Hearing Check Failed: Using Laser Vibrometry to Analyze the Potential for Hard Disk Drives to Eavesdrop Speech Vibrations

Payton Walker
Texas A&M University
College Station, Texas, USA
prw0007@tamu.edu

Shalini Saini
Texas A&M University
College Station, Texas, USA
s.saini@tamu.edu

S. Abhishek Anand
The University of Alabama at Birmingham
Birmingham, Alabama, USA
anandab@uab.edu

Tzipora Halevi
Brooklyn College
New York, USA
thalevi@nyu.edu

Nitesh Saxena
Texas A&M University
College Station, Texas, USA
nsaxena@tamu.edu

Abstract

Sound waves from speech can potentially induce vibrations, proportional to the speech signal, on nearby objects. Each of these objects introduces the risk for a malicious attacker to exploit the induced vibrations to eavesdrop on the speech. Such an eavesdropping attack is critical when we consider the potential for induced vibrations in standard magnetic hard disk drives (HDDs). As an instance of this threat, prior research has demonstrated that speech in certain scenarios can induce vibrations on the read/write head of an HDD in order to eavesdrop on the speech (Kwong et al.; Oakland'19).

In this paper, we revisit this line of research and aim to provide a closer investigation into whether HDDs can in fact be used as a source for eavesdropping on speech vibrations. As a foundation for our study, we utilize an effective, and robust methodology using *laser vibrometry* to measure the subtle speech vibrations induced on the read/write head. The prior study tested only a single HDD and only machine-rendered speech in a single setting with very loud speech. Our work broadens the scope of this research in many significant ways. *First*, we test multiple popular HDDs of different models and sizes to evaluate the generalizability of the overall threat. *Second*, we evaluate the threat from *live human speech* spoken near an HDD, expanding the scope of the attack to include most real-world speech settings involving normal human conversations. *Third*, we define machine-rendered speech scenarios to explore different propagation media and degrees of speech loudness.

Our findings are two-fold. *First*, we observed that live human speech traveling through the air is not generally strong enough to impact HDDs such that intelligible speech information is leaked. *Second*, most tested HDDs did not seem capable of eavesdropping on machine-rendered speech unless the speech is loud enough, or the HDD shares a surface or is in direct contact with the speaker device. This implies HDDs cannot eavesdrop live human speech.

ASIA CCS '22, May 30-June 3, 2022, Nagasaki, Japan

ACM ISBN 978-1-4503-9140-5/22/05…$15.00
https://doi.org/10.1145/3488932.3517399

CCS Concepts

• **Security and privacy → Side-channel analysis and countermeasures**; **Hardware attacks and countermeasures**; **Security in hardware**.

Keywords

side-channel attack; hard disk drives; laser vibrometry; speech eavesdropping

ACM Reference Format:
Payton Walker, Shalini Saini, S. Abhishek Anand, Tzipora Halevi, and Nitesh Saxena. 2022. Hearing Check Failed: Using Laser Vibrometry to Analyze the Potential for Hard Disk Drives to Eavesdrop Speech Vibrations. In *Proceedings of the 2022 ACM Asia Conference on Computer and Communications Security (ASIA CCS '22), May 30-June 3, 2022, Nagasaki, Japan.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3488932.3517399

1 Introduction

The magnetic hard disk drive (HDD) has been a widely trusted and implemented technology for over 60 years. Today, HDDs are used ubiquitously across the world in home and business settings and it is believed that they will remain popular for many more years to come [1, 2]. While HDD technology has advanced, it maintains the same basic function: the storage and retrieval of digital information per the user's command.

However, what if this technology were to also breach user's privacy? Specifically, what if an attacker could record the vibrations of an HDD's read/write head while sensitive speech is spoken nearby and possibly use the vibration data to infer speech (at least partially). For instance, the attacker can measure the vibrations of the read/write head by recording the position error signal (PES) data [3] (e.g., via a firmware malware) or by using some other vibration sensor attached to the head for shock detection purposes [4]. With sufficiently strong vibrations induced from external speech, these channels would allow an HDD to inadvertently act as a microphone device that records speech information, as demonstrated in prior research by Kwong et al. [5]. This attack holds a similar threat level to microphone surveillance, but can be even stealthier because the HDDs are already in place.

While previous work in this area is valuable and does demonstrate a hint at the potential of HDDs eavesdropping over speech, crucially, the underlying experimental settings seem limiting to

ascertain the true viability and generalizability of the threat [5]. **First**, the study was limited to experimenting on *one HDD*, a 1TB Seagate Barracuda 7200.12 HDD, from which the authors could identify the exposed AMUX pin needed to extract the PES data. **Second**, the only speech source tested in this study was a speaker device (*machine-rendered speech*) that played speech samples at *high loudness* levels (75 dB, 85 dB and 90 dB), well above the normal range for human conversation (40 dB to 60 dB) [6, 7]. Figure 1b depicts the single scenario tested in [5]. Therefore, the results of this prior work are likely anecdotal and as we come to find, the methodology is not reproducible or scalable for our purposes.

In this paper, we pursue a comprehensive study of the vibration impact to the read/write head of HDDs when exposed to external speech. As we want to measure the subtle vibrations that are the source of the threat, what technique could be intuitively better for such measurements than *laser vibrometry*? Thus, a **novel methodological approach** to our research is the use of a high-fidelity laser vibrometer to *investigate the limitations* of speech eavesdropping that uses vibrations of a read/write head. While other techniques exist, laser vibrometry provides a reliable method for measuring vibration data from emanations of acoustic signals. This methodology is highly robust to handle the depth and breadth of our research. Not only does the laser vibrometer measure with great precision, it is easily portable to measure different HDD models.

Using the laser vibrometry methodology, we dissect the studied threat along the following dimensions. **First**, for the assessment of the threat's generalizability, we use five HDDs of varying popular brands in our experimentation. **Second**, we explore a wide variety of eavesdropping scenarios with variable parameters: *(1) speech sources* (live human speech and machine-rendered speech), *(2) loudness levels* (normal (60 dB), loud (70 dB), and very loud (85 dB)), and (3) *transmission mediums* (aerial, shared surface, and direct contact), in order to gain a broader knowledge about the proposed threat. Assessing live human speech over the aerial medium in this context is extremely important as this represents the most natural and common scenario underlying the threat, which could put routine human conversations near HDDs at risk. We do not look to develop any particular attack instance, but rather aim to analyze whether or not such a threat is likely under realistic speech scenarios. ***To this end, we deliberately create an experimental setup that measures the potential for speech eavesdropping in more favorable settings than a real-world attacker would likely achieve.*** If eavesdropping results are poor in this set-up, we can assume they will be poorer in real-world settings, suggesting the lack of viability of such attacks. In this work we do not present a new or alternative attack methodology, but seek to explore the speech leakage potential of the vibration data that would be collected by a successful attack (i.e., PES, bugging, etc.).

Overall, the results from our study show that the chance of an HDD eavesdropping on speech information via vibrations induced on its read/write head is relatively low and applies to only some limiting scenarios. Our results also reconfirm the findings of the prior study [5] for their specific experimental settings mentioned above, based on our methodology.

Our Contributions: We summarize our key contributions and results below:

(1). Measuring Vibration Effects of Live Human Speech: We measured and observed the effect of vibrations induced by sound waves from live human speech on the read/write heads of multiple HDDs. Specifically, we defined the live speech scenario: *Live human-speech* (Figure 1a). The complete description of this scenario is found in Section 2.5. Comparing each measurement with a baseline control measurement of each HDD's natural frequency vibration (i.e., in the absence of speech signals) allowed us to determine any effects caused directly by the external speech. We performed both *time-domain* and *frequency-domain* analyses to identify what scenarios caused a clear vibrational effect. Our results indicate that live human speech, at normal conversational loudness, seems incapable of affecting the read/write head. This suggests that acoustic vibrations traveling via the aerial medium may not be strong enough at this loudness to leak speech information.

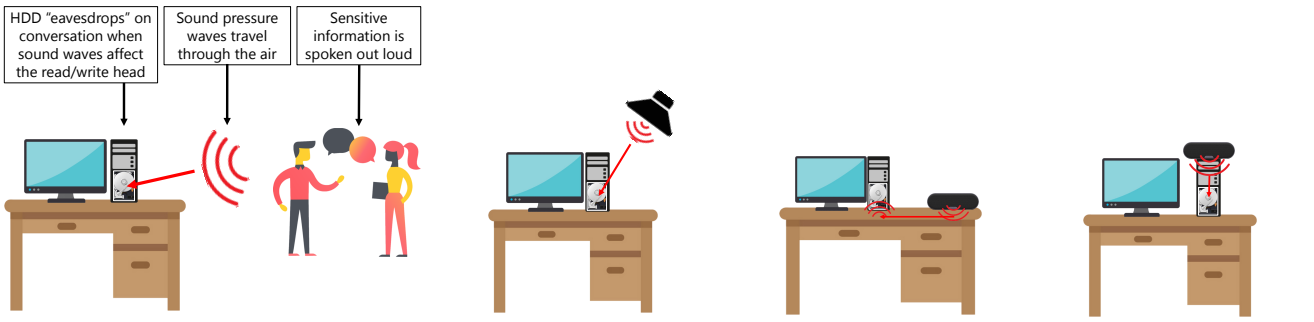
(2). Measuring Vibration Effects of Machine-Rendered Speech: We also observed the effects of vibrations induced by sound waves from machine-rendered speech through multiple propagation mediums and at multiple loudness levels. Specifically, we defined three machine-rendered speech scenarios. *Loudspeaker-Aerial* (Figure 1b), *Loudspeaker-Same-Surface* (Figure 1c) and *Loudspeaker-Touching* (Figure 1d). Full descriptions of each of the machine-rendered speech scenarios can be found in Section 2.5. Our results from the *Loudspeaker-Aerial* scenario indicate that even machine-rendered speech at normal loudness is not able to induce a vibrational effect on the read/write head of HDDs. Via the aerial propagation medium, vibrations induced by the external speech were only observed in the very loud (85 dB) setting. This observation also serves to recreate and reconfirm the results of Kwong et al. using an independent methodology. On the other hand, the *Loudspeaker-Same-Surface* scenario in the loud and very loud settings and the *Loudspeaker-Touching* scenario in all loudness settings (normal, loud and very loud) made clear impacts on the read/write heads. This suggests that only vibrations propagated via very loud speech, or through a shared surface or direct contact are strong enough to leak speech information. Table 1 provides a summary of our observed results.

(3). Laser Vibrometry as a Study Methodology: Our use of laser vibrometry to evaluate the limitations and practicality of vibration-based speech eavesdropping attacks that utilize vibrations induced on the read/write head of HDDs, is a novel contribution in our research. Laser vibrometers allowed us to measure subtle vibrations with a very high degree of precision and reveal information that would otherwise be unavailable. This is a broader methodology which may be used to assess the feasibility of other vibration-based side channel attacks (e.g., [8–10]). The use of this technique provided another unique advantage for our current research because the vibrometer is easily portable to different HDD models. Our recreation of the results of [5] serves to further confirm the effectiveness of our methodology (elaborated in Section 3).

2 Preliminaries & Attack Scenarios

2.1 Principles of Vibration and Sound

Vibration and sound are very closely related concepts. Vibrations generate sound in the form of pressure waves that propagate through the air. Conversely, these pressure waves can induce vibrations in



(a) *Live-Human-Aerial*: aerial sound waves from live human speech.

(b) *Loudspeaker-Aerial*: aerial sound waves from loudspeaker device.

(c) *Loudspeaker-SameSurf*: sound waves propagated through a shared surface with the loudspeaker device.

(d) *Loudspeaker-Touching*: sound waves propagated directly from loudspeaker device to HDD enclosure.

Figure 1: Our work studied different HDD speech eavesdropping scenarios in three volume settings (normal, loud, and very loud). *Kwong et al. only studied the Loudspeaker-Aerial scenario in the loud and very loud volume settings.

structures they encounter. A prime example of these concepts can be seen in how humans speak and hear. Vibrations of our vocal chords produce the sounds that we use to speak. Our ability to hear these sounds comes from the vibrations induced in our eardrums by the sound waves. Therefore, any structure that can record and interpret sound waves is effectively a listening device. Microphones are the main example of this as they are designed similar to the human eardrum. A microphone device contains an internal diaphragm that reacts to changes in air pressure caused by sound waves that propagate aerially. The vibrations induced on the diaphragm are transformed into an analog signal and output by the microphone.

2.2 Natural HDD Vibrations

Vibrations in HDDs can pose serious problems, affecting the position of the head and causing read/write errors [11]. Many factors have been considered for the design of HDDs in order to handle any internal vibrations. There are 3 main *internal* sources that can induce vibrations on the read/write head of a HDD [12]. First, platter wobble is caused by a small imbalance in the rotating disc platters which will induce vibrations as the disk rotates. Second, the head seek action involves the acceleration of the actuator arm at very high speeds to position the head, causing a resultant force to be applied into the HDD (e.g., inducing vibrations). Appendix Figure 7 depicts a force diagram of an HDD and how these two sources can propagate vibrations through the HDD. The last source for internal vibrations is the computer's fan. The fan unit contains a spinning component that induces some vibration into the HDD that is housed in the same compartment. Since our experiments remove the HDD from the computer casing, this source of vibration does not apply to our experimental setup. Additionally, as we design our experiments to measure an HDD that maintains an idle read/write head position, the acceleration of the actuator arm is a source of vibration that does not apply to our work. Therefore, the only internal source for vibrations that introduces variability among our tested HDDs is *platter wobble*.

2.3 Threat Model

For our study, we consider a threat model in which an attacker has, or can gain, access to the vibration data of the read/write head of a target HDD. An attacker like the one described in [5] can gain root privileges on an HDD and retrieve the PES data. The PES tracks the displacement of the HDD's read/write head from its intended track on the disk platter [13]; allowing it to measure any external vibrations induced on the head that would cause displacement. The attacker could re-flash the HDD's firmware to expose the PES or affect the HDD physically while it is in transit and inject malware or conduct a machine-in-the-middle attack to update the firmware. It is also assumed that digital signatures are not used on the HDD and that the attack is Operating System independent. The objective of the attacker is to glean vibration data about the read/write head and use it to reconstruct the human speech spoken in the vicinity of the HDD. Alternatively, a malicious attacker could utilize some other sensor technology, such as a vibration sensor used for shock prevention [4], to record vibration data from the read/write head of the victim HDD. The defining characteristic of our attacker is the ability to measure the vibration leakage of the read/write head – by any viable method available.

2.4 Experimental Attack Parameters

Sound Pressure Level: The loudness of a sound, or Sound Pressure Level (SPL), is measured in decibels (dB). The SPL of *Normal* conversation is estimated between 40 dB to 60 dB [6, 7]. Therefore, any noise above 70 dB may be considered *Loud* in terms of human conversation. As in previous work [5], greater dB levels were explored; so we analyze sounds $\geq$ 85 dB, termed the *Very Loud* setting.

Speech Sources: The two main categories of speech are *live human* and *machine-rendered*. Live human speech includes original speech produced from the vocal chords of a human. Machine-rendered speech refers to a speaker device playing the audio of human speech. Both sources are very similar and can project acoustic sound waves that can be interpreted by listening devices such as a microphone or the human ear. Therefore, both sources have a similar potential to leak the same speech information. For our methodology, we designated a live human speaker and a loudspeaker device for the live human and machine-rendered speech sources, respectively.

Sound Wave Transfer Mediums: To understand the effect of sound waves on the read/write head of HDDs, we considered the medium used to travel to the HDD and how different mediums

may induce different effects. We defined three transfer mediums with different efficiencies for energy transmission and observed how the vibrational impact of sound waves differed. The *aerial* medium involves sound waves traveling through the air and is the least efficient at transmitting speech energy. This represents the natural medium of human speech as people talk to each other. We also defined the *same surface* medium to represent the scenario of a speaker device sharing a surface with an HDD which is more efficient at energy transmission than the aerial medium. Vibrations from the machine-rendered speech of a speaker device can propagate along the shared solid surface and potentially induce a greater vibrational effect than the aerial medium. Lastly, we considered the most severe scenario of propagated vibrations and defined the *touching* medium in which a loudspeaker device playing speech is in physical contact with an HDD. This medium has the highest efficiency for energy transmission because the vibrations from the speaker device propagate directly without dampening.

2.5 Experimental Attack Scenarios

We conceptualized different scenarios to investigate the effect of sound wave vibrations induced on read/write heads. Particularly, we considered the three test parameters described above: 1) Sound Pressure Level, 2) Speech Source, and 3) Transfer Medium. Some of the scenarios in our experimental design mimic the designs used in [5] in which machine-rendered speech is played aerially through a loudspeaker device in a loud or very loud SPL setting. Expanding on this research, we considered factors not previously studied.

Live Human Speech: We designed a human speaker scenario in which the speaker talks near the HDD. As the speech originates from a live human speaker, the transfer medium in this scenario is the air. Therefore, we termed this scenario as *Live-Human-Aerial* and illustrated it in Figure 1a. To determine the threat to regular human conversation, the live speech is spoken in the normal SPL range. This setup mimics a normal conversation between humans in the presence of an HDD and best represents the real-world threat that would be faced.

Machine-Rendered Speech: We also designed a set of three machine-rendered speech scenarios in which a portable loudspeaker device is the speech source. One scenario is defined for each propagation medium, termed; *Loudspeaker-Aerial*, *Loudspeaker-Same-Surface* and *Loudspeaker-Touching*. The key advantage that a loudspeaker device has over a live human speaker is that it can reach volumes above the range of the human voice, allowing the device to project stronger sound waves. Thus, machine-rendered speech can achieve all three SPL settings.

The first machine-rendered speech scenario that we defined is the *Loudspeaker-Aerial* scenario. This scenario refers to speech audio played through a loudspeaker device and traveling aerially to reach the HDD and is depicted in Figure 1b. This scenario mimics the aerial medium of live human speech and represents a situation in which a speaker device (i.e., smart phone, portable speaker) is held in the hand of the user as sound waves travel to the HDD.

Next, the *Loudspeaker-Same-Surface* scenario involves speech audio played through a loudspeaker device that shares a common solid surface with the HDD and is shown in Figure 1c. This scenario reveals the effects of a solid transfer medium on the propagation of sound wave induced vibrations. An example of this would be if a person in their office placed their cellphone on their desk before playing a voicemail. If both the cellphone and the computer are on the same desk, there is potential leakage via the vibrations propagating from the phone to the computer.

Lastly, we define the *Loudspeaker-Touching* scenario in which the loudspeaker device playing speech audio is in physical contact with the HDD. This scenario is depicted in Figure 1d. Although less likely to occur in a practical setting, this scenario was used to investigate how directly propagated vibrations impact the HDD.

3 Our Methodology

Overview of Laser Doppler Vibrometry: The Doppler effect refers to the change in a wave's frequency as it encounters a moving object [14]. As the object moves closer or farther from the source of the sound wave, the received wave frequency either increases or decreases respectively. In the implementation of a laser vibrometer (LDV), the frequency of the light beam shifts in proportion to its velocity as it is reflected off of a moving object. The velocity information is recorded in the frequency and measured by the LDV. Through this process, an LDV can be used to measure vibrational displacement. Additionally, the standard USB data acquisition system used with an LDV device can process the voltage signal generated by the interferometer and digital decoding electronics. The signal is created by converting the wave frequency shifts. Laser Doppler vibrometry is currently used in many applications and fields of research because it offers the highest resolutions for vibration measurements [15].

Our Implementation: For each of the five HDDs used in our study, we implemented the methodology described in [5] for identifying the exposed PES pin. We connected to each HDD's serial diagnostic port and toggled the AMUX signal ON and OFF while using an oscilloscope to observe the output from each exposed pins of the circuit board. A diagram depicting this set of experiments can be found in Appendix Figure 8. Unlike the work in [5], we found that none of the exposed pins on our HDDs output the expected signal or changed values when we toggled the AMUX command. The lack of pin access across the HDD models we selected inspired us to explore another methodology.

We chose to use laser vibrometry to collect vibration data in our study for significant reasons. First, the LDV technology is highly portable to work for the generalized scope of our study. Secondly, the LDV devices we use (detailed in Section 4.1), have very high accuracy and resolution in comparison to other known techniques (such as PES). The standard width of a single track on an HDD platter, measured in the radial direction, is approximately 200-250 nm [16]. With displacement resolutions of 0.1 pm (or 0.0001 nm) and 0.3 pm (or 0.0003 nm) [17, 18], we determined that two of the vibrometer setups we used are more than sufficient for detecting the vibrations of the read/write head that would be recorded by other state-of-the-art methods. To take our measurements, the read/write head of the HDD had to be exposed. In our experiments we revealed the head by removing the front casing of the HDD and measuring while the head was in an idle position.

As it was used in [5], the PES readings are just linear modulation values that can be written to .WAV format. This is the same method

used by microphones when converting analog data (vibrations of the diaphragm component) to a digital audio signal. Therefore, an LDV device is a viable alternative for collecting vibration data for our feasibility study. We are not claiming LDV can be used in place of PES to achieve the same attack. We acknowledge that any vibration sensor could be used to capture the same information as the PES (raw vibration data), but chose LDV for its superior sampling frequency and portability to measure multiple HDDs. Notably, we are not presenting our methodology for an actual attack, but simply as a means to capture the same information (or better) as the attack previously introduced in [5] so that we may assess what speech leakage, if any, exists.

4 Experiments and Data Collection

4.1 Vibrometer Equipment Used

We utilized three different models of laser vibrometers supplied by the company Polytec; the PDV-100 Portable Digital Vibrometer [19], the OFV-5000 Modular Vibrometer [17] and the Vibroflex vibrometer [18]. The PDV-100 vibrometer can measure vibrational velocity in the 0-22kHz range and is therefore sufficient for capturing speech signals in the audible frequency spectrum. We used the PDV-100 vibrometer for initial experiments to establish our baseline for measuring the vibrational effects of speech signals. We expanded its use to collect initial data from one HDD for all of the threat scenarios that we considered.

The second model we used was the OFV-5000 modular vibrometer. Specifically, we used the OFV-5000 controller paired with the MLV-I-120 sensor head, the VX-08 decoder, and the MLV-O-SRI short distance lense. This setup introduced greater levels of measurement sensitivity, and can also measure vibrational velocity in the 0-22kHz frequency range. This model has a displacement resolution of 0.1 pm. We use the OFV-5000 setup for the majority of our data collection so that we can accurately determine if the presence of speech can cause a vibrational effect on the read/write head to the extent that speech information could be leaked.

The third model that we used was the Vibroflex vibrometer with the VFX-I-120 sensor head, the VX-08 decoder, and the VFX-O-SRI short distance lense. This setup has similar precision as the OFV-5000 setup with a displacement resolution of 0.3 pm. We used the Vibroflex setup to collect data from different HDDs in the loud and very loud settings from [5] to confirm our methodology by recreating prior results.

4.2 Hard Disk Drives Tested

We tested our scenarios using five HDDs of varying models and sizes. We chose standard HDDs from popular manufacturers that are readily available from all major vendors and are similar to the one used in [5]. Additionally, we acquired the exact same HDD model that was used in [5] to obtain results for comparison. We used the Seagate Barracuda (250 GB) [20], Seagate Barracuda (80 GB)[21], Seagate Barracuda 7200.12 (1TB - from [5]) [22], Hitachi Deskstar (80 GB) [23], and Fujitsu mini (120 GB) [24] HDDs. Each of the HDDs are SATA type and use standard spinning disk platters and read/write heads. Appendix Section A.1 contains all physical specifications for the HDDs in Table 3.

Giant Magnetoresistive (GMR) head type is currently the most common read/write head type found in magnetic HDDs. SMR is the next closest type but it adds another complexity which could cause some resistance in being adopted faster and replacing GMR heads. Additionally, the move from longitudinal (LMR) to perpendicular magnetic recording (PMR) technology has influenced new read/write recording methods. Currently, there are two types of PMR; conventional PMR (CMR) and shingled magnetic recording (SMR). These versions have a similar recording type but isolate read/write operations between adjacent tracks differently. Unlike CMR, HM-SMR heads are not drop-in replacements for traditional drives and require system software modifications [25, 26]. Therefore, for our study we selected HDDs that have GMR heads with a PMR recording method. Advanced head type technologies, including TDMR, HAMR and MAMR, are still evolving and do not have enough market presence yet for a generalizability study. Challenges like disk material, heat dissipation, and complex production cycles have not been completely resolved [27, 28].

4.3 Experiment Setup

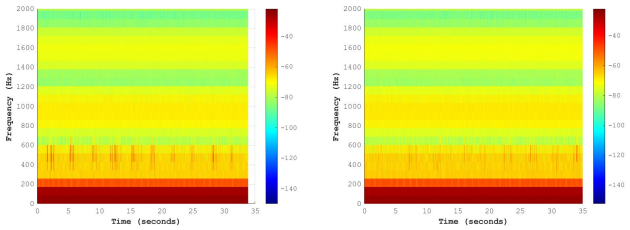
For our experimental attack model, we defined the four different scenarios described in Section 2.5. The *Live-Human-Aerial* scenario required the human speaker to read the audio sample transcription at the loudness level of normal conversation. A digital sound level meter was held above the HDDs read/write head to ensure the speech loudness remained around 60 dB at the point of measurement. We recruited a Male volunteer as our live human speaker and he was instructed to talk in the direction of the HDD with his mouth at a 0.3 meter distance for all experiments.

For the first machine-rendered speech scenario, *Loudspeaker-Aerial*, the portable loudspeaker was held up by the experimenter, at a distance of 0.3 meters, and directed at the HDD. All experiments for this scenario were done in a consistently similar setting in regards to the loudspeaker's position and the evaluators involvement. In a practical situation, human speech and loudspeaker output is not directed towards an HDD; but rather towards the other person participating in the conversation or out into the open space. This would cause dampening of the sound waves so we directed our external speech at the HDD in order to *maximize the possible impact from the sound waves*.

The second machine-rendered speech scenario, *Loudspeaker-Same-Surface*, introduced a shared surface between the speech source and the HDD. Our experimental setup placed the portable loudspeaker device on the same table as the HDD, at a distance of 0.3 meters. Again, we pointed the loudspeaker towards the HDD to maximize the vibrational effect.

The third machine-rendered scenario, *Loudspeaker-Touching*, represented the greatest threat instance where the vibrations from the loudspeaker propagate directly into the HDD without having to pass through other decoupled materials or the air. The setup placed the loudspeaker on its back, with speakers facing up, and the HDD sitting on top. This ensures the most severe propagation of vibrations into the HDD. Appendix Figure 9 provides images of our experimental setup for the machine-rendered speech scenarios.

For each machine-rendered scenario described above, the laser vibrometer (PDV-100 or OFV-5000 for experiments with normal



(a) Female Speaker (b) Male Speaker

Figure 2: Frequency Spectrum graphs comparing vibration measurement samples for the Harvard Sentences F1 (Female) and M25 (Male) speech samples in the Loudspeaker-Touching (Loud SPL) scenario.

loudness speech; PDV-100, OFV-5000, or Vibroflex for experiments with loud speech; Vibroflex for experiments with very loud speech) was attached to a tripod stand and positioned above the HDD, facing downwards. The vibrometer's laser was pointed directly at and focused onto the topmost read/write head of each HDD. Although some HDDs tested had multiple, stacked read/write heads, the mechanical coupling of the heads should result in an even propagation of induced vibrations throughout. Therefore, measuring the topmost head does not introduce a bias. Inline with the set-up used in HDD natural vibration measurement studies via vibrometry [29], the laser intersected the read/write head at the perpendicular angle to minimize the measurement error margin. For each experiment, the read/write remained idle in a fixed position. This scenario is more favorable for the feasibility of the attack as the read/write head is more affected by external vibrations. The vibrometer measurement was manually started and stopped to encompass the entirety of each speech sample played or spoken. Again, the digital sound level meter was held above the read/write head in each machine-rendered scenario to ensure the speech audio was in the correct decibel range for each SPL setting that we defined.

We had to expose the disc platters and read/write head of each HDD to perform our experiments. This was done by either completely or partially removing the front casing of the HDDs. None of the HDDs used were filled with Helium so typical operations were not affected by this setup. We recognize that removing the front casing of the HDD makes it more vulnerable to background noise that could affect its ability to act as a microphone. However, we believe that in this case any such background noise would not affect the results of our study as we intentionally injected noise around the HDD. Therefore, any minor background noise would have a negligible effect on the read/write head.

The scenario tested in these experiments (i.e., eavesdropping on HDD with front case removed) is *not intended to recreate a real-world scenario*, but rather to create a favorable scenario where speech vibrations are induced at a higher magnitude, and measured with greater precision, than would likely occur with a real-life attacker.

4.4 Data Collection

All data collection was performed in a quiet office space in order to minimize the effects of any external noise. Data collection was performed both in the presence and absence of speech samples in order to establish a baseline for future comparison in our analyses.

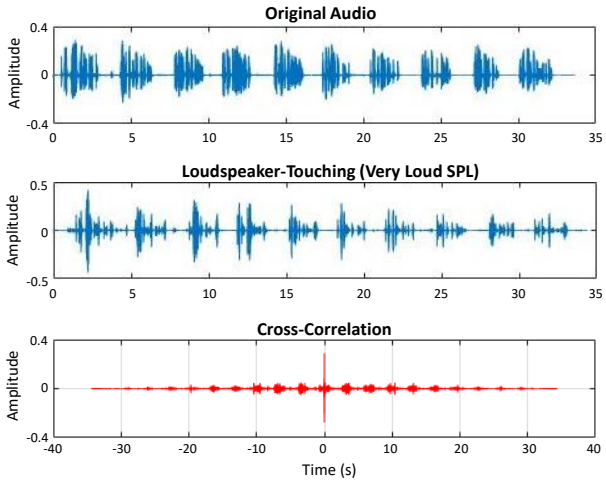
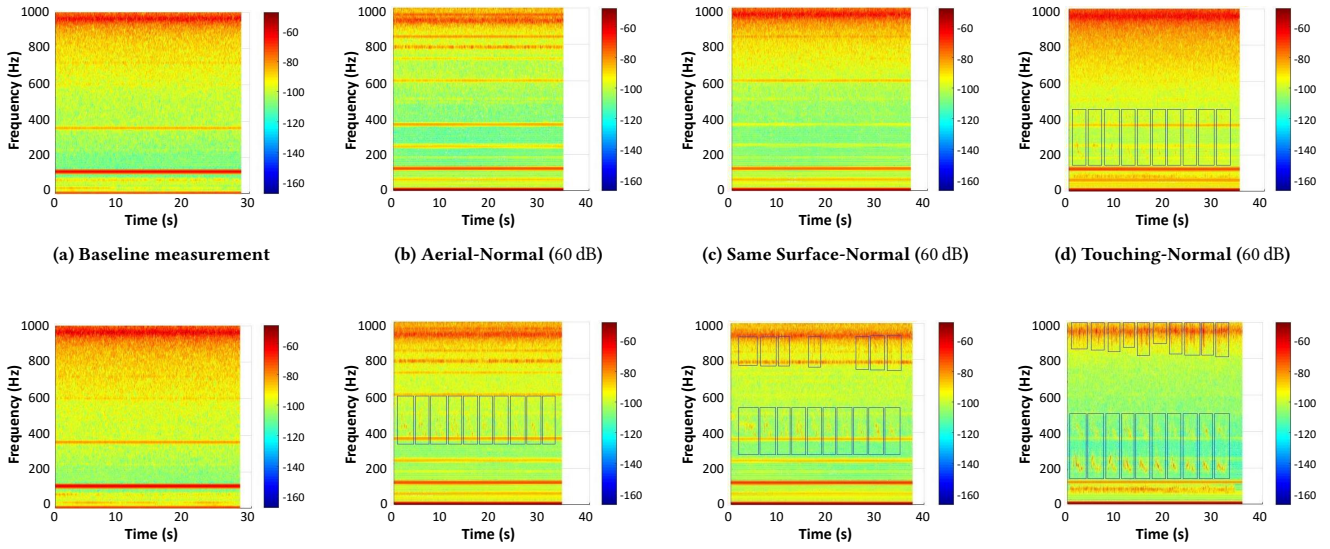


Figure 3: The top graph is the time domain of the original audio. The middle graph is the time domain of the reconstructed audio. And the bottom graph is the cross-correlation between the two signals. Notice the peak in the cross-correlation graph at lag=0 which indicates some correlation between the signals (i.e., some amount of speech information may be contained in the signal). *Graphs generated from Seagate Barracuda 7200.12 1TB HDD data.

Speech Dataset: We used a collection of Harvard sentences sample phrases from the IEEE Recommended Practices for Speech Quality Measurements [30]. These speech samples were chosen because the sentences recorded are all phonetically balanced to use specific phonemes at the same frequency that they appear in the English spoken language. The set of Harvard sentences have been widely used in standardized testing of telephone, cellphone, and Voice over IP systems. The sentences are divided into 72 lists that each contain 10 phrases/sentences. The provided speech sample recordings are per list, meaning each sample contains a single speaker saying each of the 10 phrases from that list. We utilized the recordings of three of these lists (F1-F3) from one female speaker for a total of 60 recorded sentence samples for the machine-rendered speech experiments. Additionally, the Male live human speaker read the transcription of these speech samples for the Live-Human-Aerial experiments. We performed an experiment to compare female and male speaker samples from the Harvard sentences dataset. Figure 2 shows the frequency spectrum graphs generated for the F1 and M25 samples in the Loudspeaker-Touching (Loud SPL) scenario. We can see that a similar frequency range is captured for both samples, but the female speaker frequencies related to the speech are slightly stronger. We believe this may be due to the higher frequencies present in female speech which make them more distinguishable among the other vibrational noise that is induced. Therefore, we focus our experiments on a set of female speech samples in order to observe greater levels of speech leakage. However, in real-world settings we believe the vibration-based eavesdropping attack can have similar potentials for success for both female and male speakers.

5 Signal Analysis of Collected Data

The speech samples used for each scenario consisted of 10 different sentences spoken in succession with an average total sample length of around 30 seconds. With a vibrometer sampling frequency of 44kHz, approximately 1.3 million raw data points were collected for each measurement. Vibration data was measured in the time



(a) Baseline measurement (b) Aerial-Normal (60 dB) (c) Same Surface-Normal (60 dB) (d) Touching-Normal (60 dB)

(e) Live human speech (f) Aerial-Loud (70 dB) (g) Same Surface-Loud (70 dB) (h) Touching-Loud (70 dB)

Figure 4: Frequency spectrum graphs of the Hitachi 80 GB HDD. We identify some scenarios that likely contain information leakage: Loudspeaker-Touching scenario in the normal SPL setting, and all scenarios in the loud SPL setting, contain distinct frequency markers (outlined in blue boxes) that correspond to the 10 spoken sentences in the original audio.

domain and stored as raw ASCII data (referred to as vibration signal) and .wav sound file conversion (referred to as reconstructed audio). Before performing our different components of analyses, we implemented data pre-processing techniques to enhance any speech captured by the vibrometer measurements, and to reduce any background noise that was present in the signal.

Time-Domain Analysis of Reconstructed Audio: The laser vibrometer equipment used can measure vibration in both the time and frequency domains. For our research, we measured vibration in the time domain and stored both the raw time domain data and the generated time domain graphs. To recognize the effects of the external speech, we first measured the natural vibrations of the HDDs (i.e., in absence of speech). Comparing the amplitudes of vibration measurements in the time domain between the scenarios with and without speech allowed us to identify where, if at all, the external speech induced an effect.

To pre-process the data, we considered speech enhancement routines from the speech processing toolbox, VOICEBOX [31], implemented in Matlab. Specifically, we were interested in the routines *specsub*, *spendred*, *ssubmmse*, and *ssubmmsev*. The *specsub* routine performs speech enhancement using spectral substraction. The *spendred* routine performs speech enhancement and dereverberation. And, the *ssubmmse* and *ssubmmsev* routines use minimum-mean square error (MMSE) criteria for speech enhancement with *ssubmmsev* additionally implementing voice activity detection (VAD) based noise estimations. Running each routine on the converted speech samples and comparing the filtered samples, we determined the best performing routine. Specifically, the *ssubmmse* routine had the greatest speech enhancement and static noise reduction. Therefore, *ssubmmse* was used for noise filtering and we refer to the post-processed samples as "enhanced" audio.

For our time domain analysis, we compared the time domain graphs of the original audio played in our experiments to the time domain graphs of the enhanced audio reconstructed from our vibrometer measurements. Our analysis starts in the time domain because it allowed us to compare the presence of speech features in the raw signals. If speech features are strong enough in the reconstructed signal, we would find consistent patterns that match the original speech signal. Here, we will look for peaks (caused by speech) in the enhanced signal that align with the peaks in the original signal. For the 10 different sentences in each audio file, we would expect to see 10 different peaks in the enhanced signal that correspond to the 10 sentences in the original file. An example of this comparison is shown in Figure 3. In the enhanced signal we can clearly see the 10 unique peaks throughout the signal that align with the 10 peaks in the original audio.

Frequency-Domain Analysis of Vibration Signal: For our frequency domain analysis, we utilized the raw vibration data collected from the vibrometer because it contained the full spectrum of captured frequencies. With this, we were able to generate the full spectrum graphs for each measurement to identify where the relevant frequency markers appear – we identified that vibrations induced on the HDD by the external audio appeared within the 150-1000Hz frequency range. Therefore, we focused on this frequency band for our analysis. We continued our analysis in the frequency domain because speech signals can be obfuscated such that the characteristic patterns of the original speech are no longer identifiable in the time domain. However, an obfuscated signal may still contain frequencies unique to the original speech that could be used to extract certain speech information. Therefore, we found it important to explore both the time and frequency domains to fully understand the potential for speech information leakage.

Matlab provides convenient tools for visualizing the frequency spectrum of time domain data. Therefore, we generated frequency spectrum graphs for additional analysis. The spectrum graphs are heat maps of the present frequencies in each scenario and were

used to easily identify any frequency markers directly caused by exposure to the speech signals. Again, we looked for 10 distinct frequency signatures that correspond to the 10 sentence utterances in the original audio. Figure 4 shows the frequency spectrum graphs for each experiment scenario from the Hitachi HDD. We can see that some of the tested scenarios contain clear frequency signatures in the spectrum graphs that may indicate information leakage.

Cross-correlation Analysis of Input vs. Reconstructed Audio: To confirm observations from the time domain graph and frequency spectrum graph comparisons, we must use a quantifiable measure. Cross-correlation can determine how similar two signals are so we implement this measure to compare our reconstructed samples with the original speech audio. Similarly, we compare an audio sample of natural HDD vibrations (no speech information) with the original audio to determine a baseline correlation. We reason that if the correlation between our reconstructed samples and the original speech audio is similar to the correlation between natural vibration noise and the original speech audio, it will indicate a lack of speech information. Before determining the similarity between the samples, we further processed the enhanced audio files described above to isolate any existing speech characteristics for the cross-correlation analysis. For this, we applied a low bandpass filter on the enhanced audio to capture frequencies between 150-1000 Hz. We also re-sampled the original audio so that it would have the same sampling rate as the enhanced audio file. Lastly, we aligned the signals before performing cross-correlation.

We computed the cross-correlation between the time domain data of the original and enhanced audio samples. The graphs included in this paper were generated for measurements using the Harvard Sentence Set, F1 audio. We use the graphs for the F1 audio as a concrete example throughout the paper, but the other data collected using the Harvard Sentences F2 and F3 samples as the original audio achieved the same results. Cross-correlation computes the dot-product of two signals as a function of time. As a result, the sliding nature of the algorithm will obtain the maximum output value when the peaks and troughs in each of the signals best align with each other. Since we aligned our signals before computing the cross-correlation, we would expect to see a peak in the correlation graph at the point lag=0. In the cross-correlation graphs in Figures 12c & 12d we can see the large peak indicating strong signal correlation. Our cross-correlation analysis is in line with the signal analysis performed in [5].

Speech Intelligibility Metric Analysis: As an additional step to evaluate the intelligibility of speech in our recovered samples, we utilized two different speech intelligibility metrics. The Perceptual Evaluation of Speech Quality (PESQ) is a measure for assessing speech quality in telephone networks introduced by Rix et al. [32], and used in [5]. PESQ has a value range of [0 - 4.5] with 4.5 indicating the highest quality of intelligible speech. We also used Short-term Objective Intelligibility (STOI) from a work by Taal et al. [33]. This metric is designed for short clips of time-frequency weighted speech audio in noisy environments and has a value range of [0 - 1] with 1.0 indicating the highest quality of speech. We calculated scores for sentence samples collected from each experimental scenario. We selected the highest quality sample, among the Seagate Barracuda samples, for each sentence set. Each of the 10 sentences in the selected samples were isolated and received individual metric scores for a total of 30 scores averaged per scenario.

Automatic Speech Recognition Analysis: To assess the potential for machine-recognition, we performed an Automatic Speech Recognition (ASR) analysis of our enhanced audio samples. Using the online Google Speech-to-Text (STT) interface [34], we input a selection of our enhanced samples that performed best in our live human analysis. We determined ASR success by how well it could transcribe the original speech in our enhanced samples.

Specialized Classification Analysis: We conducted an exploratory analysis of the machine learning classification potential of our recovered samples for achieving speech recognition. We focused on a specific setting for each speech source/propagation medium. Specifically, we compiled 10 samples of the F1 sentence set, across multiple HDDs, for each of the selected scenarios. The samples were all processed to isolate each of the 10 sentences, resulting in 10 samples per sentence (100 total samples per scenario). We generated Mel-frequency cepstrum (MFCC) feature sets for each sample because they are the most commonly used for speech recognition systems. We tested four learning models (Naïve Bayes, Logistic Regression, MultiClass, and Random Forest) using 10-Fold CV.

6 Results

From our analysis, we note the significant observations made and discuss those results in this section. We organize our results into 3 sub-sections: (1) scenarios that showed information leakage, including scenarios that recreate the results of [5]; (2) scenarios that did not show information leakage, and (3) our human intelligibility study. A higher level summary of our overall results is depicted in Table 1, and the detailed results are discussed in this section. In Table 1, we say that information leakage for a particular setting is *Not Likely* if information leakage could not be identified for any of the tested HDDs, *Likely* if information leakage can be proven for at least one, and up to half, of the tested HDDs, and *Very Likely* if proven for more than half of the tested HDDs. Due to space limitations, only the most significant graphs/figures supporting our analysis/results are presented in the paper as illustrative examples (and some in appendix); for full set of graphs, we refer to our website: https://sites.google.com/view/hearing-check-failed. In all cases, our results have been consistently validated via both our quantitative (correlation-based) analysis and qualitative (inspection-based in time and frequency domains) analysis.

6.1 Leakage Present / Recreated Prior Results

Through our research, we identified some of the tested scenarios in which information leakage was present and verifiable. In the Loudspeaker-Aerial scenario, we observed that information leakage was likely when the external speech was in the loud SPL setting (>70 dB). This is one of our insights that was empirically demonstrated in our work via the LDV methodology. In this scenario we observed vibration responses in two of the tested HDDs (Seagate (250 GB) and Hitachi) that were confirmed with our cross-correlation analysis. Examples of the vibration responses we looked to find in the spectrum graphs from this scenario are displayed in Figure 4f. In the Loudspeaker-Same-Surface scenario, we also observed that information leakage was likely (occurring in the same two HDDs) when the external speech was in the loud SPL setting

Table 1: Summary of results for the different experimental parameters tested. *Scenarios labeled "Not Likely" did not indicate leakage in any of the HDDs tested in our study. The bold text identifies experimental parameters that recreate the setup in [5].

Speech Source	Transfer Medium	Loudness Level	Information Leakage?	HDDs Used	Vibrometers Used
Live Human	Aerial	Normal (60 dB)	Not Likely	Seagate (80 GB), Seagate (250 GB), Hitachi, Fujitsu	PDV-100, OFV-5000
Loudspeaker Device	**Aerial**	Normal (60 dB)	Not Likely	Seagate (250 GB), Hitachi, Fujitsu	PDV-100, OFV-5000
		Loud (70 dB)	**Likely**	Seagate (80 GB), Seagate (250 GB), Hitachi, Fujitsu, **Seagate 7200.12**	PDV-100, OFV-5000, Vibroflex
		Very Loud (85 dB)	**Likely**	Seagate (80 GB), **Seagate 7200.12**	Vibroflex
	Shared Surface	Normal (60 dB)	Not Likely	Seagate (80 GB), Seagate (250 GB), Hitachi, Fujitsu	PDV-100, OFV-5000
		Loud (70 dB)	Likely	Seagate (80 GB), Seagate (250 GB), Hitachi, Fujitsu, Seagate 7200.12	PDV-100, OFV-5000, Vibroflex
		Very Loud (85 dB)	Very Likely	Seagate (80 GB), Seagate 7200.12	Vibroflex
	Direct Contact	Normal (60 dB)	Very Likely	Seagate (80 GB), Seagate (250 GB), Hitachi, Fujitsu	PDV-100, OFV-5000
		Loud (70 dB)	Very Likely	Seagate (80 GB), Seagate (250 GB), Hitachi, Fujitsu, Seagate 7200.12	PDV-100, OFV-5000, Vibroflex
		Very Loud (85 dB)	Very Likely	Seagate (80 GB), Seagate 7200.12	Vibroflex

and confirmed this observation with our cross-correlation analysis. Figure 4g shows examples of the frequency spectrum graphs from the Loudspeaker-Same-Surface scenario that indicate information leakage. For the Loudspeaker-Touching scenario, we found that information leakage was very likely in both the normal and loud SPL settings. The time domain graphs and frequency spectrum graphs for the data collected in both loudness settings showed distinct peaks/frequency markers that corresponded to the 10 sentence utterances in the original file. Figures 4d & 4h and Appendix Figures 11e and 11h show the clear frequency signatures induced in the Loudspeaker-Touching scenario. Cross-correlation graphs for this data confirm the information leakage.

As was reported in the study of Kwong et al., our experiments showed that speech emanating from a loudspeaker device, at 85 dB, and traveling through the air can induce vibrations on the read/write head of a Seagate Barracuda 7200.12 HDD and leak speech information. We recreated this setup (specifically the loudness level of the audio) in our work as the Loudspeaker-Aerial scenario in the *very loud* SPL setting. We visually inspected the speech presence in this scenario from our time and frequency domain analyses, and went on to confirm this with our cross-correlation analysis that revealed a large peak in the cross-correlation graph, indicating a strong correlation between the reconstructed and original signals. Therefore, we were successfully able to recreate the results of Kwong et al. and show the presence of speech information leakage at the dB level that they used in their experiments. The time domain graphs for the Loudspeaker-Aerial scenario (alongside time domain graphs of the original audio, control measurement, and live human speech scenario data for comparison) are shown in Appendix Figure 10. This confirms that our LDV methodology can achieve the same results as PES, validating our other data.

Lastly, the data collected in the very loud SPL setting for the Loudspeaker-Same-Surface and Loudspeaker-Touching scenarios revealed speech information leakage was very likely. These observations were confirmed via the cross-correlation calculations – we see peaks at lag=0 in the graphs. The bottom graph in Figure 3 shows an example of a cross-correlation graph generated from Loudspeaker-Touching data in the very loud SPL setting, and the large peak at lag=0 indicating strong correlation.

Table 2: Average speech intelligibility metric scores calculated for each speech source-propagation medium scenario. Metric scores for raw microphone recordings (no noise, fully intelligible) are included for comparison.

Speech Source	Prop. Medium	SPL Level	PESQ	STOI
Live Human	Aerial	Normal	1.4	0.27
Loudspeaker	Aerial	Normal	1.4	0.28
		Loud	1.5	0.26
		Very Loud	1.5	0.30
	Same Surface	Normal	1.4	0.27
		Loud	1.7	0.25
		Very Loud	1.7	0.36
	Touching	Normal	1.8	0.52
		Loud	1.6	0.48
		Very Loud	2.1	0.57
Microphone Recording (Raw)		Loud	3.5	0.93

6.2 Leakage Absent

As our experiments encompass a broad array of attack scenarios, we also identified some scenarios in which information leakage appears to be absent in the vibrometer collected data. Most notably, we found information leakage in the Live-Human-Aerial scenario was not likely. None of the HDDs tested in this scenario showed any indication of speech presence in their time domain and frequency spectrum graphs. This was further supported with cross-correlation calculations from which we confirmed there were no peaks in the correlation graph, indicating that the reconstructed audio is not correlated to the original audio (i.e., no information leakage). Again, this is a new insight that we empirically demonstrated with many varying parameter settings and the LDV methodology. Figure 5 shows the cross-correlation graphs generated from each HDD in the Live-Human-Aerial scenario.

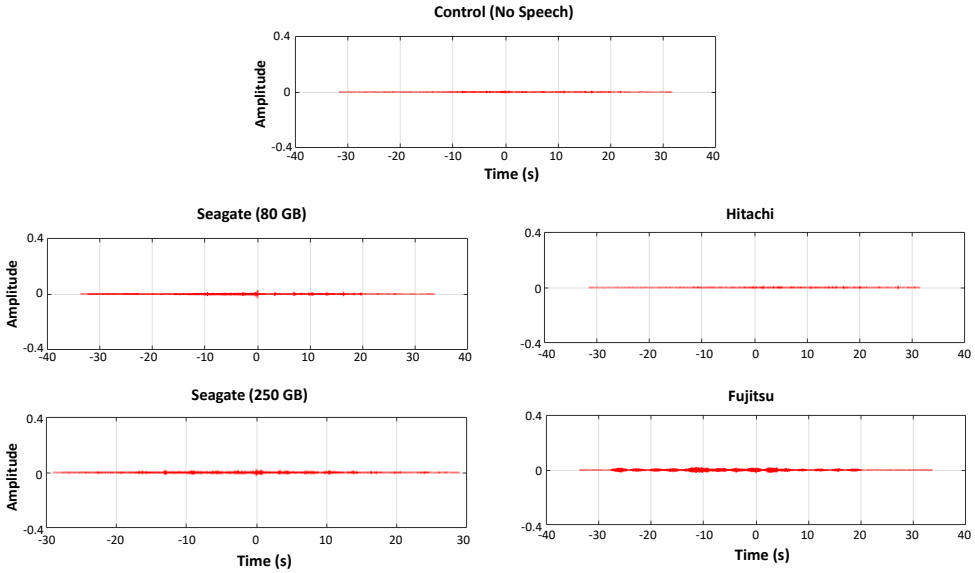


Figure 5: Cross-correlation between original audio and reconstructed audio from the *Live-Human-Aerial* scenario of each HDD. The correlation between the original audio and a control (No Speech) measurement, taken from the Seagate Barracuda 80 GB HDD, was included for comparison. These graphs use the same y-axis scale as the correlation graph in Figure 4 to allow for peak size comparison.

Similar to our observations of the Live-Human-Aerial scenario data, the data collected in the Loudspeaker-Aerial and Loudspeaker-Same-Surface scenarios, in the *normal SPL setting*, also showed no evidence of speech information leakage for all of the tested HDDs. Figures 4b & 4c and Appendix Figures 11c & 11d show frequency spectrum graphs from two of the HDDs that we tested. For each of these HDDs, there was no clear response observed so the potential for information leakage in an attack scenario is not likely. The absence of speech information in all samples was confirmed with cross-correlation analysis. The resulting cross-correlation graphs did not contain any peaks that would indicate information loss.

6.3 Speech Intelligibility Metric Analysis

Here we describe the results from our speech intelligibility metric analysis. Table 2 displays the averaged PESQ and STOI scores calculated for each experimental scenario. First, we see low PESQ scores for nearly all scenarios compared to the score generated for the raw microphone recordings. If a PESQ score of 3.5 (out of a possible 4.5) represents fully intelligible speech, we can see the lack of speech leakage in our recovered samples. Almost all scenarios and loudness levels obtained PESQ scores less than half of the microphone sample score. We do see the Touching-Very Loud scenario achieved a PESQ score over 2.0, which is an indicator of greater speech leakage potential in this single scenario.

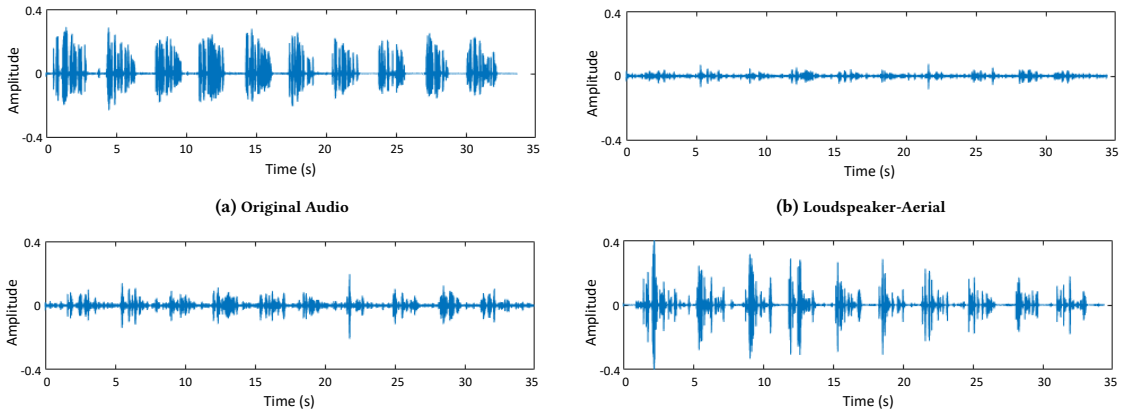
Continuing, we also used the STOI metric to evaluate speech intelligibility. STOI is specialized for shorter, noisy audio clips and may be a more accurate metric for determining perceived speech intelligibility. We see that the raw microphone recordings obtained an STOI score of 0.93 (out of 1.0), confirming highly intelligible speech. In contrast, we observed much lower scores for recovered samples from most of the experimental scenarios. Both the Live Human and Loudspeaker-Aerial scenarios (at all loudness levels) received STOI scores of $\leq$0.30 indicating very low speech intelligibility. We see similarly low scores for the Loudspeaker-Same Surface samples, except in the Very Loud setting that achieved a score of 0.36. Although this is still indicative of low intelligibility, we see a decent increase in the score compared to the lower volume levels. Lastly, the Loudspeaker-Touching scenario shows a more significant improvement in STOI scores. Even at the lower volume levels we see scores around 0.50, with a maximum score of 0.57 achieved in the Very Loud setting. While these scores are lower than the microphone recordings, they are larger than the other scores achieved which indicates a real potential for speech leakage.

6.4 Automatic Speech Recognition Analysis

Our ASR analysis revealed Google STT was unsuccessful at identifying the original speech in our selected samples and the majority of samples resulted in no textual output. However, some samples from the Loudspeaker-Same Surface and Loudspeaker-Touching scenarios did result in transcription attempts (i.e., some textual output), although incorrect. The words that were "transcribed" were not from the original speech so ASR was still unsuccessful. These results do indicate more speech leakage potential for these propagation media when the speech is Loud or Very Loud. And although recognition accuracy by ASR was still 0%, the textual output suggests speech content was detected.

6.5 Specialized Classification Analysis

The classification results that we observe support our previous conclusions about when leakage is present and absent. First, we find that none of the classifiers were able to achieve accuracies above the random guessing rate (10%) for the samples recovered from Live Human speech at the normal loudness level. Next, we find that the recovered Loudspeaker-Aerial samples in the Loud SPL setting achieved an accuracy slightly above random guessing



(a) Original Audio

(b) Loudspeaker-Aerial

(c) Loudspeaker-Same-Surface

(d) Loudspeaker-Touching

Figure 6: Time domain graphs for the original audio and data collected from the Seagate Barracuda 7200.12 1TB HDD in each of the different propagation medium scenarios, in the very loud SPL setting. Notice how the degree of information leakage, related to the amplitude of the signal peaks, increases from the Aerial to Same-Surface to Touching scenarios.

for the Logistic Regression classifier (12%). However, the other classifiers had accuracies below that of random guessing. For the Loudspeaker-Same Surface samples in the Loud SPL level, we find that the Logistic Regression and MultiClass classifiers achieved accuracies above the random guessing rate (19% and 12%), but the other classifiers performed worse than random guessing. Lastly, the recovered samples from the Loudspeaker-Touching scenario, in the Loud SPL setting, actually demonstrated real potential for speech leakage. We find that three of the classifiers achieved accuracies of at least double the random guessing rate, with the maximum accuracy observed being four times the random guessing rate (21%, 20%, and 40%). Although 40% classification accuracy is not significant for demonstrating a successful speech recognition attack, compared to other accuracies we can clearly see the potential for speech leakage in the Touching scenarios, while others have little potential.

7 Summary & Further Insights

We observed that the potential for speech information leakage was directly related to some of the parameters that we explored in our study. Both the *transfer medium* by which the sound waves propagated and the *loudness* (dB) of the external speech appeared to have significant influence on the strength of the induced vibrations (i.e., the amount of information loss). The difference between transfer mediums was seen by comparing the results of our analysis across the different mediums for each HDD. For all HDDs tested, we can describe the strength of induced vibrations (i.e., amount of information leakage) in terms of the propagation medium as: *Direct Contact* > *Shared Surface* > *Aerial*. The time domain graphs displayed in Figure 6 show the increasing strength of induced vibrations across these propagation mediums.

Similarly, we observed a consistent trend in the relation between speech loudness and information leakage for all HDDs tested. As expected, we found that a louder speech source will induce stronger vibrations on an HDD. From our experiments, we can describe the strength of induced vibrations in terms of the loudness of the original speech as: *Very Loud* > *Loud* > *Normal*. The very loud SPL setting (85 dB) consistently produced data with the most information leakage, while the normal SPL level (60 dB) produced data with little or no information leakage, depending on the scenario. The trends defined above for both transfer medium and loudness can be seen in the time domain and frequency spectrum graphs. Further analysis via cross-correlation and human listening, to determine speech intelligibility, confirmed these observed trends. Appendix Figure 12 displays the cross-correlation graphs created for the original audio vs. data collected from the Seagate 80 GB HDD in the control setting (no speech) and in the normal, loud, and very loud SPL settings. We clearly see that the peak size at lag=0 increases from No Speech to Normal to Loud to Very Loud SPL setting, demonstrating that increased volume of the source speech will cause greater information leakage (i.e., higher correlation). Our results for the transfer medium and loudness level parameters are inline with the results reported in [5]. We recreated their specific results by using the same exact HDD in our Loudspeaker-Aerial scenario at the very loud SPL setting. We replicated their correlation analysis and confirmed leakage in the settings of [5].

8 Discussion & Future Work

Limitations: The conclusions of our work are limited to the real-world speech scenarios represented in our experiments, and the attacks that specifically target HDD vibrations. We make no claims of presence/absence of speech leakage in attack scenarios where speech vibrations are recorded off of an object other than an HDD, or from the speech source itself. The favorable conditions we maintained in our experiments such as the open HDD case with exposed read/write head and source speech at a close distance are meant to allow for higher quality vibration measurements (than an attacker could make). Additionally, we find that the internal vibrational noise created when the HDD is powered on is one of the greatest hindrances to capturing vibrations proportional to the source speech. We acknowledge that our results are only applicable to HDD speech eavesdropping attacks under certain settings.

Potential Future Directions: There are many factors that determine how an object is affected by acoustic vibrations including

physical structure and natural vibrations that exist within that object. This encourages us to investigate this subject further and elaborate on the work we have done in this project. Specifically, we can conduct additional experiments with other HDD models and explore different solid transfer mediums. We will look to determine what materials, if any, are more conducive to propagating vibrations. Laser vibrometry allows us to explore the vibration domain in extensive detail and inspires new research objectives. For example, the threat of DDoS attacks against HDDs from sound waves [35] could be fully explored using our laser vibrometry methodology.

9 Other Related Works

Information leakage via vibration measurements has been studied in prior works. In [8], Marquardt et. al showed how vibrations recorded by a smartphone's accelerometer can be used to determine the text typed on the smartphone. In [36], [37] and [9], authors performed similar research on inferring a user's touchscreen inputs on an Android device by utilizing the vibrations recorded by motion sensors. Michalevsky et. al. [10] showed the gyroscope of a mobile phone may be sensitive enough to react to speech signals. Further research into this vulnerability, performed by Anand et al. [38], found that live human speech and machine-rendered speech were not able to induce a vibrational effect on the motion sensors of an Android phone across the aerial medium. Interestingly, the insights gained from our paper support the study of [38], but our work focuses on an independent application domain involving the potential of HDDs eavesdropping over speech. Additionally, our work confirms this observation for an attack scenario that utilizes vibration sensors with much greater resolution (up to 44.1 kHz) than is found in MEMS (100 Hz to 200 Hz). While higher sensor fidelity can certainly lead to increased eavesdropping potential from vibration data, our study demonstrates that the threat to live speech is still very low even when finer grained data can be acquired.

The use of vibrations for sound reconstruction has also been explored. In [39], Davis et al. use a high speed camera to capture footage of varying objects exposed to sound such as an empty bag of chips and a plant. From the footage, vibration data was extracted and used to reconstruct the external sound. Cordourier et al. demonstrate [40] that human speech can be reconstructed using nasal vibrations measured by a pair of smart glasses. Therefore, if an accurate vibration measurement of the sound waves of speech can be made, it would be feasible to reconstruct speech. The success of this is largely determined by both the quality of the speech and the quality of the vibrations measured. Uniquely, our work has performed a broad investigation of multiple key parameters that affect the success of speech eavesdropping. Doing this allows our study to better generalize the feasibility of this attack.

10 Conclusion

In this work, we have analyzed the vibrational impact induced on the read/write heads of different HDD models and observed that extracting speech information in some settings is difficult, even with speech enhancement procedures. Our results suggest that in realistic scenarios, standard magnetic HDDs are unlikely to leak sensitive speech information from an aerial source via vibrations that are induced on the read/write head. Using the LDV methodology, we empirically prove the effect of different parameters (i.e., volume, propagation medium) on eavesdropping success under certain standard conditions. We recreated the results of previous work on this attack to confirm the validity of our vibrometer methodology.

Considering the higher-risk threat model used in our study, we suspect that under live settings (where the favorable conditions of our experimental setup do not exist) the severity of vibrational impacts would be lessened. Therefore, we believe that sound waves are less likely to induce vibrations that could leak speech information, *unless the sound waves can propagate through a solid shared medium or direct contact* and are *above the normal loudness range for human conversation*. In this light, it seems that the potential threat of HDDs eavesdropping is not as viable as previously suggested.

11 Acknowledgments

We would like to thank the company Polytec for lending us the laser vibrometer equipment used for this research. We would also like to thank the reviewers for providing valuable feedback for how to improve this paper. This research is partially supported by the National Science Foundation (NSF) under the grants: CNS-1714807, CNS-2030501, CNS-2139358.

References

[1] T. Coughlin. (2018) Digital storage projections for 2019, part 1. [Online]. Available: https://www.forbes.com/sites/tomcoughlin/2018/12/21/digital-storage-projections-for-2019-part-1/#77bfba674428
[2] J. McKane. (2018) Why hard drives won't be replaced by ssds any time soon. [Online]. Available: https://mybroadband.co.za/news/hardware/251995-why-hard-drives-wont-be-replaced-by-ssds-any-time-soon.html
[3] S. Xiong and D. B. Bogy, "Position error signal generation in hard disk drives based on a field programmable gate array (fpga)," *Microsystem Technologies*, vol. 19, pp. 1307–1311, 2013.
[4] M. J. McCaslin, V. Thaveeprungsriporn, and S.-H. Hu, "Flexure based shock and vibration sensor for head suspensions in hard disk drives," August 2012, uS Patent 2012/0200959 A1. [Online]. Available: https://patents.google.com/patent/US20120200959A1/en
[5] A. Kwong, W. Xu, and K. Fu, "Hard drive of hearing: Disks that eavesdrop with a synthesized microphone," in *40th IEEE Symposium on Security and Privacy*, 2019.
[6] E. Sengpiel. (2011) Decibel table–loudness comparison chart. [Online]. Available: http://www.siue.edu/~gengel/ece476WebStuff/SPL.pdf
[7] E. P. Department. (2017) Characteristics of sound and the decibel scale. [Online]. Available: http://www.epd.gov.hk/epd/noise_education/web/ENG_EPD_HTML/m1/intro_5.html
[8] P. Marquardt, A. Verma, H. Carter, and P. Traynor, "(sp)iphone: Decoding vibrations from nearby keyboards using mobile phone accelerometers," in *Proceedings of the 18th ACM Conference on Computer and Communications Security*, 2011, pp. 551–562.
[9] E. Owusu, J. Han, S. Das, A. Perrig, and J. Zhang, "Accessory: Password inference using accelerometers on smartphones," in *Proceedings of the Twelfth Workshop on Mobile Computing Systems & Applications*, 2012, pp. 9:1–9:6.
[10] Y. Michalevsky, D. Boneh, and G. Nakibly, "Gyrophone: Recognizing speech from gyroscope signals," in *23rd USENIX Security Symposium (USENIX Security 14)*, 2014, pp. 1053–1067.
[11] (2019) Vibration of hard disk drives. [Online]. Available: https://www.me.washington.edu/research/faculty/ishen/hdd_vibration
[12] B. Kelly. (2016) Everything you need to know about hard drive vibration. [Online]. Available: https://www.ept.ca/features/everything-need-know-hard-drive-vibration/
[13] A. Rubtsov. (2016) Hdd inside: Tracks and zones. how hard it can be? [Online]. Available: http://hddscan.com/doc/HDD_Tracks_and_Zones.html
[14] Polytec. (2019) Laser doppler vibrometry. [Online]. Available: https://www.polytec.com/us/vibrometry/technology/$laser-doppler-vibrometry/
[15] ——. (2019) Optical vibration measurement per laser vibrometry. [Online]. Available: https://www.polytec.com/us/vibrometry/areas-of-application/
[16] Wikipedia. (2019) Hard disk drive platter. [Online]. Available: https://en.wikipedia.org/wiki/Hard_disk_drive_platter
[17] Polytec. (2019) Ofv-5000 modular vibrometer. [Online]. Available: https://www.polytec.com/us/vibrometry/products/single-point-vibrometers/ofv-5000-modular-vibrometer/

[18] ——. (2019) Vibroflex. [Online]. Available: https://www.polytec.com/eu/vibrometry/products/single-point-vibrometers/vibroflex/

[19] ——. (2019) Pdv-100 portable digital vibrometer. [Online]. Available: https://www.polytec.com/us/vibrometry/products/single-point-vibrometers/pdv-100-portable-digital-vibrometer/

[20] Cnet. (2019) Seagate desktop hdd st250dm000 - hard drive - 250 gb - sata 6gb/s specs. [Online]. Available: https://www.cnet.com/products/seagate-desktop-hdd-st250dm000-hard-drive-250-gb-sata-6gb-s/

[21] ——. (2019) Seagate barracuda 7200.7 (80gb, sata-150) specs. [Online]. Available: https://www.cnet.com/products/seagate-barracuda-7200-7-80gb-sata-150/

[22] ——. (2019) Seagate barracuda 7200.12 (1tb, sata-600) specs. [Online]. Available: https://www.cnet.com/products/seagate-barracuda-7200-12-1tb-sata-600/

[23] ——. (2019) Hitachi deskstar 7k80 hds728080pla380 - hard drive - 80 gb - sata-300 specs. [Online]. Available: https://www.cnet.com/products/hitachi-deskstar-7k80-hds728080pla380-hard-drive-80-gb-sata-300/

[24] ——. (2019) Fujitsu mobile mhy2120bh - hard drive - 120 gb - sata-150 series specs. [Online]. Available: https://www.cnet.com/products/fujitsu-mobile-mhy2120bh-hard-drive-120-gb-sata-150-series/

[25] (2019) Ultrastar dc hc600 smr series. [Online]. Available: https://www.westerndigital.com/products/data-center-drives/ultrastar-dc-hc600-series-hdd

[26] M. Schulz-Narres. (2018) Pmr, smr, cmr, i-just-want-a-hdd-mr. [Online]. Available: https://blag.nullteilerfrei.de/2018/05/31/pmr-smr-cmr-i-just-want-a-hdd-mr/

[27] A. Shilov. (2019) 16 tb mamr hard drives in 2019: Western digital. [Online]. Available: https://www.anandtech.com/show/13764/western-digital-2019-16tb-hdd-mamr-hamr

[28] D. G. (2018) Hard drive revolution in 2019? technology clash: Hamr vs mamr. [Online]. Available: https://www.cloudberrylab.com/resources/blog/hamr-vs-mamr-new-hdd-technology/

[29] H. Min, X. Huang, and Q. Zhang, "Active control of flow-induced vibrations on slider in hard disk drives: Experimental demonstration," *IEEE Transactions on Magnetics*, vol. 49, pp. 3038–3041, 2013.

[30] I. S. on Subjective Measurements, "Harvard sentences," *IEEE Recommended Practices for Speech Quality Measurements*, vol. 17, pp. 227–246, 1969. [Online]. Available: https://www.cs.columbia.edu/~hgs/audio/harvard.html

[31] M. Brookes. (2017) Voicebox: Speech processing toolbox for matlab. [Online]. Available: http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html

[32] A. Rix, J. Beerends, M. Hollier, and A. Hekstra, "Perceptual evaluation of speech quality (pesq)-a new method for speech quality assessment of telephone networks and codecs," in *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.01CH37221)*, vol. 2, 2001, pp. 749–752 vol.2.

[33] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "A short-time objective intelligibility measure for time-frequency weighted noisy speech," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2010, pp. 4214–4217.

[34] (2020) Speech-to-text. [Online]. Available: https://cloud.google.com/speech-to-text

[35] M. Shahrad, A. Mosenia, L. Song, M. Chiang, D. Wentzlaff, and P. Mittal, "Acoustic denial of service attacks on hard disk drives," in *Proceedings of the 2018 Workshop on Attacks and Solutions in Hardware Security*, 2018, pp. 34–39.

[36] Z. Xu, K. Bai, and S. Zhu, "Taplogger: Inferring user inputs on smartphone touchscreens using on-board motion sensors," in *Proceedings of the Fifth ACM Conference on Security and Privacy in Wireless and Mobile Networks*, 2012, pp. 113–124.

[37] L. Cai and H. Chen, "Touchlogger: Inferring keystrokes on touch screen from smartphone motion," in *Proceedings of the 6th USENIX Conference on Hot Topics in Security*, 2011, pp. 9–9.

[38] 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)*, 2018, pp. 1000–1017.

[39] A. Davis, M. Rubinstein, N. Wadhwa, G. J. Mysore, F. Durand, and W. T. Freeman, "The visual microphone: Passive recovery of sound from video," *ACM Trans. Graph.*, pp. 79:1–79:10, 2014. [Online]. Available: http://doi.acm.org/10.1145/2601097.2601119

[40] H. A. C. Maruri, P. Lopez-Meyer, J. Huang, W. M. Beltman, L. Nachman, and H. Lu, "V-speech: Noise-robust speech capturing glasses using vibration sensors," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, pp. 180:1–180:23, 2018. [Online]. Available: http://doi.acm.org/10.1145/3287058

A Appendix

A.1 Hard Disk Drive Information and PES Experiment

Table 3: Specifications of the HDDs tested in our study

Manufacturer	Model (Capacity)	Interface (Spindle Speed)	W/D/H (Weight)
Seagate Technology LLC	ST31000528AS 7200.12 (1 TB)	Serial ATA-600 (7200 rpm)	4 in / 5.8 in / 1 in (1.37 lbs)
Seagate Technology LLC	ST380013AS (80 GB)	Serial ATA-150 (7200 rpm)	4 in / 5.8 in / 1.03 in (1.4 lbs)
Seagate Technology LLC	ST250DM000 (250 GB)	Serial ATA-600 (7200 rpm)	4 in / 5.8 in / 0.8 in (0.92 lbs)
Hitachi Global	HDS728080PLA380 (80 GB)	Serial ATA-300 (7200 rpm)	4 in / 5.7 in / 1 in (1.30 lbs)
Fujitsu Company, LTD	MHY2120BH (120 GB)	Serial ATA-150 (5400 rpm)	2.8 in / 3.9 in / 0.4 in (0.223 lbs)

Figure 7: Force diagram showing the natural vibrations internal to HDDs.

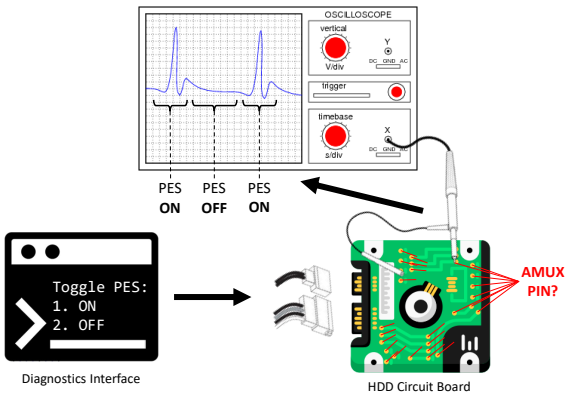
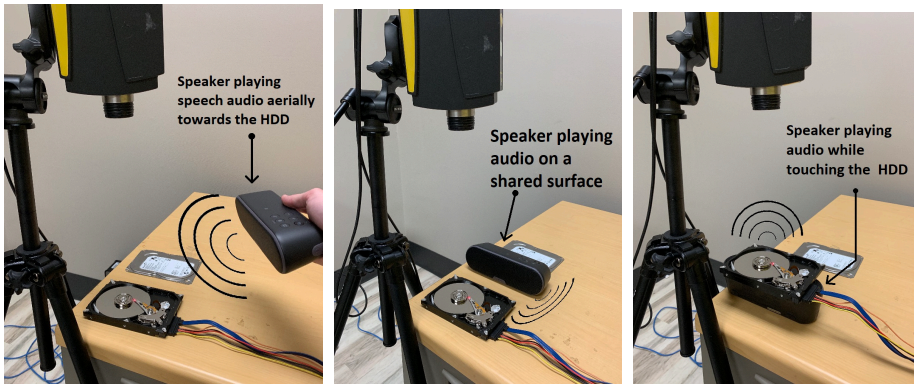


Figure 8: Diagram depicting initial experiment conducted to attempt finding the exposed AMUX pin on an HDD motherboard. All HDDs in this study were used in this experiment and observing the oscilloscope output, we found that none of the HDDs had the AMUX pin exposed for reading PES data.

A.2 Setups of the Experimental Speech Source Scenarios



(a) Aerial Setup (b) Same-Surface Setup (c) Touching Setup

Figure 9: Images of the 3 scenario setups that use the loudspeaker as the speech source.

A.3 Comparison of Time-Domain Graphs

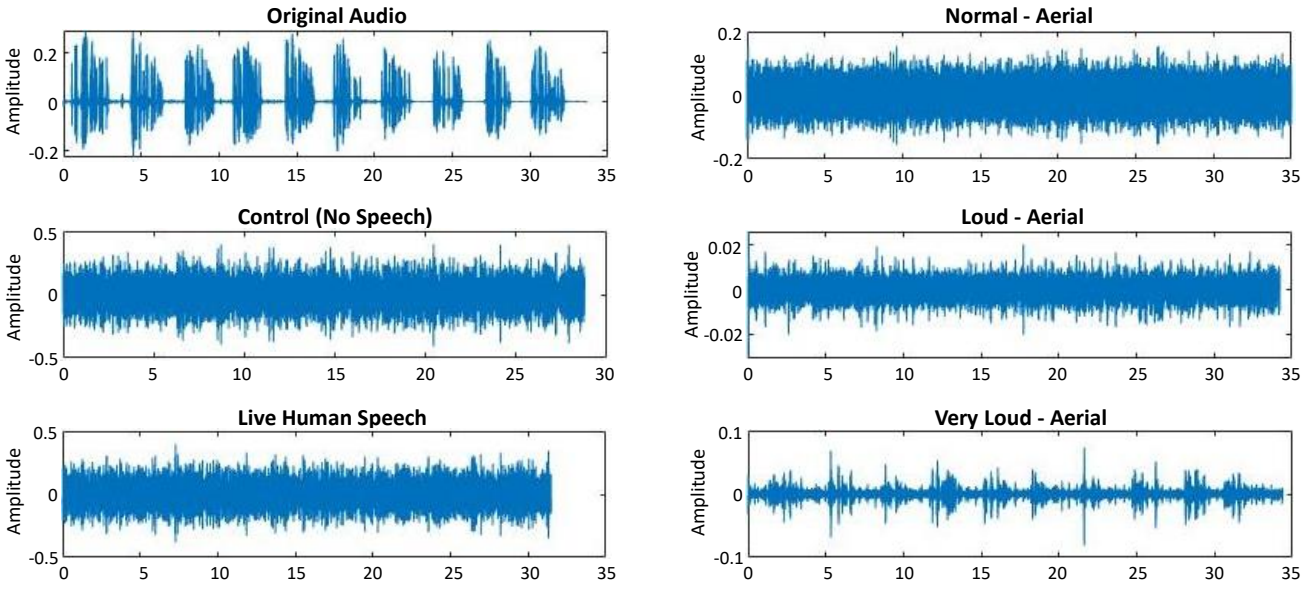
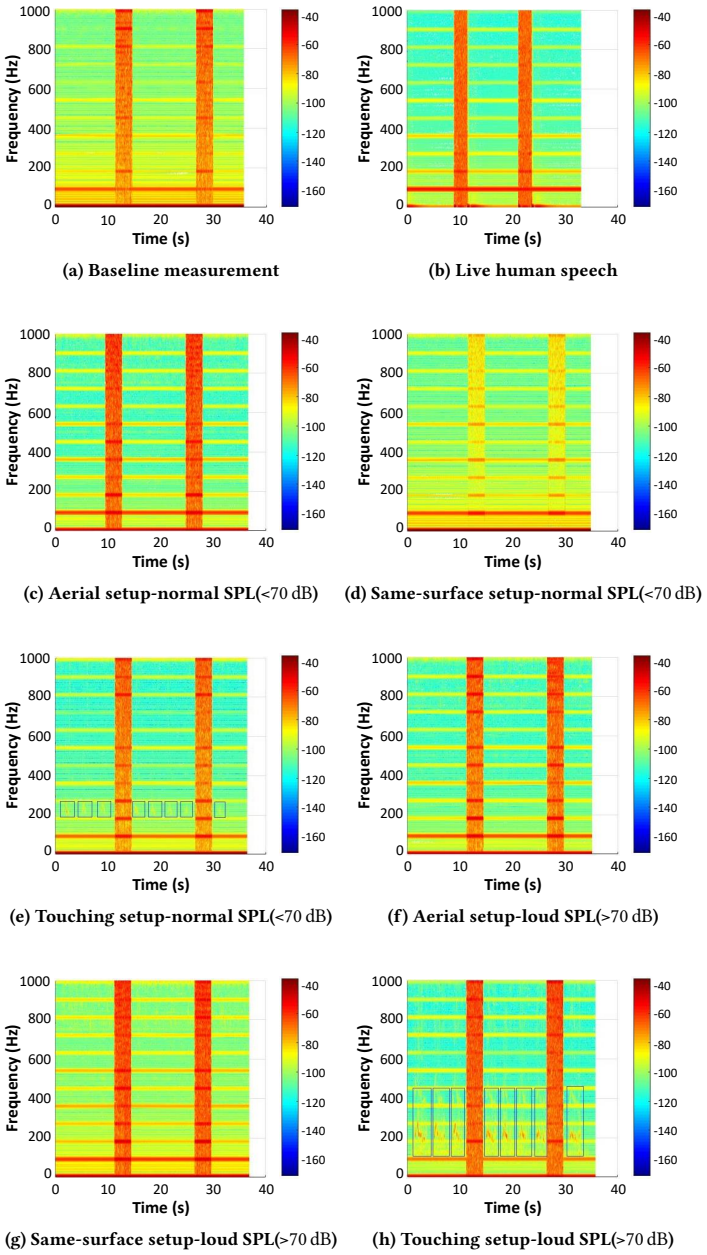


Figure 10: Side-by-side time domain graphs for original vs. reconstructed audio from each Aerial scenario. Aligning peaks in the Very Loud-Aerial graph indicate information leakage. Control, Live Human, Normal-Aerial, and Loud-Aerial data was collected from Hitachi HDD and Very Loud-Aerial data was collected from Seagate Barracuda 7200.12 1TB.

A.4 Additional Graphs



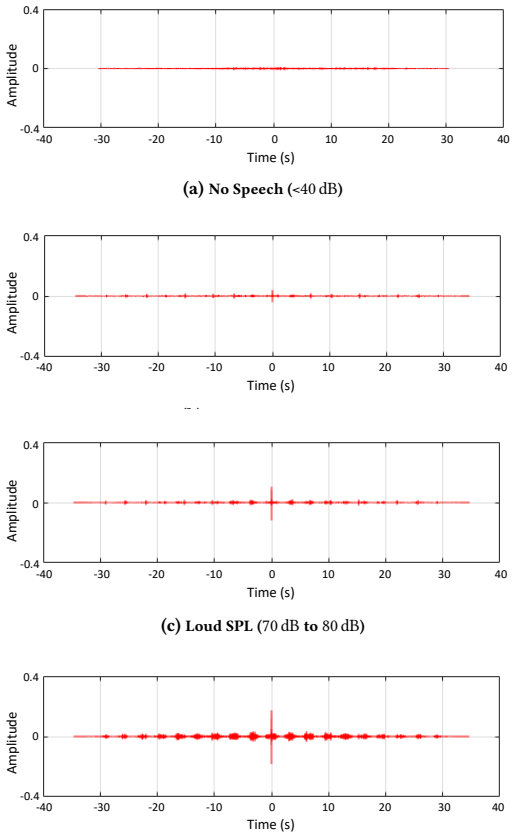
(a) Baseline measurement (b) Live human speech

(c) Aerial setup-normal SPL(<70 dB) (d) Same-surface setup-normal SPL(<70 dB)

(e) Touching setup-normal SPL(<70 dB) (f) Aerial setup-loud SPL(>70 dB)

(g) Same-surface setup-loud SPL(>70 dB) (h) Touching setup-loud SPL(>70 dB)

Figure 11: Frequency spectrum graphs of the Fujitsu 120 GB mini-HDD reveal that only the Loudspeaker-Touching scenarios induce a noticeable frequency change - outlined by the blue boxes.



(a) **No Speech** (<40 dB)

(c) **Loud SPL** (70 dB **to** 80 dB)

(d) **Very Loud SPL** (>85 dB)

Figure 12: Cross-correlation graphs from the Seagate 80 GB HDD in each of the different loudness settings, for Loudspeaker-Touching scenario, vs. original audio. The degree of information leakage, related to the amplitude of the central peak at lag=0, increases from *No Speech* to *Normal* to *Loud* to *Very Loud* SPL settings.