The Little Seal Bug: Optical Sound Recovery from Lightweight Reflective Objects

Ben Nassi[1, 2], Raz Swissa[1], Jacob Shams[1], Boris Zadov[1], and Yuval Elovici[1]

[1]Ben-Gurion University of the Negev, Beersheba, Israel
[2]Cornell Tech, New-York, USA
{nassib, razsw, jacobsh, zadov, elovici}@post.bgu.ac.il, bn267@cornell.edu

Abstract*—In recent years, various studies have demonstrated methods to recover sound/speech with an optical sensor. Fortunately, each of these methods possess drawbacks limiting their utility (e.g., limited to recovering sounds at high volumes, utilize a sensor indicating their use, rely on objects not commonly found in offices, require preliminary data collection, etc.). One unaddressed method of recovering speech optically is via observing lightweight reflective objects (e.g., iced coffee can, smartphone stand, desk ornament) with a photodiode, an optical sensor used to convert photons to electricity. In this paper, we present the 'little seal bug' attack, an optical side-channel attack which exploits fluctuations in air pressure on the surface of a shiny object occurring in response to sound, to recover speech optically and passively using a photodiode. These air pressure fluctuations cause the shiny object to vibrate and reflect light modulated by the nearby sound; as a result, these objects can be used by eavesdroppers (e.g., private investigator, surveilling spouse) to recover the content of a victim's conversation when the victim is near such objects. We show how to determine the sensitivity specifications of the optical equipment (photodiode, ADC, etc.) needed to recover the minuscule vibrations of lightweight shiny objects caused by the surrounding sound waves. Given the optical measurements obtained from light reflected off shiny objects, we design and utilize an algorithm to isolate the speech contents from the optical measurements. In our evaluation of the 'little seal bug' attack, we compare its performance to that of related methods. We find eavesdroppers can exploit various lightweight shiny objects to optically recover the content of conversations at equal/higher quality than prior methods (fair-excellent intelligibility) while doing so from greater distances (up to 35 meters) and lower speech volumes (75 dB). We conclude that lightweight shiny objects are a potent attack vector for recovering speech optically, and can be harmful to victims being targeted for sensitive information conveyed in a spoken conversation (e.g., in cases of corporate espionage or intimate partner violence/surveillance) when seated at a desk near a lightweight reflective object.

2023 IEEE Security and Privacy Workshops (SPW) | 979-8-3503-1236-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/SPW59333.2023.00032

I. Introduction

'Great Seal Bug' [1], a.k.a., 'the Thing', was the first covert listening device that utilized passive techniques to transmit an audio signal for the purpose of speech eavesdropping.[1] It consisted of a passive device that was concealed inside a gift (a picture of the Great Seal of the United States) which was given by the Soviet Union to the United States Ambassador to the Soviet Union in 1945. The concealed passive device, which is considered a predecessor of radio frequency identification (RFID) technology, became an operative listening device when it was activated by the Soviets who 'illuminated' it using electromagnetic energy from an external source. Since the device was passive and considered quite innovative at the time (eight decades ago), it took the Americans six years to determine its real purpose as a listening device, when it was accidentally found by a British embassy radio operator.

Fig. 1: A photodiode is mounted on a telescope (left) and directed at a shiny object on a desk (right). The photodiode is used to obtain optical measurements from the light reflected from the shiny object's minuscule vibrations induced by sound.

Well-known incidents[2] and various studies [2], [3], [4], [5], [6], [7], [8], [9], [10], [11] published over the years have shed light on the practicality of speech eavesdropping. The incidents and studies showed how far motivated entities are willing to go, in order to recover the content of speech. Moreover, the incidents proved that compromised devices can be used for eavesdropping, via non-acoustic data obtained from: (1) an integrated sensor [2], [3], [4], [5], [6], [7], [9], [10], [11] (e.g., using a smartphone's motion sensor data or a robotic vacuum cleaner's LiDAR data) or (2) emanations from the device [10], [11], [8], [12], [13] (e.g., electromagnetic radiation emitted from a PC's hard disk and earphones or light emitted from speakers).

In order to prevent eavesdroppers from recovering the content of conversations from compromised devices, organizations implement policies aimed at preventing employees and visitors from using their electronic devices on the organization's premises. As a result, eavesdroppers have sought new methods for recovering speech that do not rely on a compromised device, and in recent years, several methods have been pro-

[1] https://en.wikipedia.org/wiki/The_Thing_(listening_device)

[2] https://en.wikipedia.org/wiki/Covert_listening_device

DOI 10.1109/SPW59333.2023.00032

posed (e.g., the visual microphone [14], Lamphone [15], and the laser microphone [16], [17]). While the studies presenting these methods improved understanding of the privacy risks posed by objects located in proximity to potential victims, the proposed methods suffer from at least one of the following disadvantages: (1) some methods are limited to recovering speech at a high volume (>85 dB), limiting their effectiveness at recovering speech from lower volumes; (2) some methods rely on spying equipment, limiting their use in countries that restrict the sale of this equipment; (3) some methods require an active laser beam to be directed at objects located near the target, a fact that increases the likelihood of detection via dedicated sensors, and (4) some methods rely on the presence of objects that are not commonly used in offices today (e.g., a hanging light bulb).

In this paper, we present the 'little seal bug', an optical eavesdropping method aimed at recovering speech from lightweight shiny objects via minuscule vibrations that occur when sound (air pressure) hits such objects' surfaces. We show how eavesdroppers (e.g., private investigator, nosey spouse) can exploit these objects ('little seal bugs') as optical implants when they reflect light, in order to recover the content of conversations by analyzing optical data obtained by a photodiode directed at such objects. To accomplish this, we first analyze the movement of various shiny objects and show that their vibrations can be captured by a photodiode. Based on our findings, we suggest a sound recovery model that recovers speech from light obtained from reflective objects. Finally, we compare the performance of the proposed 'little seal bug' attack to the performance of two state-of-the-art methods (visual microphone [14] and Lamphone [15]). We show that the proposed attack can be used to recover the content of a victim's conversation held when the victim is seated at a desk, with fair intelligibility, from a distance of 35 meters.

In this paper, we make the following contributions: (1) We raise awareness regarding the fact that lightweight shiny objects can be exploited as optical implants for the purpose of recovering speech (hence their name 'little seal bugs'). Such objects, which may be purchased by potential victims for personal use/decoration or received as swag at conferences, are often placed on desks. By virtue of their presence on desks, these objects may behave as diaphragms and vibrate in response to conversations (e.g., virtual meetings and phone calls) that take place at the desk. Moreover, when light is reflected from their surface, it is modulated by vibrations resulting from the speech of the conversation, a fact that can be exploited to recover the content of the conversation by using a remote photodiode. (2) We present a speech recovery method that does not suffer from the disadvantages of existing methods: the 'little seal bug' attack is capable of recovering speech from objects (e.g., a smartphone stand, an empty beverage can, desk ornaments) that are commonly placed on desks (as opposed to Lamphone [15] which relies on the presence of a hanging light bulb) and at lower volume than prior methods (as opposed to the visual microphone [14] which is limited to recovering speech at an average volume level of 95 dB). In addition, while Lamphone utilizes an office lamp/light as a diaphragm and transducer for recovering sound, the 'little seal bug' attack utilizes a lightweight shiny object as a diaphragm, and an external light source as a transducer. We demonstrate that objects which don't produce their own light can be used as a diaphragm to optically recover sound, as long as they are lightweight and reflective, generalizing the attack demonstrated in Lamphone beyond lamps which produce their own light.

The rest of the paper is structured as follows: In Section II, we review existing methods for eavesdropping. In Section III, we present the threat model. In Section IV, we analyze the response of a shiny weight to sound, and in Section V, we describe the steps performed to recover sound from the optical measurements obtained from a shiny object. In Section VI, we evaluate the 'little seal bug' attack's performance on the task of recovering sound from various objects and distances. In Section VII we present countermeasure methods against the 'little seal bug' attack, and we discuss limitations of the 'little seal bug attack' in Section VIII. We discuss the findings of this research in Section IX.

II. Related Work

In this section, we review related work in the field of speech recovery using optical sensors. The first group of optical eavesdropping methods are *active* methods, methods which require a laser beam directed at an observed object in order to obtain the required optical data to recover sound. These methods recover sound using a laser beam directed at an object, where the beam is directed either from a dedicated laser [16], [17], [18], or LiDAR sensor [9]. The laser beam is reflected off of an object, back to a sensor (e.g., a video camera [18], LiDAR [9], laser transceiver [16], [17]) which then converts the beam into an audio signal. This conversion process takes advantage of miniscule object vibrations detected by analyzing the laser's speckle patterns [18] or raw LiDAR returns [9]. Active methods are limited in their practical use for visual eavesdropping since they rely on a laser beam [16], [18], [9], [17], which increases the likelihood of detection (even if the beam is invisible) by using a dedicated optical sensor to detect (1) the frequency/wavelength of the laser, or (2) the sudden appearance of light in a room or on a window. The 'little seal bug' is a passive method, not relying on active beams (e.g., laser), avoiding detection by optical sensors.

The second group of methods are *passive* methods, methods which obtain optical data from passive observation of the targeted objects. These passive observation methods utilize a range of optical sensors, including photodiodes [13], [19], high-speed video cameras [14], and electro-optical sensors [15]. These methods take advantage of optical information produced as a side effect of speech, such as fluctuating power LED intensity correlated to a desktop speaker's power consumption [13] or minuscule vibrations of various office objects, e.g., a water bottle or bag of chips [14] or a hanging light bulb [15]. While these methods demonstrate the ability to recover speech passively, they are limited in one of the following ways: they rely on (1) a very high sound level (over 95 dB, on average) well beyond the sound level of speech and virtual meetings (e.g., [14]); or (2) hanging and desk lamp light

bulbs [15] which emanate light. In addition, some methods: (3) can only be used to recover the content of virtual speakers in meetings, since they rely on leakage from devices (e.g., [13], [19]); or (4) rely on post-processing analytics obtained from a pre-recorded sample (e.g., [14]).

We present the 'little seal bug' attack as an abstraction of Lamphone. In order to capture and process speech, a microphone consists of three parts: (1) a diaphragm, (2) a transducer, and (3) an ADC. Lamphone utilizes an office lamp/light as both a diaphragm and a transducer, and an ADC to convert electrical signals, obtained from an electro-optical sensor, to digital signals. In the 'little seal bug' attack, lightweight shiny objects serve as diaphragms, vibrating according to the speech in the room, while an external light source serves as the transducer, with an ADC to convert electrical signals obtained from the photodiode to digital signals. We demonstrate that objects which don't produce their own light can be used as a diaphragm to recover speech, as long as they are lightweight and reflective (compared to desk lamps/lights, which produce their own light).

III. Threat Model

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

Objective. The 'little seal bug' attack is a method to recover speech using a photodiode mounted on a telescope directed at a lightweight, shiny object (e.g., an empty soda can). Conversations which take place in the vicinity of the object cause the object to vibrate. In order to recover the conversation's contents, the minuscule vibrations are observed by the photodiode, and processed from an optical signal into an acoustic signal, representing the recovered conversation.

Victim. We consider potential victims of the 'little seal bug' attack to be anyone who is exchanging sensitive information in a verbal conversation. A victim could be: (1) a worker at a company, (2) a surveiled spouse [20], or (3) any other person who an attacker has an interest in obtaining their sensitive information.

Setup. We consider the following setup to be susceptible to surveillance using the 'little seal bug' attack; The victim is located in their home/office and is engaged in verbal conversation where sensitive information is being shared. The room the victim is located in contains a shiny, lightweight object/ornament located on a table/desk near the victim. The conversation (in-person or virtual) is being held in an illuminated room, where the shiny object can be observed from outside the room (e.g., through a window).

Attacker. We consider a potential attacker to be someone with an interest in accessing a victim's sensitive information by executing the 'little seal bug' attack in order to surveil the victim and obtain the content of their conversation. Actors with these interests can range from (1) a business competitor, (2) private investigators, (3) a surveilling spouse/family member [20], or (4) thieves/other malicious actors.

Security Risks. The victim's security risk from the 'little seal bug' attack is proportional to the value the attacker places on the sensitive information the victim may share in an observed conversation. Attacker interests can include: (1) obtaining organizational passwords/operational information, (2) stealing a competitor's intellectual property (IP) and/or 'trade secrets', (3) blackmail, determining a spouse's infidelity, and/or obtaining information for divorce proceedings, or (4) obtaining a person's plans/schedule to avoid or intercept them [20].

Assumptions. We assume a victim (person) that is located in their house and seated at a desk exchanging/sharing 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 'little seal bug', in the form of a lightweight shiny object, which is located 25-50 cm away from the victim, a reasonable distance in this setting (the depth of a standard desk is 60 cm). We consider the eavesdropper to be a malicious entity interested in recovering speech from the victim's conversation by performing the 'little seal bug' attack. We assume that the eavesdropper is located within 35 meters of the target room.

Components. The 'little seal bug' attack consists of the following primary components: (1) Telescope - This piece of equipment is used to focus the field of view on the 'little seal bug' from a distance. (2) Photodiode - This sensor is mounted on the telescope, and consists of a semiconductor device that converts light into an electric current. The current is generated when photons are absorbed by the photodiode. Photodiodes are used in many consumer electronic devices (e.g., smoke detectors, medical devices). (3) Sound recovery model - This model receives an optical signal as input, and outputs the recovered acoustic signal. The eavesdropper can implement such a model with dedicated hardware (e.g., using capacitors, resistors). Alternatively, the eavesdropper can use an ADC to sample the photodiode, and process the data digitally using a laptop; in this study, we use the 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 'little seal bug'. These fluctuations cause the object to vibrate, resulting in a pattern of displacement over time that the eavesdropper measures with the photodiode, which is directed at the object via the telescope. The analog output of the photodiode 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. 2 outlines the threat model.

In general, microphones rely on three components (a diaphragm, transducer, and ADC). In the 'little seal bug' attack, the shiny object serves as a diaphragm, which vibrates when sound waves hit its surface. The transducer consists of an external light source and the remote photodiode, which is used to convert the vibrations of the lightweight shiny object (the diaphragm) to optical measurements, using the light reflected from the surface of the shiny object. An ADC is used to convert the electrical signal to a digital signal (as in standard microphones).

Significance. The significance of the 'little seal bug' attack with respect to methods presented in other studies is that the 'little seal bug': (1) is an external method that relies on a line of sight between the photodiode and the 'little seal bug' (unlike

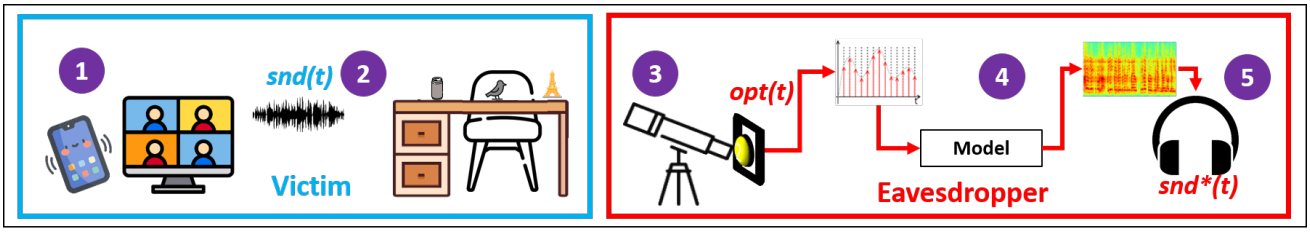

Fig. 2: The 'little seal bug' attack's threat model: The sound $snd(t)$ from the victim's conversation (1) creates fluctuations on the surface of a lightweight reflective object (e.g., an empty iced coffee can, desk ornaments) placed on a desk (2). The eavesdropper directs a photodiode at the object via a telescope (3). The optical signal $opt(t)$ is sampled from the photodiode via an ADC (4) and processed, using an algorithm to recover the acoustic signal $snd^*(t)$ (5).

other methods that require eavesdroppers to compromise a device located in physical proximity of the victim in order to obtain data and exfiltrate it [5], [2], [4], [11], [10], [7], [6], [8]), (2) recovers intelligible audio signals, so it is not limited to classifying isolated words that appear in a precompiled dictionary (unlike [2], [5], [4], [10]), and (3) can be used to recover the content of physical and virtual meetings (in contrast to TEMPEST attacks that can only be used to recover the content of virtual conversations [12], [13], [21], [22], [23]).

The methods most related to ours are the laser microphone [16], the visual microphone [14], Lamphone [15], and the Glowworm attack [13], all of which are also passive optical methods for sound recovery. Unlike those methods, the 'little seal bug' attack can recover speech: (1) from reflections of light on objects that are not electronic (as opposed to Lamphone [15] and the Glowworm attack [13] which recover sound from electronic devices that emit light, respectively speakers and light bulbs), (2) from objects which are commonly placed on desks (as opposed to Lamphone which relies on a hanging light bulb), (3) at a sound level of 75 dB, which is a significantly lower volume than prior works (as opposed to the visual microphone [14] and other methods [8], [14], [11], [10] that are limited to recovering speech at higher volumes), (4) using a photodiode, which is a passive sensor that does not provide any indication regarding its use (as opposed to the laser microphone [16] which relies on a laser transceiver), and (5) is composed of hardware (ADC, photodiode) that is not associated with spying (as opposed to the laser microphone [16]).

The 'little seal bug' attack presents an abstraction of the Lamphone threat model. In addition to desk lamps, which produce their own light, lightweight reflective objects can also be used as a diaphragm to optically recover sound, with an external light source acting as a transducer. Since these objects are more ubiquitous than desk lights/lamps in office settings (e.g., iced coffee can) this presents a more accessible threat model to attackers.

IV. Reflective Objects As Microphones

In this section, we describe the series of experiments we performed which were aimed at: (1) explaining why lightweight reflective objects can be used to recover sound, (2) determining the specifications of the equipment needed to recover speech from a shiny object, (3) characterizing the optical signal obtained by a photodiode when shiny objects vibrate in response to sound, and (4) analyzing the effect of ambient factors on sound recovery. In the experiments described in this section, we chose to use a simple shiny object (a light weight) as the lightweight reflective object; the use of such a generic object allowed us to investigate whether a photodiode can be used to successfully recover sound from shiny objects.

A. Shiny Objects as Microphones

In this subsection, we show that shiny objects can be exploited as optical microphones. First, we demonstrate that a shiny object vibrates according to the sound waves that hit its surface. Then, we show that the vibration of a shiny object can be recovered from the light reflected from its surface by using a photodiode.

Experimental Setup: A wire was used to attach a shiny 50 gram weight, purchased from Amazon [24] to the upper edge of a stand. We attached a gyroscope [25] to the bottom of the weight and connected the gyroscope to a Raspberry Pi 3. We then sampled the gyroscope via the Raspberry Pi at 4000 Hz (see Fig. 3). Finally, we created an audio file of a frequency scan (chirp/sweep) from 200-1500 Hz and played the audio file, via speakers which were placed 10 cm from the weight, at an average volume level of 75 dB.

Results & Conclusions: Fig. 3 presents spectrograms of (1) the frequency scan and (2) the weight's vibrations, extracted from the gyroscope's measurements. As can be seen from the spectrograms, the weight vibrates based on the nearby sound from the frequency scan. We note that, as can be seen in the spectrograms, there are distortions in the reconstructed audio from the gyroscope measurements, and we demonstrate in the following sections signal processing and denoising methods to help recover less distorted audio.

In the next experiment, we show that the vibrations of the weight can be captured using a photodiode when light is shining on the object.

Experimental Setup: Extending the setup used in the previous experiment, we directed a telescope (Meade 8" ACF) at the weight from a distance of 10m, at the same height as the weight. We mounted a photodiode (the Thorlabs PDA100A2 [26]) to the telescope. The voltage was obtained from the photodiode using a 24-bit ADC NI-9234 card [27] at a sampling rate of 4 KHz and processed in a LabVIEW script

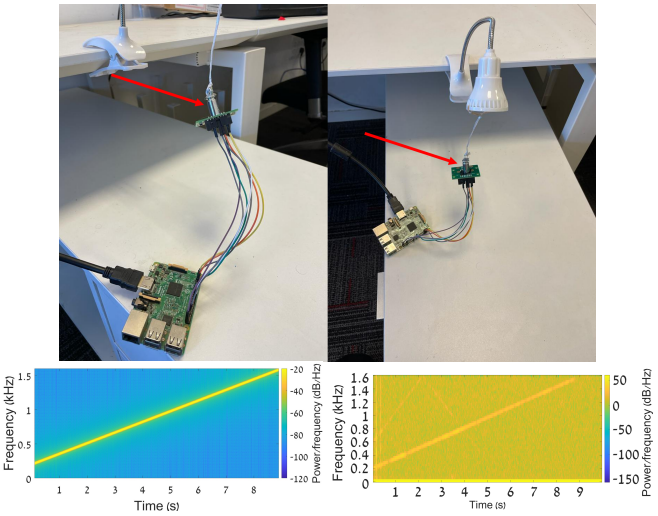

Fig. 3: Top: The gyroscope attached to the weight (indicated by the red arrow). The gyroscope is sampled by a Raspberry Pi 3. Bottom: The spectrogram of the original frequency scan audio file (left) and the spectrogram extracted from the gyroscope measurements when the frequency scan was played by nearby speakers (right).

that we wrote. 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 speakers which were placed 10 cm from the weight, at an average volume level of 75 dB. We then obtained the optical signal via the photodiode when the lights in the room were on and off, using three different weights: weights of 10, 50, and 100 grams. Finally, we covered the weights with black tape and repeated the abovementioned experiment again when the lights in the room were on. The lights in the room were standard LED office ceiling lights. The experiment was conducted without external light (e.g., sunlight) entering the room.

Results & Conclusions: Fig. 4 presents the signal-to-noise ratio (SNR) obtained from the optical measurements acquired in the experiments. The following observations can be made by analyzing the SNR values: (1) When the lights are on and the surface of the weights is not covered with black tape, the vibrations of the weights can be spotted in the optical measurements (the SNR is positive). (2) However, if light is not reflected from the weights (because the lights are off or the surface of the weight is covered with black tape), an object's vibrations (the weights in our case) cannot be identified in the optical measurements (as can be seen in Fig. 4, the SNR is zero in the experiments conducted when the lights in the room were off and when the surface of the weights was covered with black tape). (3) The SNR increases with lighter weights, but the unique behavior of the SNR is maintained across all of the weights used. (4) The response is not the same across the spectrum and decreases as a function of the frequency.

We note that there is a "peak" in the SNR around 900Hz in the optical measurements obtained when the lights are on. This could be caused by various factors associated with the physical characteristics of the hanging weight and connected wire (e.g., a mechanical resonant frequency around 900Hz, tension/elasticity of the wire, etc.) which results in a higher SNR in the optical signal around this frequency.

Based on these experiments, we made the following conclusions: (1) When light hits a reflective object, the reflection of the light from the object is modulated by the object's movement, which is associated with the nearby sound. (2) In some cases, an equalizer is required to balance an unequal response across the spectrum. (3) The zero SNR value obtained in the dark rules out another possible explanation as to why we could spot the vibrations of the weights in the optical measurements, which is that the measurements obtained by the photodiode were affected by EMR emitted from the speakers; clearly, the optical measurements were not affected by any possible side effects; if they were, the SNR in the dark would not be zero.

B. Specifications of the Equipment Required for Speech Recovery

Next, we describe our experiments aimed at measuring the vibrations of shiny objects when sound waves hit their surface. Based on the results, we establish criteria for the sensitivity specifications of the equipment needed to recover sound from shiny objects' vibrations.

Experimental Setup: A wire was used to attach a shiny 50 gram weight to the upper edge of a stand. We attached a gyroscope [25] to the bottom of the weight and connected the gyroscope to a Raspberry Pi 3. Then, we sampled the gyroscope via the Raspberry Pi at 1600 Hz (see Fig. 3). We placed speakers in front of the weight (a few centimeters away) and played various sine waves (200, 250, 300, 350,...., 800 Hz) from the speakers at 75 dB. Finally, we obtained measurements from the gyroscope when the sine waves were played.

Results: We calculated the peak-to-peak values for each of the three angles measured by the gyroscope for every 1800 consecutive measurements (one second of sampling). Based on the known formula of the spherical coordinate system [28], we calculated the 3D vector (x,y,z) that represents the peak-to-peak vibration for each of the axes. We calculated the Euclidean distance between this vector and the vector of the initial position. As can be seen from the results, which are presented in Fig. 5, the total movement of the weight is very miniscule (0.3-1.3 mm).

In the next experiment, we examine how eavesdroppers can determine the sensitivity of the equipment needed to recover sound based on shiny objects' minute vibrations (0.3-1.3 mm).

Experimental Setup: We directed a telescope (Meade 8" ACF) at a shiny weight (at the same height as the weight)and mounted a photodiode (the Thorlabs PDA100A2 [26]) to the telescope. The voltage was obtained from the photodiode using a 24-bit ADC NI-9234 card [27] at a sampling rate of 1600 Hz and processed in a LabVIEW script that we wrote. The internal gain of the photodiode was set at 50 dB. Finally, we placed the telescope at various distances (1, 2, 3, 4, 6, 7, 9 meters) from the weight and measured the voltage obtained from the photodiode for each distance. The lights in the room were standard LED office ceiling lights. The experiment was conducted without external light (e.g., sunlight) entering the room.

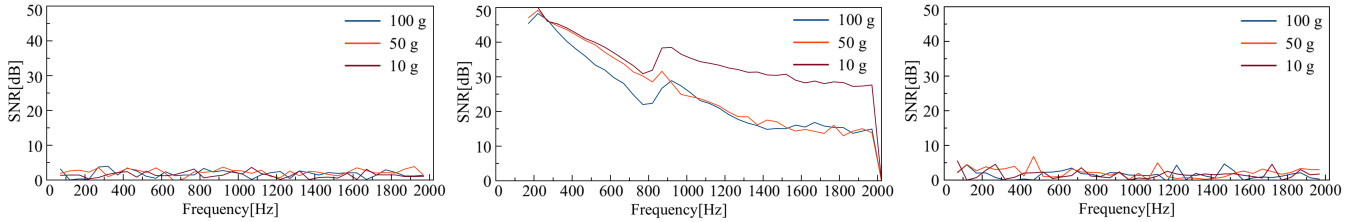

Fig. 4: The SNR obtained from the weights when the lights in the room were off (left) and on (middle) and when the surface of the weights was covered with black tape (right).

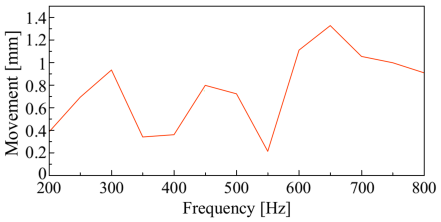

Fig. 5: The movement of the weights as a function of the frequency of the sound played.

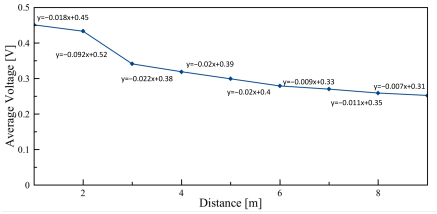

Fig. 6: The output of the photodiode (voltage) as a function of its distance from the weights.

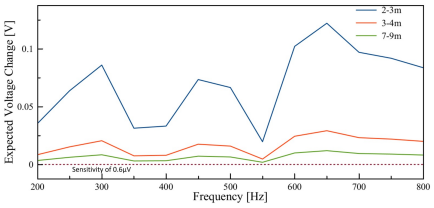

Fig. 7: The expected voltage of the photodiode across the spectrum from various distances.

Results: Based on the results, we computed the linear equation between each two consecutive points. The gradient of each linear equation represents the expected differential in voltage for a given differential in distance between the photodiode and the weight. The results of this experiment and the linear equations are presented in Fig. 6.

Based on the linear equations, we calculated the expected voltage for each expected movement of the weight (which changes the distance between the photodiode and the weight) in the 200-800 Hz spectrum for a sound level of 75 dB (using the results obtained from the previous experiment). The expected voltage changes as a function of the frequency for three distance ranges (2-3, 3-4, and 7-9 meters) are presented in Fig. 7.

We now explain how to use the data presented in this figure in order to determine which frequencies can be recovered from the optical measurements obtained for a sound level of 75 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 $[-5, 5]$ volts (e.g., the card used in our experiments) provides a sensitivity of: $\frac{10}{2^{24}-1} \approx 0.6$ µV. As can be seen in Fig. 7, a sensitivity of 0.6 µV is sufficient to recover the entire spectrum for all three ranges, since the expected voltage for each frequency is greater than the sensitivity of the ADC. This calculation can be used by eavesdroppers to determine the specifications of the equipment required to recover speech from a desired distance.

C. Characterizing the Optical Signal

In this experiment, we examine the characteristics of the optical signal when no sound is played, with the aim of profiling the optical signal in order to filter out any side effects that are not associated with sound from the recovered audio signal.

Experimental Setup: We mounted a photodiode (the Thorlabs PDA100A2 [26]) to the telescope (Meade 8" ACF) directed at the weights from a distance of 10m, and at the same height as the weights. The voltage was obtained from the photodiode using a 24-bit ADC NI-9234 card [27] at a sampling rate of 2.5 KHz and processed in a LabVIEW script that we wrote. We obtained five seconds of optical measurements from the photodiode when no sound was played near the weights when the lights in the room were turned on. The lights in the room were standard LED office ceiling lights. The experiment was conducted without external light (e.g., sunlight) entering the room.

Results & Conclusions: The FFT graph extracted from the optical measurements when no sound was played is presented in Fig. 8.

As can be seen in the FFT graph, peaks appear around 100 Hz, 200 Hz, etc. Since the optical measurements were obtained via a photodiode directed at an object that reflected the light in the room, the light frequency (100 Hz) and its harmonics are added to the optical measurements. The optical phenomenon that occurs at 100 Hz (which was captured by the photodiode) is the result of power net harmonics. The LED bulb in the room uses DC voltage which is converted from AC. A diode bridge is integrated in 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 which creates a periodic phenomenon of 100 Hz, 200 Hz, 300 Hz, etc. We concluded that bandstop filtering would be required to eliminate side effects which are not the result of the sound that we want to recover yet significantly impact the optical signal.

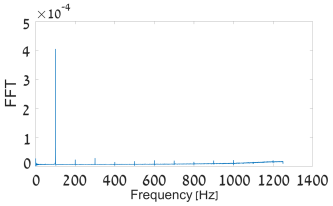

Fig. 8: The FFT of the optical signal when no sound is played (the baseline).

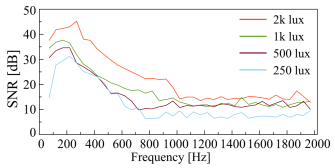

Fig. 9: The SNR as a function of the light reflected from the weight.

D. Analyzing the Effect of Ambient Factors

In this experiment, we analyze the effect of ambient factors that may exist in the target room. First, we examine how the SNR of the optical signal obtained from a weight is affected by the intensity of the light reflected from the weight.

Experimental Setup: In this case, we made one change to the experimental setup used to obtain optical measurements in the experiments described in Subsection IV-A: we measured the amount of light that hits the surface of the shiny object using a professional lux meter (this corresponds to the amount of light reflected back from the object to the photodiode). We played the same frequency scan from Subsection IV-A via the speakers near a 50 gram weight in four experiments, varying the intensity of the light reflected on the object in each experiment (250, 500, 1000, 2000 lux). The light was output by an LED flashlight. The experiment was conducted without external light (e.g., sunlight) entering the room. We obtained the optical measurements via the ADC at a sampling rate of 4 KHz.

Results & Conclusions: Fig. 9 presents the SNR obtained from the optical measurements in the four experiments.

As can be seen, the intensity of the light reflected from the weight has a strong effect on the SNR of the optical measurements. Unsurprisingly, the SNR improves when light of a greater intensity hits the surface of the weight.

V. Recovering Speech

In this section, we leverage the findings presented in Section IV and explain how to recover audio from measurements obtained from a photodiode directed at a shiny object. Throughout this section, we consider $snd(t)$ as the sound played inside the victim's room, $opt(t)$ as the optical signal obtained via a photodiode directed at a shiny object, and $snd^*(t)$ as the audio signal recovered from $opt(t)$. The following steps are performed to recover speech from the optical measurements:

Bandstop Filters. As discussed in Section IV and seen in Fig. 8, the optical signal consists of side effects that are not the result of the sound played, e.g., the harmonics of 100 Hz (200 Hz, 300 Hz, etc.). We applied bandstop filters across 100 Hz, 200 Hz, 300 Hz, and 400 Hz using a band of 5 Hz across the filtered frequency. We did not apply bandstop filterers beyond 400 Hz, because the added noise (i.e., the side effects) is very low (the same noise level as the signal).

Scaling. Scaling is a simple method used in audio processioning for speech enhancement. We scale the values of $opt(t)$ to the range of [-1,1].

Spectral Subtraction. Spectral subtraction is a method for the restoration of the power or magnitude spectrum of a signal with additive noise, in which an estimate of the average noise spectrum is subtracted from the noisy signal spectrum. This method is widely used in speech processing, since it can adaptively estimate the average noise spectrum from a baseline signal (e.g., a signal that was recorded in silence) or the signal itself. We used spectral subtraction to denoise our signal by estimating the average noise spectrum from the signal itself [29].

Equalizer. Equalization is the process of adjusting the balance between frequency components within an electronic signal. It is used to amplify the response of weak frequencies. We designed the equalizer based on the spectral response presented in Fig. 9 which shows that the spectral response decreases as a function of the frequency (the SNR of the higher frequencies is weaker than the SNR of lower frequencies). Based on this observation we designed and utilized the equalizer presented in Fig. 11.

The techniques used in this study to recover speech are commonly used in the area of speech processing; we utilized them for the following reasons: (1) the techniques rely on a speech signal obtained from a single channel; if eavesdroppers have the ability to use additional sensors for sampling, allowing them to obtain several signals via multiple channels, other methods can also be applied to recover an optimized signal; (2) the techniques don't require prior data collection to create a model; other novel speech processing methods use neural networks to characterize/profile the noise in order to optimize the speech quality, however to create robust models, such neural networks require a large amount of data for the training phase, a requirement that may be offputting to eavesdroppers; and (3) the techniques are adaptive and can be applied to recover sound from various shiny objects, which may behave differently (e.g., require different equalizers) or produce different noise levels/distributions.

VI. Evaluation

In this section, we evaluate the performance of the 'little seal bug' attack in terms of its ability to recover sound from light reflected from various objects. We also evaluate the 'little seal bug' attack's performance at various distances. We compare the 'little seal bug' attack's performance to two state-of-the-art sound recovery methods (the visual microphone [14] and Lamphone [15]) by replicating the experimental setups reported in the original works.

A. Metrics

The reader can assess the quality of the recovered sound visually by analyzing the extracted graphs (spectrograms), qualitatively by listening to the recovered audio signals online,[3,4,5] 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

[3] https://youtu.be/FvxxJ0Ieccc
[4] https://youtu.be/Ff-y7izdGgs
[5] https://youtu.be/9F7REbJJdLs

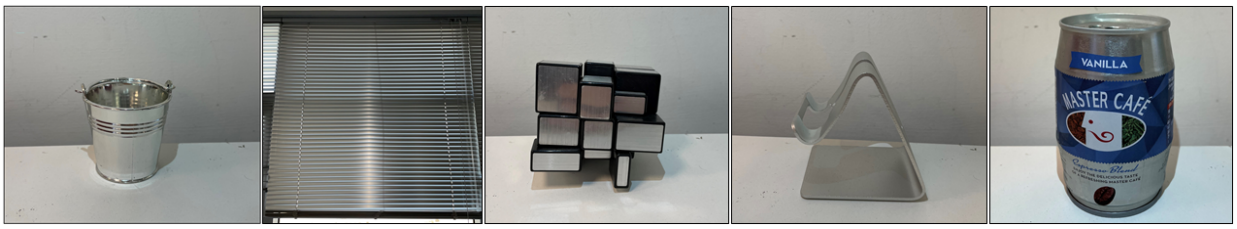

Fig. 10: The lightweight reflective objects used to recover sound (from left to right): decorative bucket, Venetian blinds, Rubik's Cube, smartphone stand, and iced coffee can.

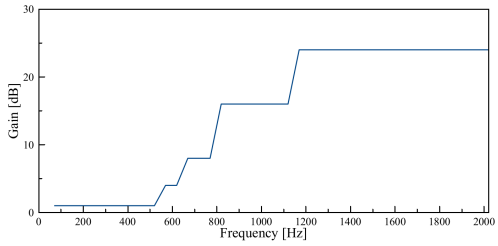

Fig. 11: The equalizer function used.

comprehensible speech is in given conditions [30], measured according to the metric suggested by [31], and (2) NSNR (NIST Speech SNR) - the speech-to-noise ratio, which is the logarithmic ratio between speech power and noise power estimated over 20 consecutive milliseconds [32]. Higher intelligibility/NSNR values indicate better sound quality.

B. Sound Recovery Setup

We used the following setup to recover sound in all of the experiments conducted in this section: a telescope (Meade 8" ACF) was directed at various lightweight reflective objects. We mounted a photodiode (Thorlabs PDA100A2 [26]) to the telescope. The output of the photodiode (the voltage associated with the light intensity) was sampled with a 24-bit ADC NI-9234 card. The sampling frequency of the ADC was configured at 2 KHz. We used Logitech Z200 speakers, which were placed on a dedicated stand, to produce the sound; the sound level was measured with a professional decibel meter. The acoustic signals were recovered using a MATLAB script that we wrote which isolates the acoustic signal from the optical measurements (see Section V). In the rest of this section, we refer to this setup as the eavesdropping equipment.

In our evaluation we recovered speech from a variety of lightweight reflective objects: a Rubik's Cube, a decorative bucket, a smartphone stand, and an empty iced coffee can, as well as an item often used in offices to protect individuals' privacy: Venetian blinds. The objects are presented in Fig. 10.

In the following experiments, we describe how we replicate the experimental setups of the visual microphone [14] and Lamphone [15] studies.

C. Recovering Speech from Various Objects

In this section, we demonstrate the ability to recover speech from various shiny objects, utilizing six sentences from the

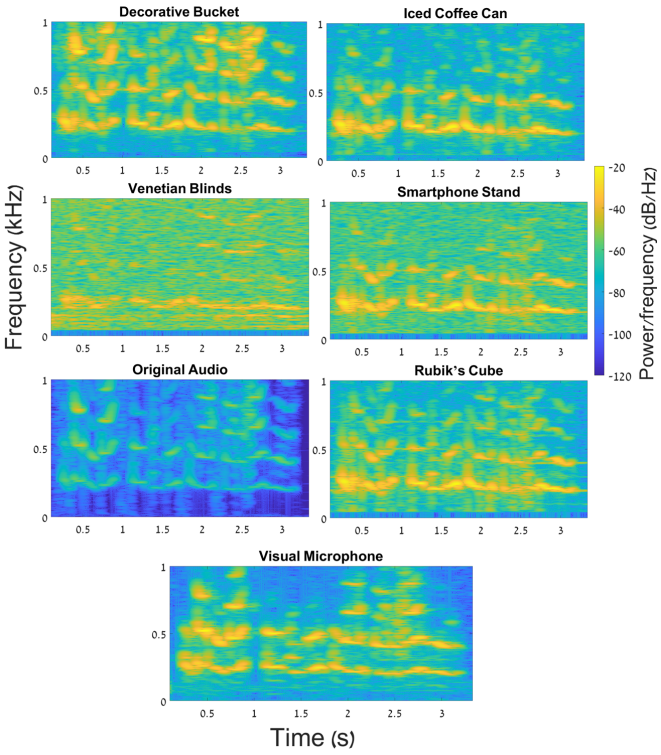

Fig. 12: Recovery of the sentence "She had your dark suit in greasy wash water all year" by fadg0,sa1 from various objects, as well as the recovery from the visual microphone [14].

TIMIT repository [33] recovered by the visual microphone in [14] from a distance of 2.5 meters.

Experimental Setup: We replicated the experimental setup used in [14] as follows: We placed the speakers on a dedicated stand 5 cm from various shiny objects (the exact same range used in [14]). We played the same six sentences from the TIMIT repository recovered by the visual microphone [14] via speakers at 75 dB (we note that the experiments performed in [14] were conducted with speech played at an average volume level of 95 dB). We placed the eavesdropping equipment (specified in Subsection VI-B) 2.5 meters from the lightweight reflective objects (the exact same range used in [14]) at the same height as the lightweight objects. The lights in the room were standard LED office ceiling lights. The experiment was conducted without external light (e.g., sunlight) entering the room. Our experimental setup was used to recover speech from the five objects presented in Fig. 10.

Results & Conclusions: We recovered the audio signals

		Rubik's Cube		Decorative Bucket		Smartphone Stand		Iced Coffee Can		Venetian Blinds		Visual Microphone	
	Speech	Int.	NSNR	Int.	NSNR	Int.	NSNR	Int.	NSNR	Int.	NSNR	Int.	NSNR
Female speaker - fadg0, sa1	"She had your dark suit in greasy wash water all year"	0.73	7.5	0.79	4.25	0.64	20.75	0.64	17.25	0.51	4.3	0.72	26.8
Female speaker - fadg0, sa2	"Don't ask me to carry an oily rag like that"	0.59	4	0.62	3.75	0.52	4.75	0.51	7.25	0.39	3.5	0.65	43.3
Male speaker - mabw0, sa1	"She had your dark suit in greasy wash water all year"	0.65	5.25	0.74	2.25	0.61	6.75	0.59	8.5	0.495	6.5	0.59	27.3
Male speaker - mabw0, sa2	"Don't ask me to carry an oily rag like that"	0.59	5	0.69	3.25	0.49	15	0.49	5.5	0.41	27.5	0.67	18
Male speaker - mccs0, sa1	"She had your dark suit in greasy wash water all year"	0.72	6.25	0.77	12.25	0.63	12	0.63	16.25	0.51	10.8	0.77	6
Male speaker - mccs0, sa2	"Don't ask me to carry an oily rag like that"	0.63	3	0.71	14.25	0.54	3.75	0.53	5.25	0.41	25.5	0.72	25.8
	Average	0.65	5.17	0.72	6.67	0.57	10.5	0.56	10	0.45	13.02	0.68	24.53
	STD	0.05	1.46	0.06	4.73	0.06	6.06	0.06	4.9	0.05	9.83	0.06	12.27

TABLE I: The intelligibility (Int.) and NSNR of the recovered speech using the 'little seal bug' attack and the visual microphone [14] based on sentences from the TIMIT repository.

from the optical measurements using the 'little seal bug' attack (see Section V). The recovered audio signals are available online[3,4] where they can be heard. The spectrograms extracted from the optical measurements for the six sentences recovered when using various objects are presented in Fig. 12 and Figs. 14-18 in the appendix. We evaluated the intelligibility and NSNR of the recovered signals, and the results are reported in Table I. We also downloaded the same six audio signals recovered and published in [14] and evaluated their performance based on the same metrics. The following interesting observations can be made from the results presented in Table I: The average intelligibility of the speech recovered depends, to a large extent, on the shiny object used to implement the attack. In some cases, the average intelligibility of the object is considered good (the Rubik's Cube and decorative bucket) according to [30]; in the case of the other objects, the average intelligibility is considered fair. In one case, the decorative bucket succeeds at enabling reconstruction with excellent intelligibility (>0.75) where the visual microphone [14] was unable to. The spectrograms for this sentence are displayed in Fig. 12. A similar conclusion can be made by analyzing the NSNR of the recovered speech, which ranges from 3.75-13 for the five examined objects. We note that these results were achieved at a significantly lower sound intensity than [14], without needing to pre-record the optical measurements and then reconstruct the sound afterwards. Interestingly, due to the higher sensitivity of a photodiode compared to a video camera, it is able to obtain more precise visual information from object vibrations at lower sound volume levels than a video camera. It is also worth noting how from Venetian blinds, intended to protect a person's privacy by covering their window, an eavesdropper is capable of capturing optical measurements which produce speech reconstructions with fair intelligibility. This demonstrates that Venetian blinds provide a false sense of security when obscuring a room during a meeting.

D. Recovering Speech from Various Distances

Having demonstrated speech recovery from various objects, we test the performance of the speech recovery from various distances (15, 25, 35 meters), utilizing a benchmark statement used in the paper proposing Lamphone [15] at the sound level of a virtual meeting 75 dB. The benchmark statement is

	Intelligibility		NSNR	
	Rubik's Cube	Lamphone	Rubik's Cube	Lamphone
15 meters	0.61	0.52	16.3	21
25 meters	0.55	0.49	14.5	21.8
35 meters	0.5	0.45	14.5	17.5

TABLE II: The intelligibility and NSNR results of Lamphone [15] and the 'little seal bug' attack (Rubik's Cube) for the recovery of speech from various distances.

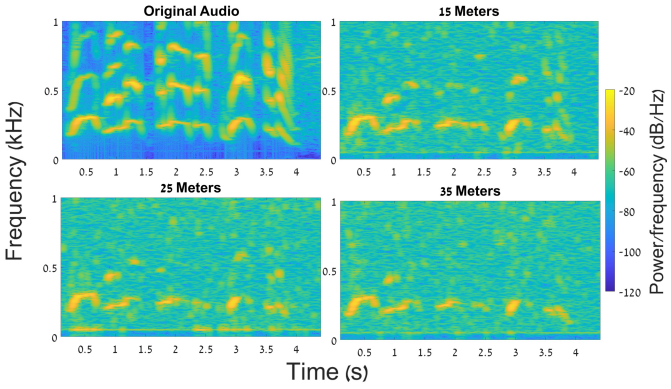

Fig. 13: Recovery of the benchmark statement, "We will make America great again", from various distances (15, 25, 35 meters) from light reflected from a Rubik's Cube when the speakers were located 25 cm from the object.

the viral Donald Trump quote "We will make America great again", widely recognizable from his 2016 US presidential campaign.

Experimental Setup: We replicated the experimental setup used in [15] as follows: We placed the eavesdropping equipment (specified in Subsection VI-B) at various distances (15, 25, 35 meters) from a Rubik's Cube that was placed on a desktop (at the exact same ranges used in [15] and at the same height as the lightweight objects); the speakers were placed at the edge of the desktop, 25 cm away (the exact same range used in [15]) from the reflective object (the distance is equivalent to half of the depth of a standard desktop). The lights in the room were standard LED office ceiling lights. The experiment was conducted without external light (e.g., sunlight) entering the room. Then, we obtained the optical measurements as the statement was played via the speakers at a volume level of 75 dB.

Results & Conclusions: We recovered the audio signals from the optical measurements. The recovered audio signals are available online[5]. The spectrograms of the speech extracted from the Rubik's Cube from various distances (15, 25, 35 meters) when objects were located 25 cm from the speakers are presented in Fig. 13. We also downloaded the recoveries of the same benchmark sentence published in [15] and evaluated their performance based on the same metrics. The intelligibility and NSNR of the recovered signals are reported in Table II. The following observations can be made from these results: (1) Although the intelligibility of the audio signals recovered from the Rubik's Cube decreases with distance, fair intelligibility (according to [30]) is achieved for the examined distances. (2) The intelligibility scores achieved by the 'little seal bug' attack are higher than Lamphone [15] in all three tested ranges. In particular, we note that from 35 meters, the 'little seal bug' attack achieves comparable intelligibility to Lamphone performed from 15 meters, less than half the distance of the 'little seal bug' attack. This indicates that lightweight shiny objects can serve as optical implants to recover sound with a higher intelligibility than desk lamps/lights, from farther distances.

VII. Countermeasures

In this section, we discuss known countermeasures against the 'little seal bug' attack. One organizational approach for preventing the 'little seal bug' attack is policy-based; organizations could prohibit employees from displaying any decorative lightweight shiny objects (e.g., statuettes) that vibrate when they are hit by sound waves on their desks. Organizations could also prohibit the use of shiny Venetian blinds on their premises; such blinds could be replaced with Venetian blinds with a matte finish or a different type of window covering. The main limitation of this approach is the fact that the 'little seal bug' attack can also be used to recover speech from an empty beverage can or smartphone stand, items which are commonly seen on desks in office settings.

Another approach is to contain the optical leakage. This can be done by installing a non-reflective/non-shiny curtain to eliminate the line of sight to objects capable of vibrating when hit by sound waves or by changing the location of sensitive conversations to a windowless room. While the latter might be effective for home offices, the number of inner rooms in office buildings is usually limited. Similarly, curtains are seen more often in homes than in offices, and even in homes, people often prefer to work in a room with natural light, as opposed to a room with the curtains drawn.

Another approach for preventing eavesdroppers from recovering speech from lightweight shiny objects is to create a safety perimeter. This can be done by installing a fence around a home/building in order to limit eavesdroppers' ability to recover sound by forcing them to perform the attack from farther away (as was shown in Section VI, the quality of the recovered speech decreases with distance).

VIII. Limitations

The 'little seal bug' attack suffers from the following limitations:

Threat Model. The attack depends on multiple non-trivial assumptions in order to successfully recover sound optically from lightweight shiny objects. In particular, the 'little seal bug' depends on: (1) the existence of a suitable lightweight shiny object near the victim (25 cm), (2) a direct line of sight to the lightweight shiny object, and (3) a sound level of 75 dB, which is higher than most normal conversations/virtual meetings.

Quality of Recoveries. The effectiveness of the attack is proportional to the quality of the equipment used by the attacker. Due to the vibrations of the lightweight reflective objects being minuscule, sensitive equipment is required to observe and process the optical measurements needed for speech recovery. In our study, the cost of the equipment came to $2,500 ($1,000 - telescope, $500 - photodiode, and $1,000 - ADC), an investment which allowed us to recover speech from a distance of 35 meters. In order to increase the attack range and maintain an SNR that allows attackers to recover speech, more sensitive and expensive professional equipment is required (e.g., a more sensitive ADC and photodiode, a professional telescope). Such equipment would enable attackers to recover speech from greater distances.

IX. Discussion

The results presented in this research show that eavesdroppers located in an adjacent building (35 meters away) can use a photodiode in order to recover the content of a virtual meeting or phone call of a victim seated at a desk on which a lightweight reflective object is placed 25 cm away from them.

The primary objectives of this study were to: (1) raise awareness regarding the fact that lightweight shiny objects, present in many home offices and considered innocuous, can serve as 'little seal bugs', optical implants exploited by eavesdroppers to recover sound, and (2) increase understanding in optical speech recovery, which became a growing field of research in the last eight years (e.g., [14], [9], [15], [13], [18]).

The risks and potential victims of the 'little seal bug' are widespread, including: (1) workers targeted by competitors for business/organizational information, and (2) spouses/family members targeted by a private investigator/nosey spouse for information on infidelity or other intimate partner violence/-surveillance motivations [20]. As long as the victim is located near a lightweight shiny object, which can be observed from an external location, the 'little seal bug' attack is a real, practical threat to potential victims' privacy. These facts may encourage malicious actors, interested in acquiring confidential information, to recover speech from shiny objects. To avoid this, such objects shouldn't be present in the victim's vicinity when discussing sensitive information.

For future research, we propose evaluating the 'little seal bug' attack's performance under different lighting conditions (LED with PWM dimming, natural lighting, etc.). In addition, we propose evaluating different intermediate mediums' effect (e.g., glass) on sound recovery as future work. Finally, we propose additional evaluations with human speakers as future work.

REFERENCES

[1] J. Landt, "The history of rfid," *IEEE potentials*, vol. 24, no. 4, pp. 8–11, 2005.

[2] Y. Michalevsky, D. Boneh, and G. Nakibly, "Gyrophone: Recognizing speech from gyroscope signals," in *23rd USENIX Security Symposium (USENIX Security 14)*. San Diego, CA: USENIX Association, 2014, pp. 1053–1067. [Online]. Available: https://www.usenix.org/conference/usenixsecurity14/technical-sessions/presentation/michalevsky

[3] Z. Ba, T. Zheng, X. Zhang, Z. Qin, B. Li, X. Liu, and K. Ren, "Learning-based practical smartphone eavesdropping with built-in accelerometer," in *Proceedings of the Network and Distributed Systems Security (NDSS) Symposium*, 2020, pp. 23–26.

[4] 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)*, vol. 00, pp. 116–133. [Online]. Available: doi.ieeecomputersociety.org/10.1109/SP.2018.00004

[5] L. Zhang, P. H. Pathak, M. Wu, Y. Zhao, and P. Mohapatra, "Accelword: Energy efficient hotword detection through accelerometer," in *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 2015, pp. 301–315.

[6] M. Guri, Y. Solewicz, A. Daidakulov, and Y. Elovici, "Speake(a)r: Turn speakers to microphones for fun and profit," in *11th USENIX Workshop on Offensive Technologies (WOOT 17)*. Vancouver, BC: USENIX Association, 2017. [Online]. Available: https://www.usenix.org/conference/woot17/workshop-program/presentation/guri

[7] N. Roy and R. Roy Choudhury, "Listening through a vibration motor," in *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '16. New York, NY, USA: ACM, 2016, pp. 57–69. [Online]. Available: http://doi.acm.org/10.1145/2906388.2906415

[8] 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: IEEE Computer Society, may 2019. [Online]. Available: https://doi.ieeecomputersociety.org/10.1109/SP.2019.00008

[9] S. Sami, Y. Dai, S. R. X. Tan, N. Roy, and J. Han, "Spying with your robot vacuum cleaner: Eavesdropping via lidar sensors," in *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*, ser. SenSys '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 354–367. [Online]. Available: https://doi.org/10.1145/3384419.3430781

[10] G. Wang, Y. Zou, Z. Zhou, K. Wu, and L. M. Ni, "We can hear you with wi-fi!" *IEEE Transactions on Mobile Computing*, vol. 15, no. 11, pp. 2907–2920, Nov 2016.

[11] T. Wei, S. Wang, A. Zhou, and X. Zhang, "Acoustic eavesdropping through wireless vibrometry," in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '15. New York, NY, USA: ACM, 2015, pp. 130–141. [Online]. Available: http://doi.acm.org/10.1145/2789168.2790119

[12] J. Choi, H.-Y. Yang, and D.-H. 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*, ser. CCS '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 1085–1101. [Online]. Available: https://doi.org/10.1145/3372297.3417241

[13] B. Nassi, Y. Pirutin, T. Galor, Y. Elovici, and B. Zadov, "Glowworm attack: Optical tempest sound recovery via a device's power indicator led," in *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, 2021, pp. 1900–1914.

[14] A. Davis, M. Rubinstein, N. Wadhwa, G. J. Mysore, F. Durand, and W. T. Freeman, "The visual microphone: passive recovery of sound from video," 2014.

[15] B. Nassi, Y. Pirutin, R. Swisa, A. Shamir, Y. Elovici, and B. Zadov, "Lamphone: Passive sound recovery from a desk lamp's light bulb vibrations," in *31st USENIX Security Symposium (USENIX Security 22)*, 2022, pp. 4401–4417.

[16] R. P. Muscatell, "Laser microphone," Oct. 25 1983, uS Patent 4,412,105.

[17] P. Walker and N. Saxena, "Laser meager listener: A scientific exploration of laser-based speech eavesdropping in commercial user space," in *2022 IEEE 7th European Symposium on Security and Privacy (EuroS&P)*, 2022, pp. 537–554.

[18] M. Sheinin, D. Chan, M. O'Toole, and S. G. Narasimhan, "Dual-shutter optical vibration sensing," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 16 324–16 333.

[19] B. Nassi, Y. Pirutin, J. Shams, R. Swissa, Y. Elovici, and B. Zadov, "Optical speech recovery from desktop speakers," *Computer*, vol. 55, no. 11, pp. 40–51, nov 2022.

[20] E. Tseng, R. Bellini, N. McDonald, M. Danos, R. Greenstadt, D. McCoy, N. Dell, and T. Ristenpart, "The tools and tactics used in intimate partner surveillance: An analysis of online infidelity forums," in *29th USENIX Security Symposium (USENIX Security 20)*. USENIX Association, Aug. 2020, pp. 1893–1909. [Online]. Available: https://www.usenix.org/conference/usenixsecurity20/presentation/tseng

[21] C. V. Wright, L. Ballard, F. Monrose, and G. M. Masson, "Language identification of encrypted voip traffic: Alejandra y roberto or alice and bob?" in *USENIX Security Symposium*, vol. 3, 2007, pp. 43–54.

[22] A. M. White, A. R. Matthews, K. Z. Snow, and F. Monrose, "Phonotactic reconstruction of encrypted voip conversations: Hookt on fon-iks," in *Security and Privacy (SP), 2011 IEEE Symposium on*. IEEE, 2011, pp. 3–18.

[23] C. V. Wright, L. Ballard, S. E. Coull, F. Monrose, and G. M. Masson, "Spot me if you can: Uncovering spoken phrases in encrypted voip conversations," in *Security and Privacy, 2008. SP 2008. IEEE Symposium on*. IEEE, 2008, pp. 35–49.

[24] "7 pcs calibration weights," https://www.amazon.com/gp/product/B08SQ2WTNY/ref=ppx_yo_dt_b_asin_title_o00_s00?ie=UTF8&psc=1.

[25] "Mpu-6000," https://invensense.tdk.com/wp-content/uploads/2015/02/MPU-6000-Datasheet1.pdf.

[26] "Pda100a2." [Online]. Available: https://www.thorlabs.com/thorproduct.cfm?partnumber=PDA100A2

[27] "Ni 9234 datasheet." [Online]. Available: https://www.ni.com/pdf/manuals/374238a_02.pdf

[28] "Spherical coordinates. encyclopedia of mathematics." http://encyclopediaofmath.org/index.php?title=Spherical_coordinates&oldid=48774.

[29] N. Upadhyay and A. Karmakar, "Speech enhancement using spectral subtraction-type algorithms: A comparison and simulation study," *Procedia Computer Science*, vol. 54, pp. 574–584, 2015.

[30] T. Houtgast and H. J. M. Steeneken, "Evaluation of speech transmission channels by using artificial signals," *Acta Acustica united with Acustica*, vol. 25, no. 6, pp. 355–367, 1971. [Online]. Available: https://www.ingentaconnect.com/content/dav/aaua/1971/00000025/00000006/art00006

[31] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An algorithm for intelligibility prediction of time–frequency weighted noisy speech," vol. 19, no. 7. IEEE, 2011, pp. 2125–2136.

[32] "Nist-snr," https://www.nist.gov/itl/iad/mig/nist-speech-signal-noise-ratio-measurements.

[33] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, and D. S. Pallett, "Darpa timit acoustic-phonetic continous speech corpus cd-rom. nist speech disc 1-1.1," *STIN*, vol. 93, p. 27403, 1993.

X. Appendix

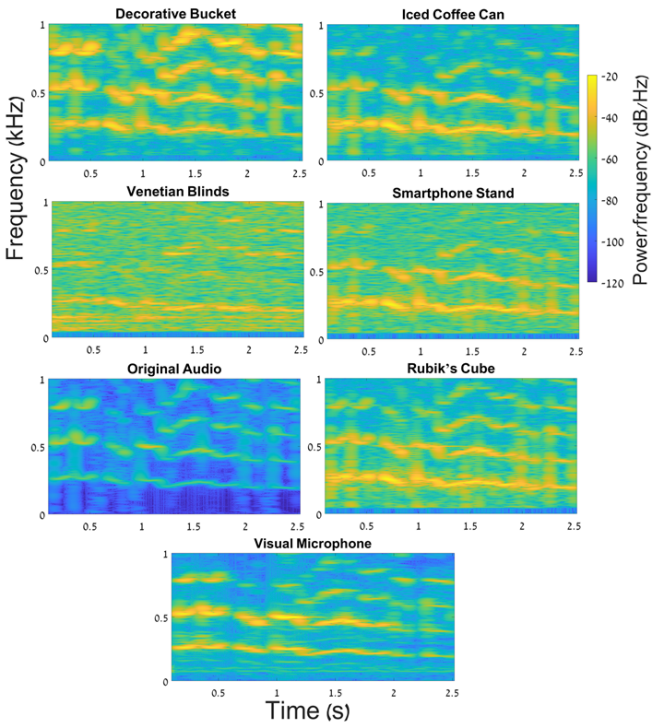

Fig. 14: Recovery of the sentence "Don't ask me to carry an oily rag like that" by fadg0,sa2 from various objects, as well as the recovery from the visual microphone [14].

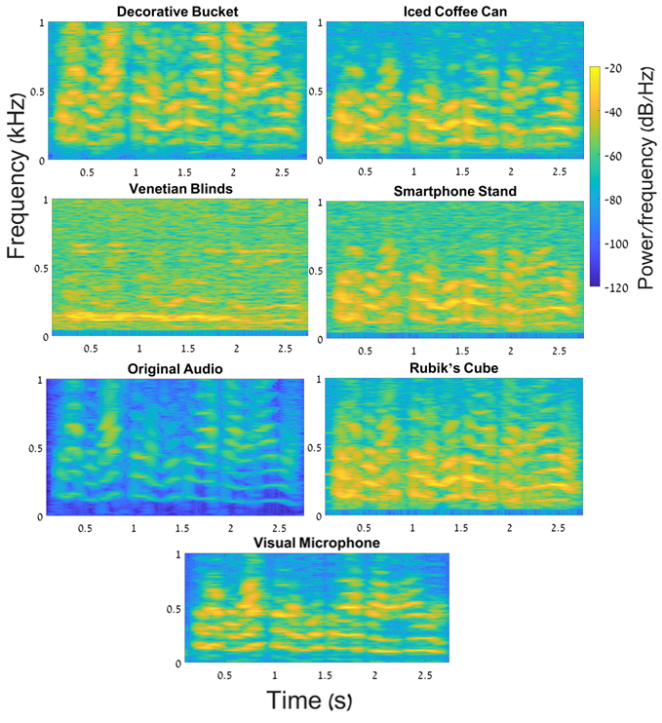

Fig. 15: Recovery of the sentence "She had your dark suit in greasy wash water all year" by mabw0,sa1 from various objects, as well as the recovery from the visual microphone [14].

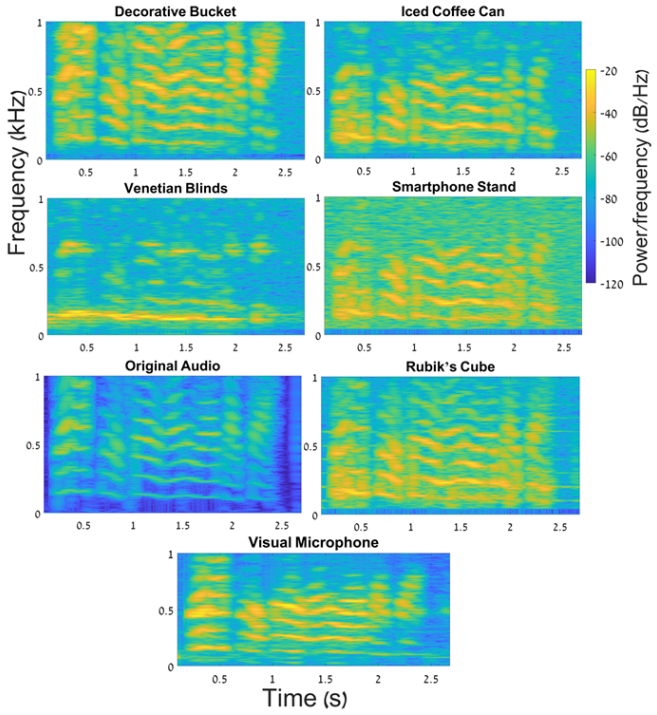

Fig. 16: Recovery of the sentence "Don't ask me to carry an oily rag like that" by mabw0,sa2 from various objects, as well as the recovery from the visual microphone [14].

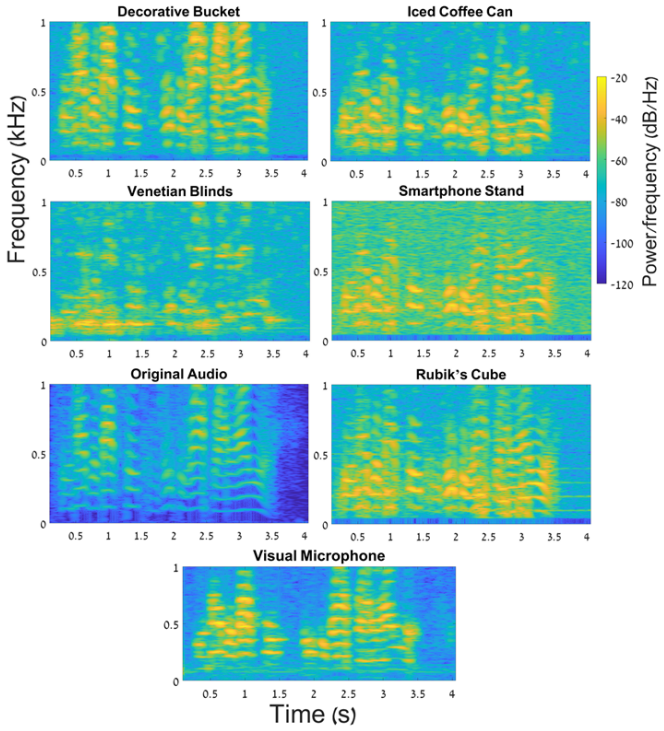

Fig. 17: Recovery of the sentence "She had your dark suit in greasy wash water all year" by mccs0,sa1 from various objects, as well as the recovery from the visual microphone [14].

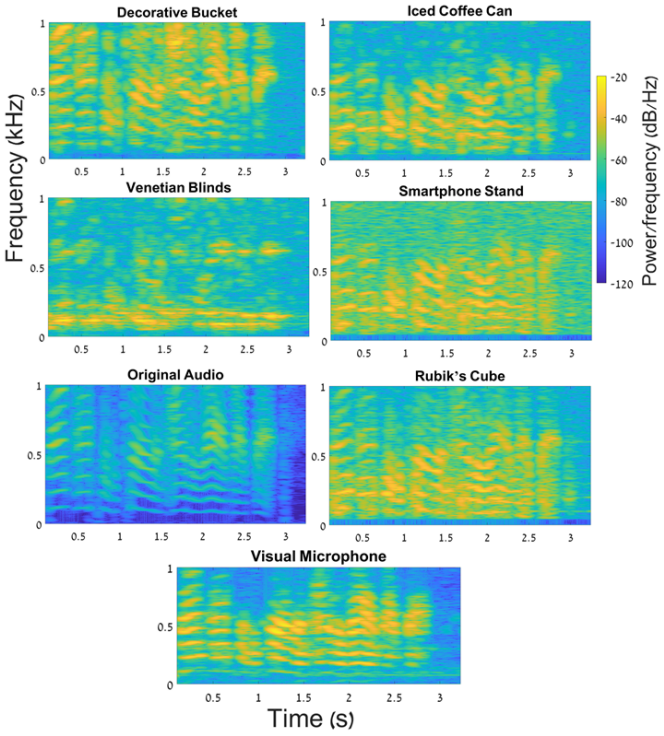

Fig. 18: Recovery of the sentence ”Don’t ask me to carry an oily rag like that” by mccs0,sa2 from various objects, as well as the recovery from the visual microphone [14].