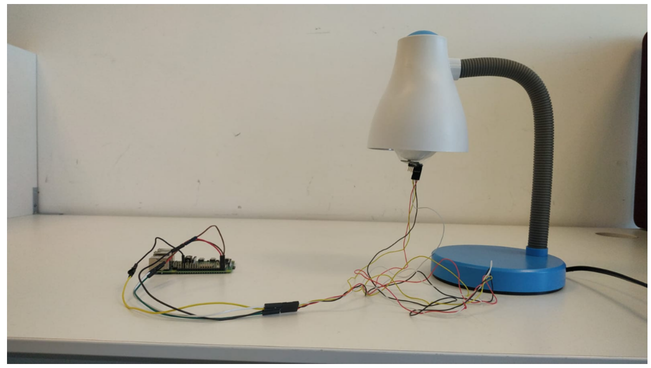

Figure 21: A gyroscope attached to the bottom of a desk lamp light bulb.

Algorithm 1 Recovering Audio from Optical Signal

INPUT: optical-stream, fs, equalizer-function
bulbFs = 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 = bulbFs; i < fs/2; i+=bulbFs) **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 + $\frac{(snd*[i]-min)*2}{max-min}$
    /*Noise reduction*/
    snd* = spectral-subtraction(snd*)
    /*Balancing*/
    snd* = equalizer(snd*,equalizer-function)
    play (snd*)

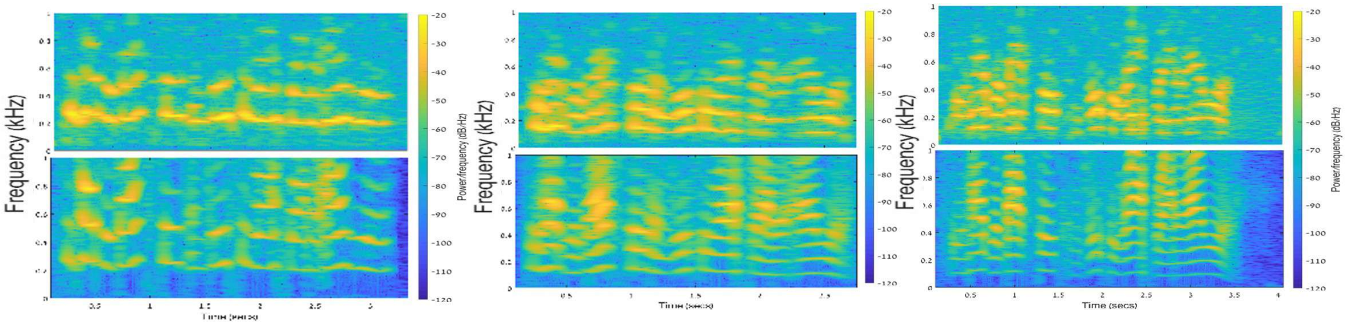

Figure 22: fadg0,mccs0,mabw0 sa1: "She had your dark suit in greasy wash water all year." - recovered (top) and original (bottom) speech.

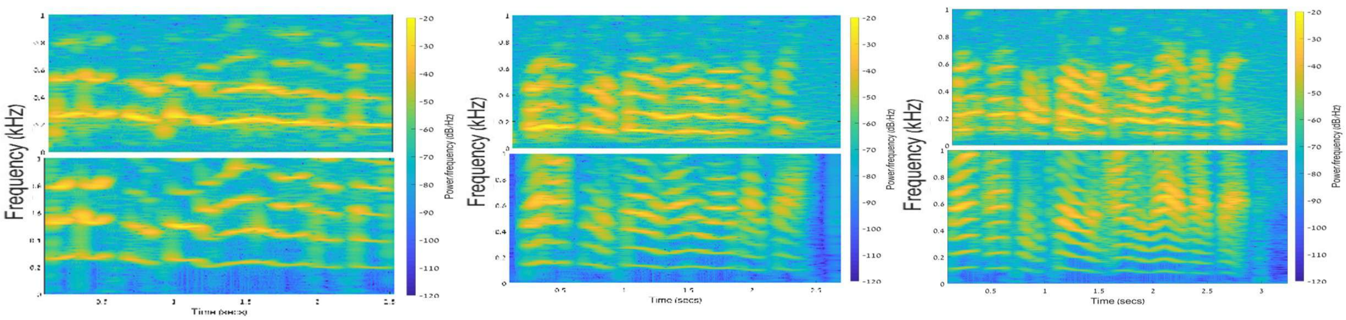

Figure 23: fadg0,mccs0,mabw0 sa2: "Don’t ask me to carry an oily rag like that."- recovered (top) and original (bottom) speech.

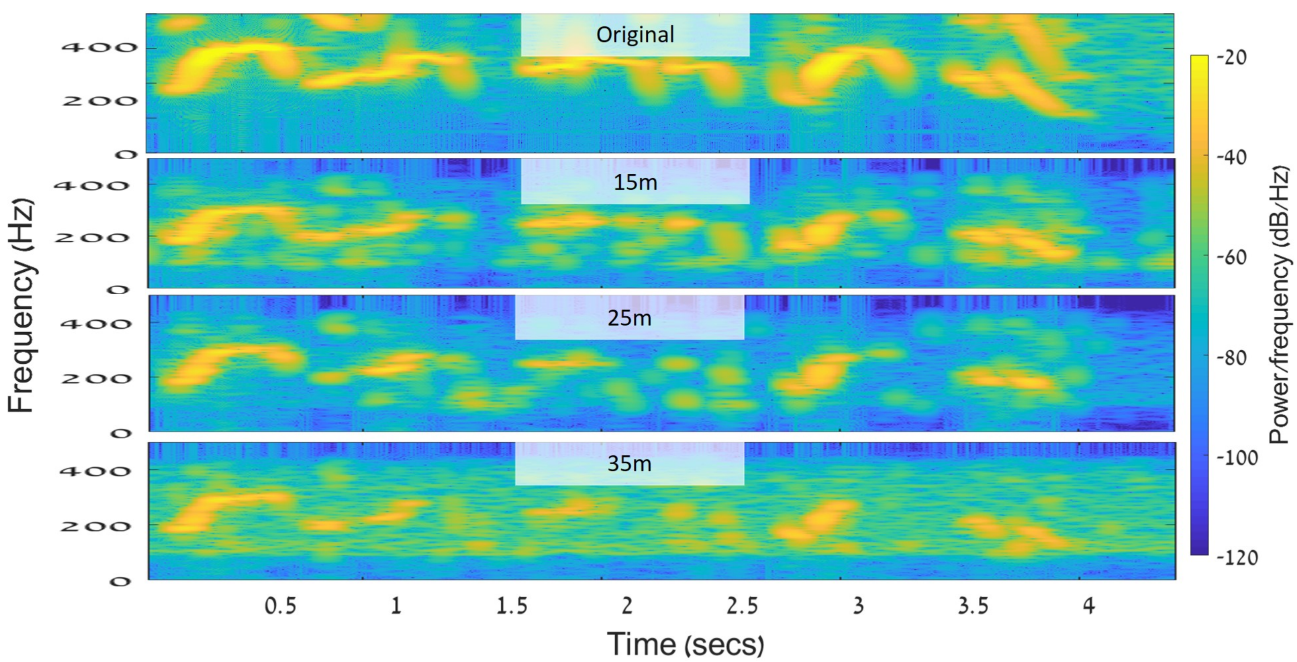

Figure 24: "We will make America great again!" - original (top) and recovered (bottom). The speakers are located 25 cm from the light bulb.

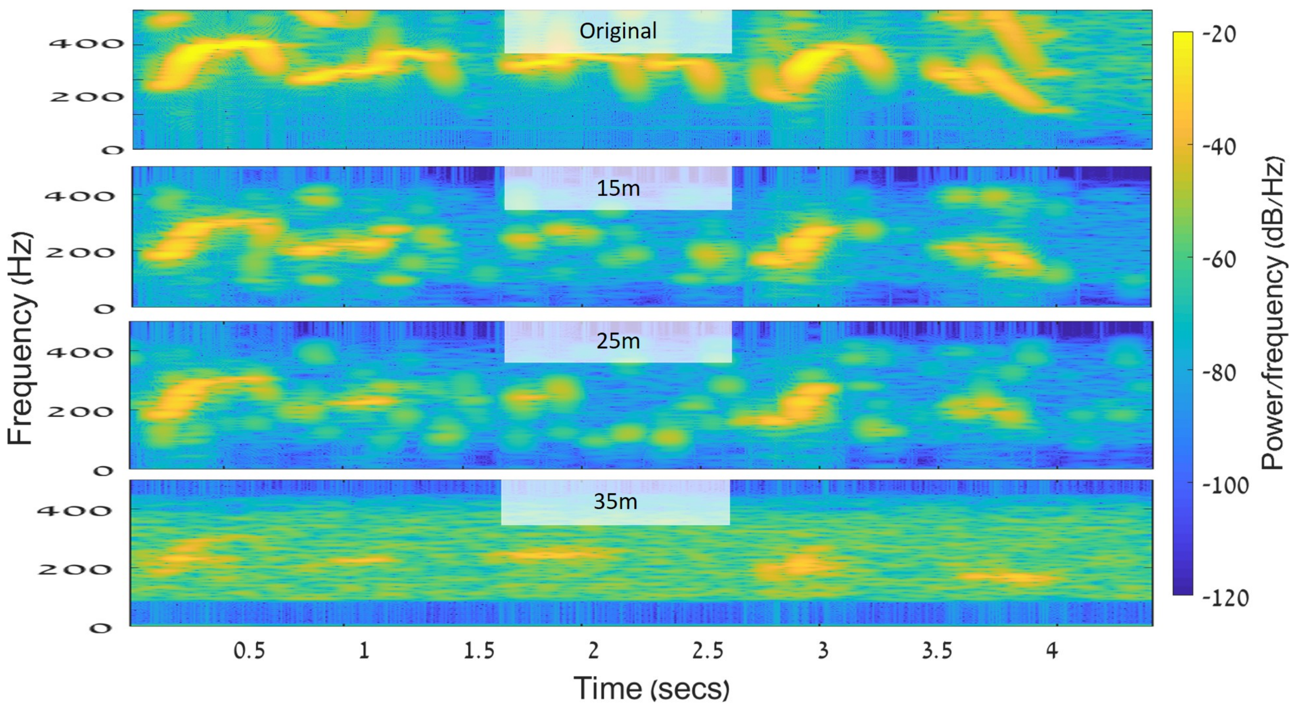

Figure 25: "We will make America great again!" - original (top) and recovered (bottom). The speakers are located 50 cm from the light bulb.