Can't Hear You: A Laser Vibrometry Study of the Viability of Hard Disk Drives to Eavesdrop Speech Vibrations

***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. Any technology that can precisely measure vibrations can be used to achieve this goal. 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 that can be used to eavesdrop on the speech (Kwong et a.; Oakland'19). This threat would clearly have devastating real-world implications due to the mass use of HDDs.**

In this paper, we revisit this line of research and aim to provide a closer investigation into whether or not HDDs can in fact be used as a source for eavesdropping on speech vibrations. As a foundation for our study, we introduce a methodology using *laser vibrometry* to measure the subtle speech vibrations induced on the read/write head. Our vibrometer methodology is effective, and robust to accommodate for the increased scope of our research. The prior study tested only a single HDD and only machine-rendered speech in a single setting with very loud speech. Using our methodology, our work broadens the scope of this research in many significant ways. *First*, we test multiple popular HDDs of different models and sizes. In doing so, we aim 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 settings involving normal human conversations. *Third*, we define other real-world scenarios using machine-rendered speech from different source locations and propagation media, and at different degrees of 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. The overall implication of our work therefore is that HDDs are not capable of eavesdropping on human speech in most real-world scenarios and the impact of the threat seems lower than perceived previously.

I. 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, the basic function has remained the same – 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, an attacker can 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].

Computers containing magnetic HDDs are abundant and widely used; and the unsuspecting nature of this threat makes any sensitive information spoken near a computer vulnerable to eavesdropping by an attacker. The success of this attack would make any HDD a potential recording device and it holds a similar threat level to an attack using microphone surveillance to eavesdrop on speech directly. However, implementing HDDs as microphones is a stealthier attack in that the recording devices (e.g., the HDDs) are already installed and hidden from the potential victim.

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, 85 and 90 dB), well above the normal range for human conversation (40-60 dB) [6], [7]. Figure 1b depicts the single scenario tested in [5].

In this paper, we broaden the scope of the prior study in significant ways and pursue a comprehensive study of the vibration impact to the read/write head of HDDs when exposed to external speech. Our goal is to better define the scope of this threat model and to determine the viability of HDDs to eavesdrop on human speech. We do not look to implement any particular attack instance, but rather determine whether or not such a threat may be possible by measuring the impact of speech vibrations on HDDs. 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 measure the vibrations of the read/write head. 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. Here, our aim is to observe how most HDDs would react to speech vibrations. **Second**, we explore a wide variety of attack scenarios with variable parameters: *(1) speech sources* (live human speech and machine-rendered speech), *(2) loudness levels* (at normal human conversation (40-60 dB), loud (70+ dB), and very loud (85+ dB)), and (3) *transmission mediums* (aerial, shared surface, and direct contact (i.e., touching)), in order to gain a broader set of knowledge about the proposed threat. Assessing live human speech that travels aerially to the HDD in this context is extremely important as this represents perhaps the most natural and common scenario underlying the threat, which could put routine human conversations near HDDs at risk.

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). 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 III-D. Comparing each measurement with a baseline control measurement of each HDD's natural frequency vibration (i.e., in the absence of speech signals) allows 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 a normal loudness for conversation, seems incapable of causing an effect on the read/write head. This suggests that acoustic vibrations traveling through the aerial medium may not be strong enough at a normal conversational loudness to leak speech information.

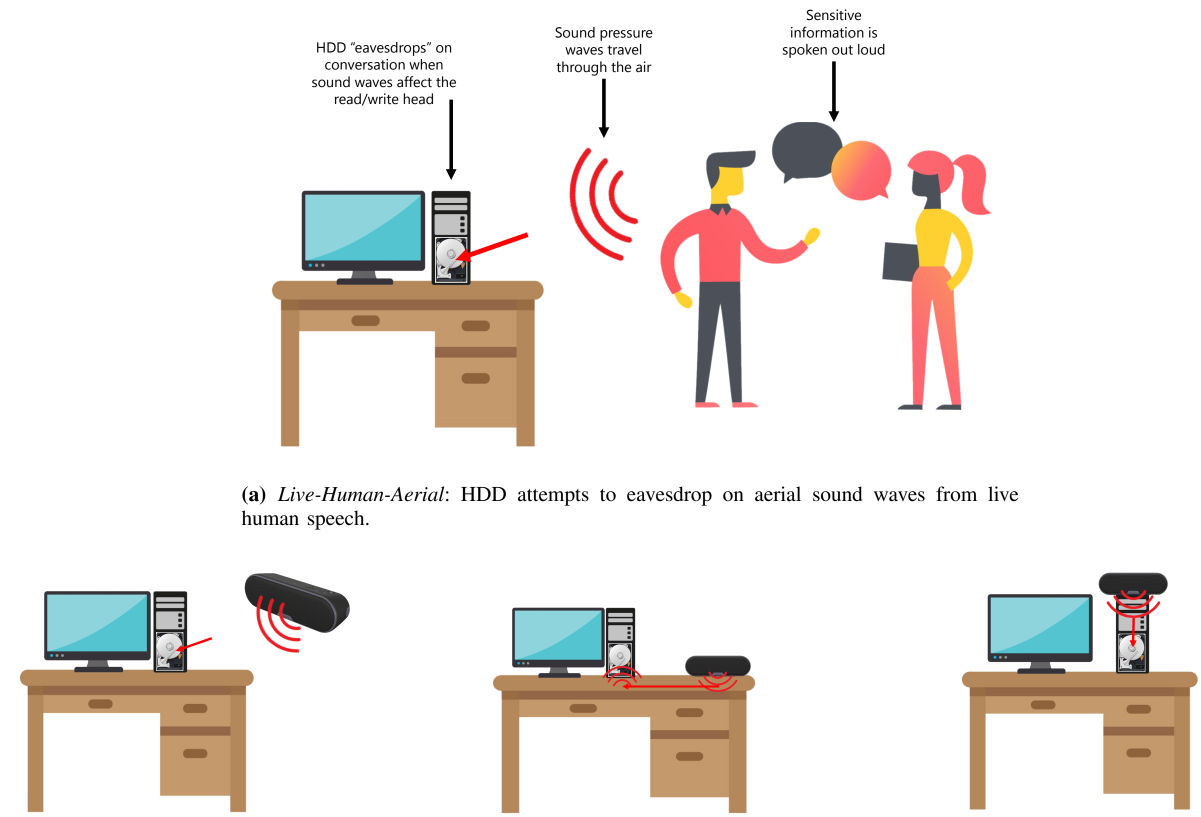
(2). Vibration Effects of Machine-Rendered Speech: We also observed the effects of vibrations induced by sound waves from machine-rendered speech through multiple propagation mediums and at multiple loudness levels. Specifically, we defined three machine-rendered speech scenarios. *Loudspeaker-Aerial* (Figure 1b), *Loudspeaker-Same-Surface* (Figure 1c) and *Loudspeaker-Touching* (Figure 1d). Full descriptions of each of the machine-rendered speech scenarios are found in Section III-D. 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 to the read/write head. Table I provides a summary of our observed results across all tested parameters.

(3). Laser Vibrometry as a Study Methodology: Our use of laser vibrometry to measure and study the effects of vibrations induced on the read/write head of HDDs by acoustic waves is a novel contribution in our research. Laser vibrometers allow us to measure subtle vibrations with a very high degree of precision and reveal information that would otherwise be unavailable without this technique. The use of this technique provides unique advantages for our current research because the vibrometer has very high precision and is easily portable. Our recreation of the results of [5] serves to further confirm the effectiveness of our methodology. Our methodology is elaborated in Section IV.

II. Threat Overview

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 [8]; allowing it to measure any external vibrations induced on the head that would cause displacement. The attacker could reflash the HDD's firmware to expose the PES. The attacker could also intercept the HDD physically while it is in transit and plant malware or conduct a man-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.

Additionally, 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 means by which an attacker gains this vibration data is irrelevant. There may be other side channel attacks or exploitable hardware within standard magnetic drives that could be used to collect the necessary



(a) *Live-Human-Aerial*: HDD attempts to eavesdrop on aerial sound waves from live human speech.

(b) *Loudspeaker-Aerial*: HDD attempts to eavesdrop on aerial sound waves from loudspeaker device.

(c) *Loudspeaker-Same-Surface*: HDD attempts to eavesdrop on sound waves propagated through a shared surface with the loudspeaker device.

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

Fig. 1: HDD speech eavesdropping attack scenarios, with two normal and loud volume settings, studied in our work (Kwong et al. only studied Loudspeaker-Aerial with loud volume settings).

vibration data. The defining characteristic of our attacker is the ability to measure the vibration leakage of the read/write head – by any viable method available.

III. Preliminaries & Attack Scenarios

A. 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 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 detect, record, and interpret the 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 propagating through the air. The vibrations induced on the diaphragm are transformed into an analog signal and then output by the microphone.

B. Natural HDD Vibrations

Vibrations in HDDs can pose serious problems, affecting the position of the head and causing read/write errors [9] . 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 [10]. 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. This acceleration causes a resultant force to be applied into the HDD, again inducing vibrations on the read/write head. Appendix Figure 8 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 is powered ON, but maintains an idle read/write head position, the acceleration of the actuator arm is also a source of vibration that does not apply to our experiments. Therefore, the only internal source for vibrations that introduces variability among our tested HDDs is *platter wobble*.

C. Experimental Attack Parameters

Sound Pressure Level: The loudness of a sound is represented by its Sound Pressure Level (SPL), and is measured in decibels (dB). The SPL of *normal* conversation is estimated between 40-60dB[6], [7]. Therefore, any noise above 70dB may be considered *loud* in terms of human conversation. As in previous work[5], greater dB levels are explored; so we will also analyze sounds that are $\geq$ 85 dB as being in 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 designate 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 consider the medium by which those sound waves travel to the HDD and how different mediums may induce different effects. We define three transfer mediums and observe how the vibrational impact of sound waves differs between them. The *aerial* medium involves sound waves traveling through the air towards the HDD. This represents the natural medium of human speech as people talk to each other. We also define the *same surface* medium to represent the scenario of a speaker device sharing a surface with an HDD. 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 consider the most severe scenario of propagated vibrations and define the *touching* medium in which a loudspeaker device playing speech is in physical contact with an HDD. Via this medium, the vibrations from the speaker device propagate directly into the HDD without initial damping from traveling through other transfer mediums.

D. Experimental Attack Scenarios

We conceptualize different scenarios to investigate the effect of sound wave vibrations induced on read/write heads. To meet the objectives of our study, we consider 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 consider factors not previously studied.

Live Human Speech: We design 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 term this scenario as *Live-Human-Aerial* and illustrate 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 design 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, and are called; *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 define is the *Loudspeaker-Aerial* scenario. This scenario refers to machine-rendered speech played through a loudspeaker device and traveling through the air 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 machine-rendered speech 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 from their friend. 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 the machine-rendered speech 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 is used to investigate how directly propagated vibrations impact the HDD.

IV. Our Methodology

In our study, we investigate the HDD vibrations induced on the read/write head from speech signals by implementing a laser vibrometry methodology. Our methodology provides a unique insight into the vibration domain features of HDDs in the presence of speech. We are able to observe these effects with great precision by utilizing this technique.

Overview of Laser Doppler Vibrometry: The Doppler effect refers to the change in a wave's frequency as it encounters a moving object [11]. 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, this effect refers to the frequency of a light beam shifting in proportion to its velocity as the light beam is reflected off of a moving object. The velocity information is recorded in the frequency of the laser beam and measured by the laser vibrometer. Through this process, a laser vibrometer can be used to measure the vibrations of an object. More specifically, it can measure the vibrational displacement, velocity, or acceleration. Additionally, the standard USB data acquisition system used with a laser vibrometer can process the voltage signal generated by the interferometer and digital decoding electronics of the vibrometer. The voltage signal is created by converting the frequency shifts recorded by the laser. Laser Doppler vibrometry is currently used in many applications and fields of research because it offers the highest resolutions for vibration measurements[12].

Our Implementation: We chose to use laser vibrometry to collect vibration data in our study for significant reasons. First, the laser vibrometer technology is highly portable to work for almost any measurement scenario and parameters. Specifically, this allows us to easily include a larger set of HDDs for experimentation. Other methods of acquiring head vibration data (i.e., firmware hacking) can be highly specific to the target HDD – making the technique difficult to scale for including different HDDs. As our study investigates the induced vibrations in 5 different HDDs, it is crucial that our methodology can easily transition for use across the different test drives.

Secondly, a laser vibrometer (such as the models we use in our study, discussed later in Section V-A), has very high accuracy and resolution in comparison to other known vibration measurement 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 [13]. With displacement resolutions of 0.1 pm (or 0.0001 nm) and 0.3 pm (or 0.0003 nm) [14], [15], we determine that the OFV-5000 and Vibroflex vibrometer setups (we elaborate on all vibrometers used in the next section) are more than sufficient for detecting the vibrations of the read/write head that would be recorded by other state-of-the-art methods.

V. Experiments and Data Collection

A. Vibrometer Equipment Used

We utilized three different models of laser vibrometers supplied by the company Polytec; the PDV-100 Portable Digital Vibrometer [16], the OFV-5000 Modular Vibrometer [14] and the Vibroflex vibrometer [15]. 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. This setup can also measure vibrational velocity in the 0-22kHz frequency range and it 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 if measured, 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 use the Vibroflex setup to collect data from different HDDs in the loud and very loud settings used in [5].

B. 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. Additionally, we acquired the exact same HDD model that was used in [5] in order to ensure our vibrometer methodology can be used to recreate the results reported in their work. We used the Seagate Barracuda (250 GB)[17], Seagate Barracuda (80 GB)[18], Seagate Barracuda 7200.12 (1TB - from [5])[19], Hitachi Deskstar (80 GB)[20], and Fujitsu mini (120 GB)[21] HDDs. The HDDs used were chosen because of their popularity and varying sizes and physical structures. Each of the five HDDs are SATA type and use standard spinning disk platters and read/write heads. Appendix Section A contains all physical specifications for each of the HDDs in Table II, and images of the inside of each HDD and their sizes are shown in Appendix Figure 7.

The move from longitudinal (LMR) to perpendicular magnetic recording (PMR) technology has influenced new HDD head types. Currently, there are two types of PMR heads; conventional PMR (CMR) and shingled magnetic recording (SMR). These versions have a similar recording type but isolate read/write operations between adjacent tracks differently. Because of added parallel layers, disk density improved to higher density HDDs. Two-dimensional Magnetic Recording (TDMR), Heat Assisted Magnetic Recording (HAMR) with laser-powered heating to write the data, Microwave-Assisted Magnetic Recording (MAMR) using a spin torque oscillator (STO) with the magnetic write head generating microwaves, and Energy Assisted Magnetic Recording (EAMR/EPMR) are newer technologies which may further influence head mechanisms to accommodate read/write operations on much higher density and compact HDDs [22].

Challenges like disk material, heat dissipation, and complex production cycles have not been completely resolved which keeps these HDDs from becoming mainstream [22], [23].

Advanced head type technologies, including TDMR, HAMR and MAMR, are still evolving and do not have enough market presence yet for a generalizability study. Giant Magnetoresistive (GMR) head type is currently the most common 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 CMR heads. Unlike CMR, HM-SMR heads are not drop-in replacements for traditional drives and require system software modifications [24], [25]. Therefore, the HDDs selected for our experiments are CMR/GMR/PMR type HDDs (80 GB – 2 TB).

C. Experiment Setup

For our experimental attack model, we defined the four different scenarios described in Section III-D. The *Live-Human-Aerial* scenario requires the human speaker to read the audio sample transcriptions at the loudness level of normal conversation. A digital sound level meter was placed beside the HDD to ensure the speech remained within 40-60dB at the point of measurement. The speaker was instructed to talk in the direction of the HDD, as if they were having a conversation with it. The speaker's mouth was located approximately 0.3 meters from the HDD throughout each experiment.

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 device's position and the evaluator's 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 damping of the sound waves by the time they reached the HDD. Therefore, we directed our external speech at the HDD in order to *maximize the possible impact from the produced sound waves*.

The second machine-rendered speech scenario, *Loudspeaker-Same-Surface*, was designed to represent the situation in which a speaker device is placed on the same surface as an HDD when the sensitive audio is played. Sharing a surface introduces a new transport medium by which vibrations can travel. When a speaker is placed on a solid surface and audio is played, the resulting vibrations propagate through that solid surface and can affect any other object that is in contact with that same surface. Our experimental setup placed the portable loudspeaker device on the same surface as the HDD, at a distance of 0.3 meters. Again, we pointed the device speakers in the direction of the target HDD to maximize the vibrational effect.

The third machine-rendered speech scenario we designed, *Loudspeaker-Touching*, involves the speaker device physically touching the HDD. In a practical situation, it would be unlikely for a speaker device to be in contact with the HDD directly. It would be more likely to find a speaker device touching the computer casing that houses the HDD. However, that situation is effectively a shared surface scenario with the physical computer casing being the shared surface. This scenario represents the greatest threat instance where the vibrations from the speaker device propagate directly into the HDD without having to pass through other decoupled materials or the air. The setup for this scenario placed the loudspeaker device on its back, with speaker facing up, and the HDD sitting on top of the speaker. This ensures the most severe propagation of vibration from the speaker to the HDD. Appendix Figure 9 provides images of our experimental setup for all three machine-rendered speech scenarios.

For each scenario described above, the laser vibrometer 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 read/write head of the HDD. Inline with the set-up used in HDD natural vibration measurement studies via vibrometry[26], the laser intersected the read/write head at the perpendicular angle to minimize the measurement error margin. For each experiment, the state of the read/write remained idle and in a fixed position. This scenario in which the read/write head remains idle throughout the entire measurement, while being exposed to external audio, is not likely in real-world scenarios. However, this scenario is more favorable for the feasibility of the attack as the read/write head is more affected by external vibrations when it is in a resting position. The vibrometer measurement was manually started and stopped to encompass the entirety of each speech sample played or spoken. Additionally, the digital sound level meter was used in each scenario to ensure the speech audio was in the correct decibel range that we defined for each SPL settings.

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 will not affect the results of our study as we are intentionally injecting noise around the HDD. Therefore, any minor background noise that is introduced will be masked by induced vibrations caused by the injected audio.

D. 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 case for future comparison in our analysis.

Speech Database: We used a collection of Harvard sentences sample phrases from the IEEE Recommended Practices for Speech Quality Measurements [27]. 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

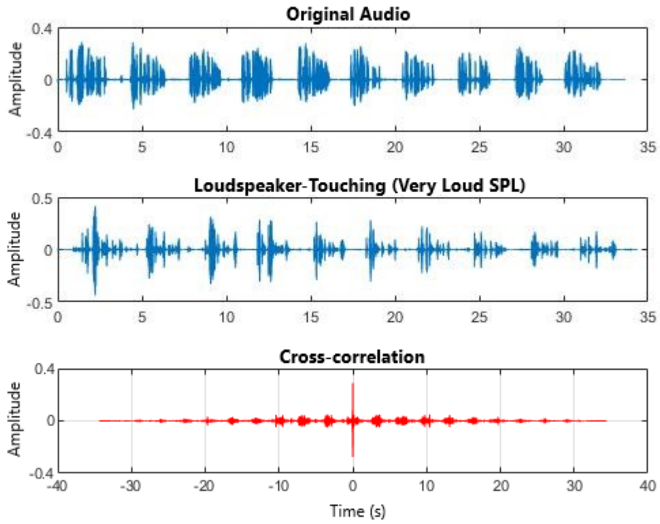


Fig. 2: The top graph is the time domain of the original audio. The middle graph is the time domain of the reconstructed audio. And the bottom graph is the cross correlation between the two signals. Notice the peak in the cross correlation graph at lag=0 which indicates some correlation between the signals (i.e., some amount of speech information may be contained in the signal). *Graphs above were generated from Seagate Barracuda 7200.12 1TB HDD data.

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 from one female speaker for a total of 60 recorded sentence samples. Three measurements were taken for each speech sample and scenario.

VI. Signal Analysis of Collected Data

The speech samples used for each scenario consist 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 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 analysis, 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.

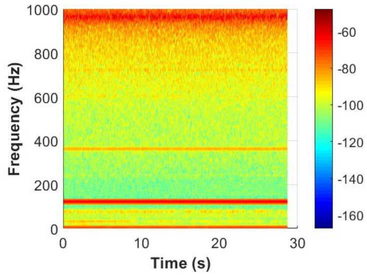
Time-Domain Analysis of Reconstructed Audio: To pre-process the data, we considered speech enhancement routines from the speech processing toolbox, VOICEBOX [28], implemented in Matlab. Specifically, we were interested in the routines *specsub*, *spendred*, *ssubmmse*, and *ssubmmsev*. The *specsub* 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 detector (VAD) based noise estimations. After running each routine on the converted speech samples and comparing the newly generated samples, we determined the best routine for our purposes. We found that the *ssubmmse* routine had the greatest speech enhancement and static noise reduction. Therefore, we chose *ssubmmse* as the noise filtering technique for our experiments. We refer to these audio files, after being processed, as the “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. 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 2. In the enhanced signal we can clearly see the 10 unique peaks throughout the signal that align with the 10 peaks in the time domain graph of the original audio.

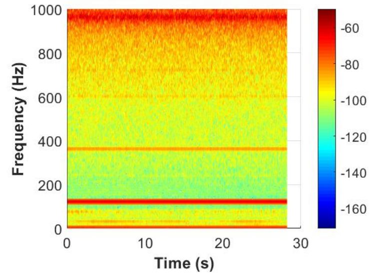
Frequency-Domain Analysis of Vibration Signal: For our frequency domain analysis, we utilized the raw vibration data collected from the vibrometer. We chose to use the raw vibration data because it contains the full spectrum of captured frequencies. With this, we are able to generate the full spectrum graphs for each measurement to identify where the relevant frequency markers appear – we identified that vibrations induced on the HDD by the external audio appear within the 150-1000Hz frequency range. Therefore, we focus on this frequency band for our analysis.

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 are used to easily identify any frequency markers directly caused by exposure to the speech signals. Similar to our time domain analysis, we can look for 10 distinct frequency signatures that may correspond to the 10 sentence utterances in the original audio. Figure 3 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.

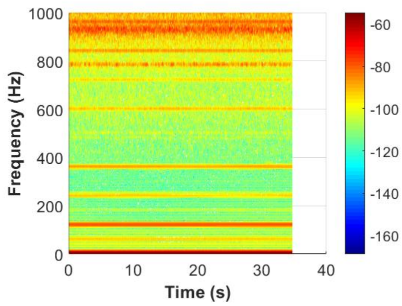
Cross-correlation Analysis of Input Audio and Reconstructed Audio: Although our observations from the time domain graph and frequency spectrum graph comparisons may indicate the presence of information leakage qualitatively, we must use a quantifiable measure to confirm these observations. Before determining the similarity between our reconstructed audio files and the original audio file, we further process the enhanced audio files described above to isolate any existing speech characteristics for the cross correlation analysis. For this, we apply a low bandpass filter on the enhanced audio



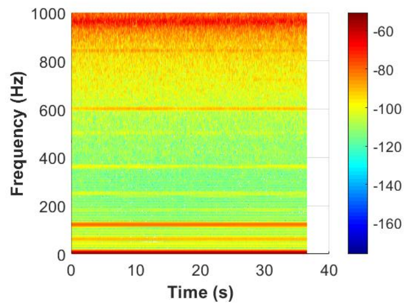
(a) Baseline measurement



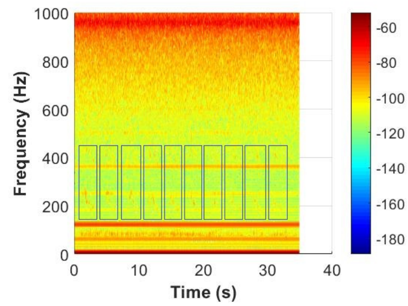
(b) Live human speech



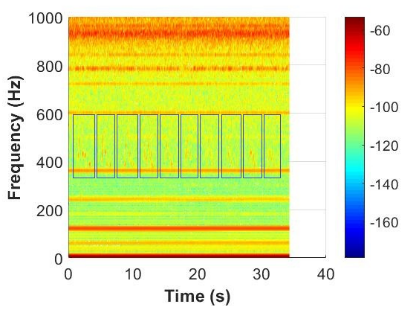
(c) Aerial setup-normal SPL(<70dB)



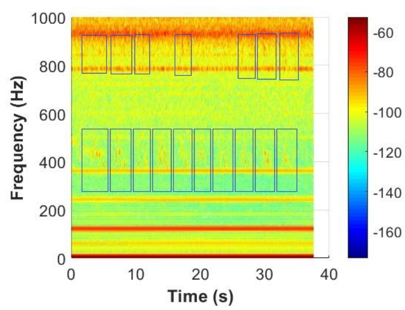
(d) Same-Surface setup-normal SPL(<70dB)



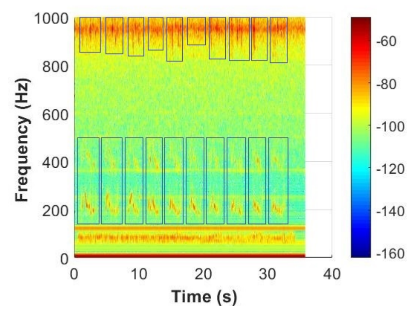
(e) Touching setup-normal SPL(<70dB)



(f) Aerial setup-loud SPL(>70dB)



(g) Same-Surface setup-loud SPL(>70dB)



(h) Touching setup-loud SPL(>70dB)

Fig. 3: Frequency spectrum graphs of the Hitachi 80GB 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.

to capture frequencies between 150-1000 Hz.[1] 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 compute the cross correlation between the time domain data of the original audio and the time domain data of each of the enhanced audio samples. The graphs included in this paper were generated when the original audio was the Female Speaker #1, Harvard Sentence Set #1 (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 Sentence Sets #2 and #3 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 align our signals before computing the cross correlation, we would expect to see a peak in the correlation graph at the point lag=0. An example of this is shown in the cross correlation graph in Figure 2 where we can see the large peak indicating strong signal correlation. Our cross correlation analysis is in line with the signal analysis performed in [5], for showing speech presence.

[1]Plotting the full frequency spectrum of the raw vibration data, we determined that the frequency markers corresponding to the external speech audio were only present between 150-1000 Hz.

Human Speech Intelligibility Analysis: The final step in our data analysis is assessing the quality of speech in our

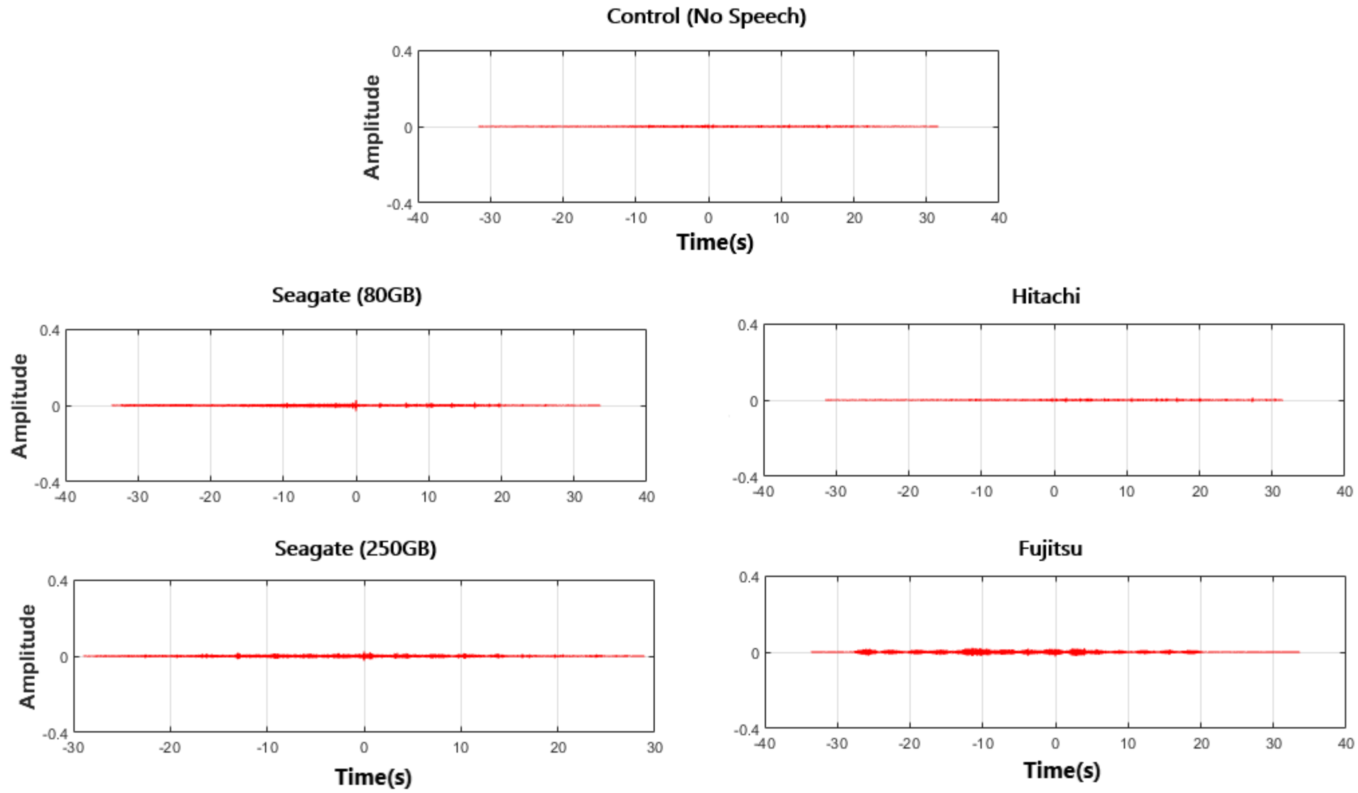


Fig. 4: Cross correlation between original audio and reconstructed audio from the *Live-Human-Aerial* scenario of each HDD. The correlation between the original audio and a control (No Speech) measurement, taken from the Seagate Barracuda 80GB HDD, was included for comparison. These graphs use the same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

enhanced, reconstructed samples. We start by investigating the intelligibility of the samples (i.e., information leakage) to live human listeners. The speech intelligibility component of our analysis will be used to confirm our observations from the time domain, frequency domain, and cross correlation analyses. Although speech recognition systems have achieved comparable performance to human listening in recent years [29], humans are still superior at comprehending speech in noisy environments [30]. Therefore, we find it important to investigate what information a human listener can glean from our reconstructed audio. We instructed each human participant to carefully listen to the enhanced audio files generated from the vibrometer data in each scenario, and to determine the level of speech intelligibility. Specifically, the human listener was asked to note the level of "speech presence" and "speech intelligibility". Speech presence refers to a listener's ability to recognize that there is speech contained in the audio, whether or not the speech can be understood. Speech intelligibility refers to a listener's ability to hear and understand the speech that is being spoken in the audio file such that the listener could transcribe or repeat the speech. The human participant made note of these properties for each of the enhanced audio files that was created from the data collected for each tested HDD, and each of the scenarios tested.

VII. 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 I, and the detailed results are discussed in this section. In Table I, 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/harddeafdrives/. In all cases, our results have been consistently validated via both our quantitative

TABLE I: Summary of results and HDDs used for experimentation for the different experimental parameters. *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
Live Human	Aerial	Normal	Not Likely	Seagate (80 GB), Seagate (250 GB), Hitachi, Fujitsu
Loudspeaker Device	**Aerial**	Normal	Not Likely	Seagate (250 GB), Hitachi, Fujitsu
		Loud	**Likely**	Seagate (80 GB), Seagate (250 GB), Hitachi, Fujitsu, **Seagate 7200.12**
		Very Loud	**Likely**	Seagate (80 GB), **Seagate 7200.12**
	Shared Surface	Normal	Not Likely	Seagate (80 GB), Seagate (250 GB), Hitachi, Fujitsu
		Loud	Likely	Seagate (80 GB), Seagate (250 GB), Hitachi, Fujitsu, Seagate 7200.12
		Very Loud	Very Likely	Seagate (80 GB), Seagate 7200.12
	Direct Contact	Normal	Very Likely	Seagate (80 GB), Seagate (250 GB), Hitachi, Fujitsu
		Loud	Very Likely	Seagate (80 GB), Seagate (250 GB), Hitachi, Fujitsu, Seagate 7200.12
		Very Loud	Very Likely	Seagate (80 GB), Seagate 7200.12

(correlation-based) analysis and the qualitative (inspection-based in time and frequency domains) analysis.

A. *Leakage Present / Recreated Prior Results*

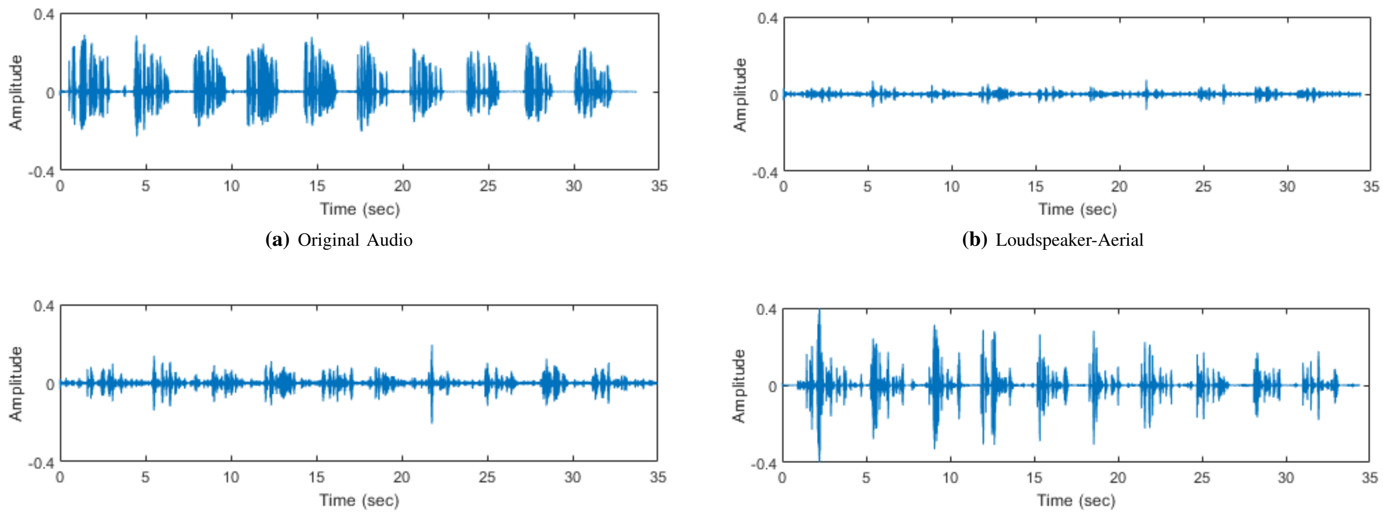
Through our research, we identified some of the tested scenarios in which information leakage was present and verifiable. In the Loudspeaker-Aerial scenario, we observed that information leakage was likely when the external speech was in the loud SPL setting (>70dB). This is one of our novel insights that was empirically demonstrated in our work via a new 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 3f. 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 and confirmed this observation with our cross correlation analysis. Figure 3g 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 present in both the normal and loud SPL settings. The time domain graphs and frequency spectrum graphs for the data collected in both loudness settings show distinct peaks/frequency markers that correspond to the 10 sentence utterances in the original file. Figures 3e & 3h 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 85dB, 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 of Section C. This confirms that our vibrometer methodology can achieve the same results as PES, validating our other data.

Lastly, the data collected for the Loudspeaker-Same-Surface and Loudspeaker-Touching scenarios, in the very loud SPL setting, were also shown to contain speech information via the cross correlation calculations – we see peaks at lag=0 in the graphs. The bottom graph in Figure 2 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 that indicates strong signal correlation.

B. *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 did not find any evidence of information leakage in the vibrometer data collected in the Live-Human-Aerial scenario for all of the HDDs tested. None of the other



(a) Original Audio **(b)** Loudspeaker-Aerial **(c)** Loudspeaker-Same-Surface **(d)** Loudspeaker-Touching

Fig. 5: 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.

HDDs tested in this scenario showed any indication of speech presence in their time domain and frequency spectrum graphs. This was confirmed 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 novel insight that we empirically demonstrated with many varying parameter settings and a new methodology. Figure 4 shows the cross correlation graphs generated with data from each HDD in the Live-Human-Aerial scenario.

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 3c & 3d 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. 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.

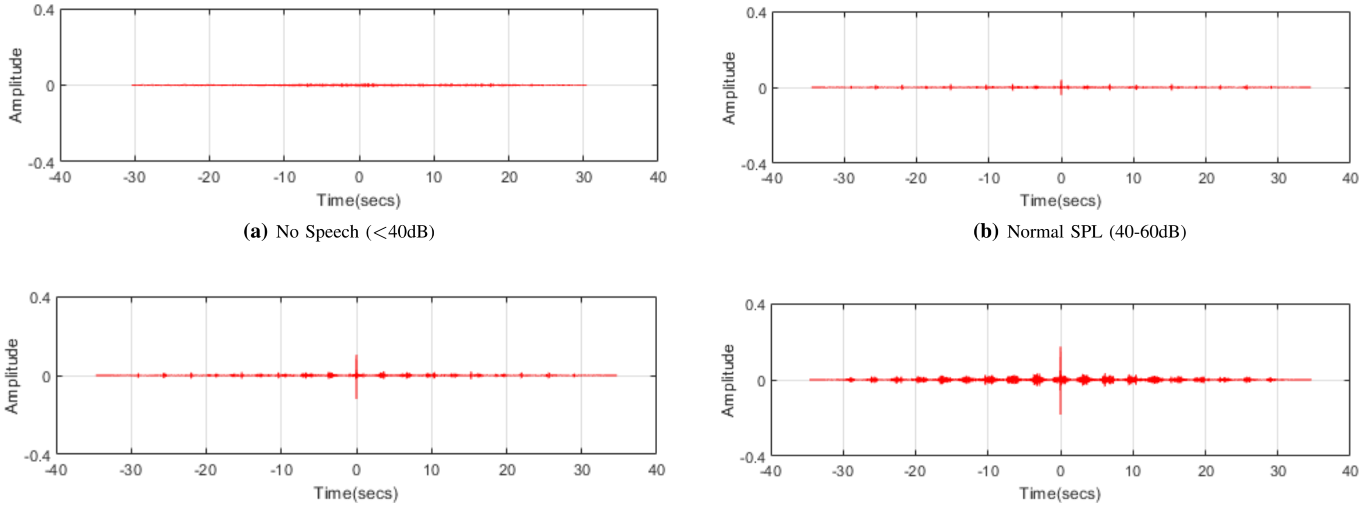
C. Human Speech Intelligibility Analysis

Here we discuss the results of our analysis of speech quality in our reconstructed and enhanced audio samples. We used a live human study to evaluate the quality of the speech signal reconstructed from the collected vibration data. We look to determine whether or not the speech signals in our reconstructed samples are strong enough to be exploited by an attacker. The results of our human speech intelligibility analysis suggest that the *volume* of the external speech and the *propagation medium* by which the sound waves travel is directly related to amount of *speech information leakage* in the collected data. The human listener observed varying levels of speech presence and intelligibility among the reconstructed audio samples of the different HDDs.

Live Human Speech Scenario: We observe that most HDDs may be incapable of eavesdropping on live human speech when spoken at normal conversational loudness. The enhanced audio file created from the vibration data collected from each of the tested HDDs reveal little to no speech information leakage in all of the HDDs. We found that speech was faintly present for the Seagate Barracuda 250GB audio sample. Although faintly present, we determined that there was no speech intelligibility as no specific words could be identified. The enhanced audio files for the other HDDs did not contain any evidence of speech presence and therefore had no speech intelligibility.

Machine-Rendered Speech Scenarios: In the *Loudspeaker-Aerial* scenario, we found the results for speech presence and intelligibility to be similar to those of the Live-Human-Aerial scenario described above in the normal loudness setting. In the normal loudness setting, the enhanced audio files for each HDD contained the faint presence of human speech. In the loud setting however, the enhanced audio files for each HDD indicated the clear presence of speech and some speech intelligibility. During the examination, we were able to recognize a few of the words spoken. Lastly, even greater speech intelligibility was observed in the Very Loud setting with the human listener being able to understand almost all of the speech in the enhacned audio file.

The results from our analysis for the *Loudspeaker-Same-*



(a) No Speech (<40dB) **(b)** Normal SPL (40-60dB)

(c) Loud SPL (70-80dB) **(d)** Very Loud SPL (>85dB)

Fig. 6: Cross correlation graphs from the Seagate 80GB HDD in each of the different loudness settings, for Loudspeaker-Touching scenario, vs. original audio. Notice how 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.

Surface scenario suggest a greater potential for speech information leakage than the aerial medium scenarios described above. In the normal loudness setting, the enhanced audio files for each HDD show the potential for some speech intelligibility. In the loud settings, the enhanced audio files are almost completely intelligible. Most of the words spoken in the audio file could be transcribed. For the very loud setting, 100% speech intelligibility was observed with the human listener able to understand all speech in the audio file.

Lastly, our analysis results for the *Loudspeaker-Touching* scenario show that direct contact as a transfer medium has the greatest potential for speech information leakage. In the normal loudness setting, the speech in the enhanced audio files for each HDD was partially intelligible. In the loud and very loud settings, the enhanced audio files were 100% intelligible. The human listener was able to recognize and transcribe all speech within the enhanced audio files.

VIII. 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 appear to have significant influence on the strength of the induced vibrations (i.e., the amount of information loss). The difference between transfer mediums is 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 5 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 range (>85dB) consistently produced the data with the most information leakage, while the normal SPL range (40-60dB) 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, confirm these observed trends. Figure 6 displays the cross correlation graphs created for the original audio vs. data collected from the Seagate 80GB 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 recreate 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 tested in [5].

IX. Discussion & Future Work

Limitations: In our research we encountered certain limitations. First, our methodology requires that the HDD tested can remain powered on with the disk spinning for the entirety of each measurement. However, we found initially that some

HDDs do not remain powered on if there are no read or write requests being made. Additionally, for accurate vibration measurements our research also requires that the HDD in each scenario does not move its read/write head for the duration of the measurement. We observed that some HDDs routinely move their read/write heads every few seconds while remaining idly powered on. This behavior was present in the Fujitsu 120GB mini-HDD with the read write head taking a new position every 12 seconds. The read/write head would remain in the new position for approximately four seconds before returning to its original position. We chose to include our measurements for the Fujitsu HDD as we can still observe vibrational effects between the short periods of distorted measurements caused by the movement of the read/write head. These short periods of displacement are clearly visible as dark red bands in Appendix Figure 11 of Section D.

Potential Future Directions: The results from our research have revealed interesting information about the relationship between speech and vibration and how vibration data can be used to extract speech information. Mainly, 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 would like to repeat our experiments with other HDD models to further generalize our investigation. Different solid transfer mediums (i.e. plastic, metal, wood) may also be explored 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 [32] could be fully explored using our laser vibrometry methodology. At what volume can sound actually induce errors in a data center? This question, and many others, can be answered using the laser vibrometry methodology.

X. Other Related Works

Information leakage via vibration measurements has been studied in prior works. In [33], Marquardt et. al showed how vibrations recorded by a smartphone's accelerometer can be used to determine the text typed on the smartphone. In [34], [35] and [36], 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. [37] 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.

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.

XI. Conclusion

In this work, we have analyzed the vibrational impact induced on the read/write heads of different HDD models when exposed to different speech sources (live human, machine-rendered), transfer mediums (aerial, shared surface, direct contact), and loudness levels (normal, loud and very loud). We recreated the results of previous work on this attack to confirm the validity of our vibrometer methodology. We used the same model of HDD and attack parameter settings as the previous work and were able to achieve the same results. Through this study, we observed that the vibrational responses varied between the tested HDDs. Additionally, gleaning speech information in some settings (aerial transfer medium, normal SPL) 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 a new methodology, we experimentally demonstrated criteria (i.e., volume, propagation medium) for eavesdropping attack success under certain standard conditions.

Considering the higher-risk threat model that we used in our study, we suspect that under live settings the severity of vibrational impacts that we observed would be lessened. Sensitive information is generally not spoken (or played back via loudspeakers) in the loud or very loud SPL ranges and HDDs are usually fully enclosed in their own casing and then again within the larger computer casing. Therefore, we believe that sound waves are less likely to induce a vibrational effect on the read/write head of a HDD, significant enough to 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 over speech signals is not as viable as perceived from the results of prior work.

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] A. Rubtsov. (2016) Hdd inside: Tracks and zones. how hard it can be? [Online]. Available: http://hddscan.com/doc/HDD_Tracks_and_Zones.html

[9] (2019) Vibration of hard disk drives. [Online]. Available: https://www.me.washington.edu/research/faculty/ishen/hdd_vibration

[10] 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/

[11] Polytec. (2019) Laser doppler vibrometry. [Online]. Available: https://www.polytec.com/us/vibrometry/technology/$laser-doppler-vibrometry/

[12] ——. (2019) Optical vibration measurement per laser vibrometry. [Online]. Available: https://www.polytec.com/us/vibrometry/areas-of-application/

[13] Wikipedia. (2019) Hard disk drive platter. [Online]. Available: https://en.wikipedia.org/wiki/Hard_disk_drive_platter

[14] Polytec. (2019) Ofv-5000 modular vibrometer. [Online]. Available: https://www.polytec.com/us/vibrometry/products/single-point-vibrometers/ofv-5000-modular-vibrometer/

[15] ——. (2019) Vibroflex. [Online]. Available: https://www.polytec.com/eu/vibrometry/products/single-point-vibrometers/vibroflex/

[16] ——. (2019) Pdv-100 portable digital vibrometer. [Online]. Available: https://www.polytec.com/us/vibrometry/products/single-point-vibrometers/pdv-100-portable-digital-vibrometer/

[17] 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/

[18] ——. (2019) Seagate barracuda 7200.7 (80gb, sata-150) specs. [Online]. Available: https://www.cnet.com/products/seagate-barracuda-7200-7-80gb-sata-150/

[19] ——. (2019) Seagate barracuda 7200.12 (1tb, sata-600) specs. [Online]. Available: https://www.cnet.com/products/seagate-barracuda-7200-12-1tb-sata-600/

[20] ——. (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/

[21] ——. (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/

[22] 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

[23] 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/

[24] (2019) Ultrastar dc hc600 smr series. [Online]. Available: https://www.westerndigital.com/products/data-center-drives/ultrastar-dc-hc600-series-hdd

[25] 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/

[26] 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.

[27] 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

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

[29] C. Forrest. (2016) Microsoft's ai can now understand speech better than humans. [Online]. Available: https://www.techrepublic.com/article/microsofts-ai-can-now-understand-speech-better-than-humans/

[30] E. Myers. (2019) The role of artificial intelligence and machine learning in speech recognition. [Online]. Available: https://www.rev.com/blog/artificial-intelligence-machine-learning-speech-recognition

[31] Rf - snr vs sinad. [Online]. Available: https://www.sharetechnote.com/html/RF_Handbook_SNR.html

[32] 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.

[33] 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.

[34] 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.

[35] 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.

[36] 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.

[37] 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.

[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

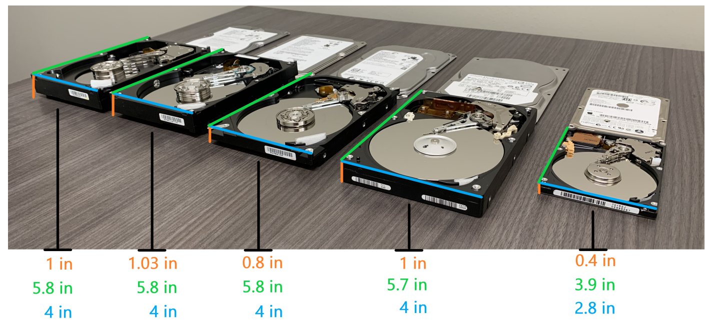
Appendix

A. Characteristics of the Tested Hard Disk Drives

TABLE II: 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	1.40 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

(a) Internal view of all HDDs tested.

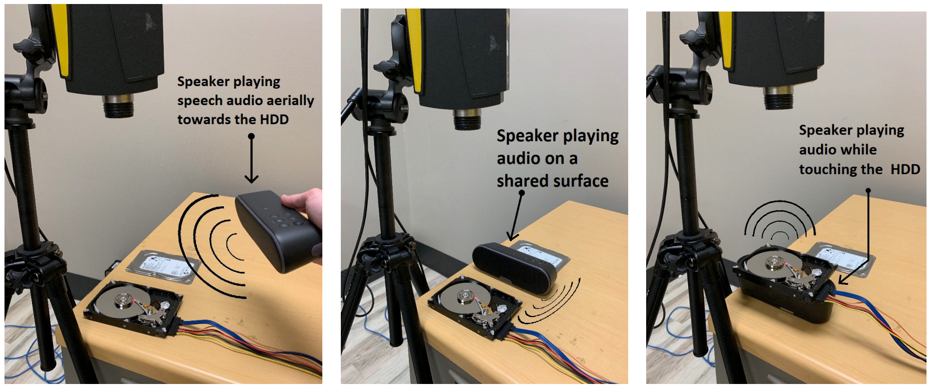


(b) Dimensions of all HDDs used in experiments.

Fig. 7: 5 tested HDDs: (from left) Seagate Barracuda 7200.12 1TB, Seagate Barracuda 250GB, Seagate Barracuda 80GB, Hitachi Deskstar 80GB, & Fujitsu mini 120GB.

Fig. 8: Force diagram showing the natural vibrations internal to HDDs.

B. Setups of the Experimental Speech Source Scenarios



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

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

C. Comparison of Time-Domain Graphs

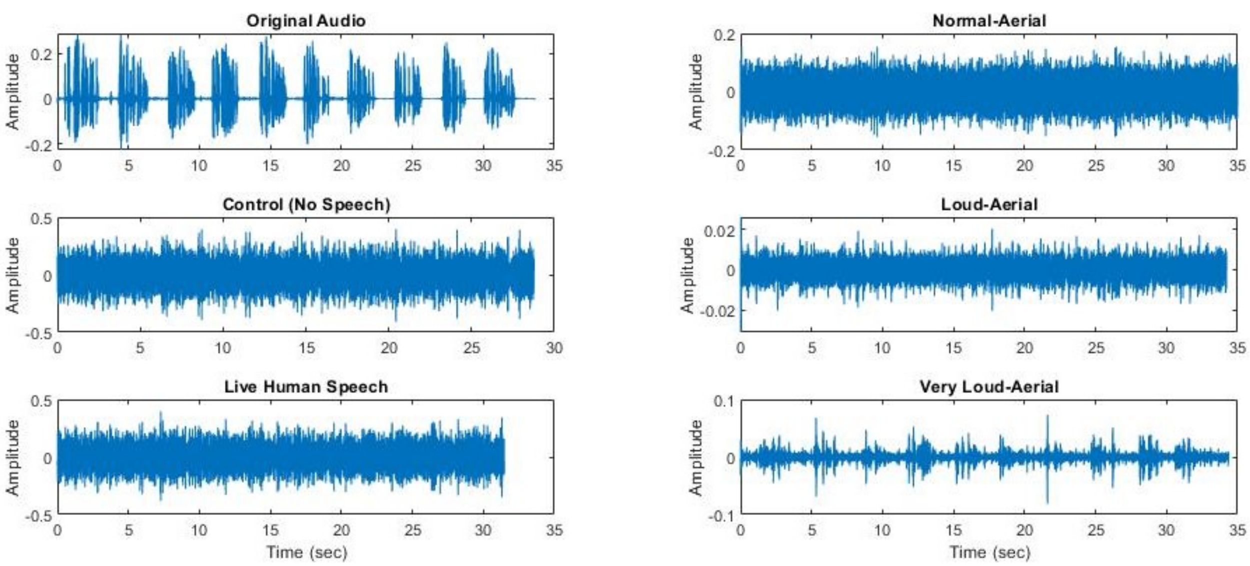
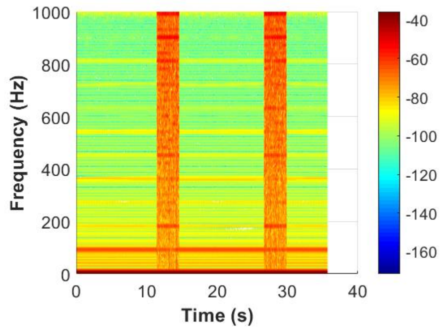
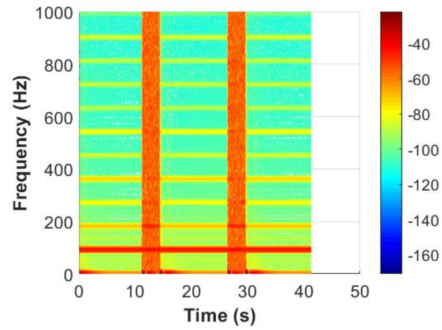


Fig. 10: Side-by-side time domain graphs for original vs. reconstructed audio from each Aerial scenario. Aligning peaks in the Very Loud-Aerial graph indicate information leakage. Control, Live Human, Normal-Aerial, and Loud-Aerial data was collected from Hitachi HDD and Very Loud-Aerial data was collected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

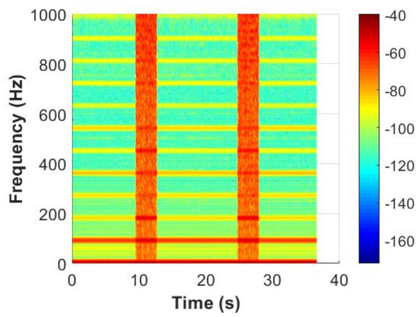
D. Frequency Spectrum Graphs of Vibrometer Measurements



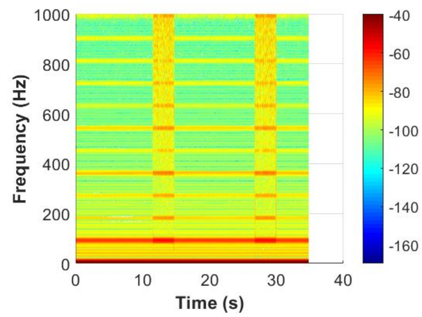
(a) Baseline measurement



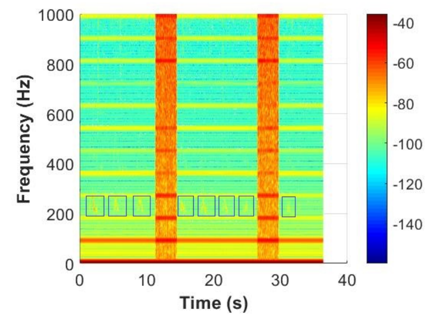
(b) Live human speech



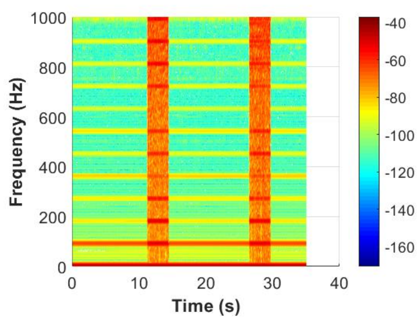
(c) Aerial setup-normal SPL(<70dB)



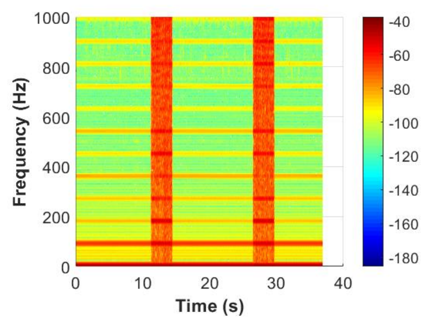
(d) Same-surface setup-normal SPL(<70dB)



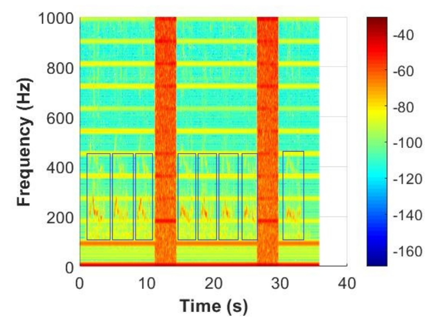
(e) Touching setup-normal SPL(<70dB)



(f) Aerial setup-loud SPL(>70dB)



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



(h) Touching setup-loud SPL(>70dB)

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