MagEar: Eavesdropping via Audio Recovery using Magnetic Side Channel

Qianru Liao[†], Yongzhi Huang[†], Yandao Huang[†§], Yuheng Zhong[†], Huitong Jin[†], Kaishun Wu[†]
[†]Shenzhen University, Shenzhen, China
[§]Hong Kong University of Science and Technology, Hong Kong
{liaoqianru2016,huangyongzhi}@email.szu.edu.cn,yhuangfg@connect.ust.hk
{2019281103,2019151067}@email.szu.edu.cn,wu@szu.edu.cn

ABSTRACT

Speakers have been widely embedded in various electronic devices as a standard configuration. The security vulnerability of microspeakers (such as earphones) is commonly overlooked because it is often assumed that soundproof boundaries, such as walls, can prevent privacy-infringing sound leakage. In this paper, we present the prototype MagEar, an eavesdropping system that leverages magnetic side-channel signals leaked by a microspeaker to recover intelligible human speech. MagEar has sufficiently high sensitivity to detect magnetic fields on the order of nanotesla, exceeding some high-precision magnetometers. It can recover high-quality audio with 90% similarity to the original audio even at a distance of 60 cm. In addition, the MagEar prototype is portable and can be hidden in a headset shell. We have implemented MagEar as a proof-of-concept system and conducted several case studies of eavesdropping on different types of speaker-embedded devices, including earphones, and we have demonstrated the ability to successfully transcribe the recovered speech using automatic speech recognition techniques even when blocked by soundproof walls. We hope that our work can push manufacturers to rethink this security vulnerability of speakers.

CCS CONCEPTS

• **Security and privacy → Human and societal aspects of security and privacy**.

KEYWORDS

Eavesdropping, Side Channel attack, Magnetic field, Mobile security, Privacy disclosure

ACM Reference Format:
Qianru Liao, Yongzhi Huang, Yandao Huang, Yuheng Zhong, Huitong Jin, Kaishun Wu. 2022. MagEar: Eavesdropping via Audio Recovery using Magnetic Side Channel. In *The 20th Annual International Conference on Mobile Systems, Applications and Services (MobiSys '22), June 25–July 1, 2022, Portland, OR, USA.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3498361.3538921

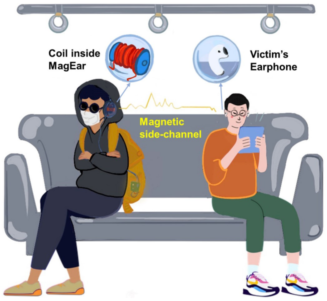


Figure 1: Illustration of MagEar. The adversary disguised MagEar as headphones to eavesdrop on a victim.

1 INTRODUCTION

Recent years have witnessed the growing proliferation of advanced information and communication technology. We are now surrounded by various electronic devices that provide comprehensive services and enrich our daily experience. Acoustic sensors (i.e., speakers and microphones) have been widely embedded in electronic devices as essential components. Yole Développement predicted that the microphone market would continue to expand and reach over 8 billion units by 2022 [2], while the microspeaker market was projected to reach 15.16 billion units by 2021, according to Technavio [1]. However, while enjoying high-quality services supported by these acoustic sensors (e.g., Voice over Internet Protocol (VoIP), remote conferencing, and online infotainment), we are also facing a looming threat of privacy disclosure. Despite many preventive measures, new eavesdropping techniques are still emerging and infringing on users' everyday privacy and security.

State-of-the-art eavesdropping systems have demonstrated the feasibility of harnessing various non-acoustic sensors in smartphones, such as motion sensors [5, 7, 13, 23, 29] and vibration motors [25], to infer sensitive user information. However, such methods assume that the adversary has gained access to the sensor reading by planting malware. The equipment provider or another relevant organization can easily impose restrictions on such hacking by deploying new policies [3]. In addition, previous works have mainly focused on how to turn a device near a victim into a microphone to record human speech. Only a few works have noted the fact that the speaker can "talk" as well. In ART [26], the authors proposed the concept of wireless vibrometry to eavesdrop on loudspeakers. They successfully demodulated a loudspeaker's vibration signal from the received signal strength (RSS) readings of received Wi-Fi packets, thereby recovering the sound of a piano and ten discrete digits. However, for small devices, such as headphones,

MobiSys '22, June 25–July 1, 2022, Portland, OR, USA
© 2022 Association for Computing Machinery.
ACM ISBN 978-1-4503-9185-6/22/06…$15.00
https://doi.org/10.1145/3498361.3538921

ART cannot restore the sound of their speakers. In this paper, we explore a new eavesdropping mechanism and report the use of a lightweight device to recover human speech from speakers, even the micro-speakers in earphones.

We present MagEar, an eavesdropping system that leverages the magnetic field radiated by a speaker to infer what the user is hearing, as shown in Figure 1. We show that the majority of commodity electronic devices that have speakers present an imminent threat of leaking private user information. The key observation lies in the sound generation process of the main component in the speaker (i.e., the transducer). The transducer is responsible for translating electrical signals (digital audio) into mechanical signals (sound). However, during this process, there is an 'intermediate product,' i.e., the changing magnetic field, which can be captured by a simple coil. An adversary may analyze the changes in the magnetic field to infer the original digital audio input. We refer to such a process as a magnetic side-channel attack.

However, we need to overcome several **challenges** before transforming the high-level idea presented above into a practical eavesdropping system.

1) The magnetic field leaked by earphones is weak, at the order of magnitude of only nanotesla (10^{-9} T). The intrinsic attenuation property of the magnetic signal also limits its leakage range. The magnetic field intensity varies inversely with the third power of the distance, i.e., $O(1/d^3)$. With such a weak magnetic field, even a high-precision commercial magnetometer can detect magnetic variations only within 5 cm. Therefore, designing a low-cost device that can eavesdrop from beyond an intimate distance (35 cm) is a considerable challenge.

2) To measure the weak magnetic field, we need to improve the system's sensitivity. However, a high sensitivity will also inevitably make the system more susceptible to noise interference. The noise in our system comes from two sources. First, circuit noise has significant energy at low frequencies that overlap with human speech. We cannot apply a simple filter to separate the effective speech signal from this noise. Second, we are surrounded by a large number of electronic devices that emit electromagnetic radiation at all times. The received signal tends to be contaminated or even overwhelmed by this surrounding noise. Therefore, the second challenge is how to eliminate these two kinds of noise and obtain clean audio.

3) Since a magnetic coil is not designed to record sound like a microphone, it cannot provide a satisfactory listening experience. The audio converted from a magnetic signal will have poor audio quality and be distorted. Investigating the differences between audio and magnetic signals and designing a reasonable repair method is another critical challenge.

We performed an evaluation using 15 types of off-the-shelf electronic devices (i.e., 10 earphones and 5 smartphones). We used these devices to play audio from an open-source corpus and eavesdropped on them using MagEar. The average MOSNet score and cosine similarity for the smartphones were 2.09 and 0.92, respectively, at a 60 cm distance. The corresponding values for the earphones were 1.83 and 0.89, respectively, at a distance of 50 cm.

The main **contributions** of MagEar are summarized as follows:

1) We propose a new side-channel attack scheme to eavesdrop on speaker-embedded devices, including earphones. The results demonstrate the alarming threat of leaking user privacy presented by the majority of COTS speaker-embedded devices.

2) Our designed MagEar system has sufficiently high sensitivity to measure nanotesla-level magnetic fields, exceeding the sensitivity of some high-precision magnetometers.

3) We redesign a coil based on the physical model to make its ability equivalent to that of a coil of three times the diameter. We also eliminate noise interference and improve the quality of the distorted audio. The audio quality of the recovered speech at a distance of 60 cm is equivalent to that of the original audio subjected to a 1/3 downsampling rate. The recovered audio can be found on our website[1].

4) We have conducted experiments in various scenarios to study the feasibility of MagEar. We have found that the eavesdropped signal can be recovered as intelligible speech and transcribed directly using a COTS automatic speech recognition service.

2 THREAT MODEL

We assume that the victim is wearing headphones or earphones and that the adversary uses a receiver coil. By measuring the leaked magnetic field near the victim's headphones and applying signal processing algorithms, the adversary can recover the contents of the victim's audio. The victim may be using headphones to make a phone call, participate in a virtual meeting, or listen to media. The audio restored by MagEar can be played directly and recognized by a human listener and can also be transcribed by existing speech-to-text API. Thus, the adversary can infer the target's private information, business details, and personal interests and hobbies from the recovered audio.

We assume that the victim and attacker are located not far from each other. For example, during the morning and evening rush hours, subway or bus carriages are often crowded with passengers, and everyone inevitably comes into physical contact with others. For passengers sitting in seats, the shoulders of adjacent passengers are often in contact. As shown in Figure 1, an attacker can hide a receiver coil inside a headphone and eavesdrop on adjacent passengers' headphones or earphones. In addition, when an attacker and victims are sitting on the same bench located in a waiting room, airport lounge, or park, the audio contents played by the victims' earphones are at risk of leakage.

On the other hand, a receiver coil can act as an eavesdropping tool without human participation. A malicious coil can be placed on any object that people may stand or sit next to. Since some passengers will commonly wear headphones on the train or bus for a long time and seldom check their seats, an adversary may embed coils in train and bus seats to record nearby audio. Employees at a company are also likely to go to an empty conference room for a personal call or a virtual meeting. In this scenario, a conference room table or chair is also a suitable place to conceal coils that can eavesdrop on private or business conversations. In addition, people sometimes engage in phone conversations at coffee shops or restaurants or while sitting on park benches, thus potentially exposing themselves to MagEar.

[1] https://github.com/lkeyee/MagEar

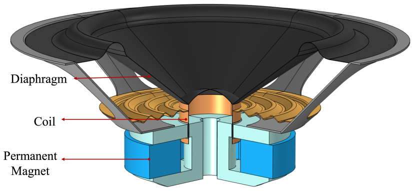


Figure 2: Simulation model of a dynamic speaker

3 BACKGROUND AND PRINCIPLE

As is well known, the speaker is the core module of an earphone, receiving the device's signal and producing sound. Although various speaker structure designs exist, nonprofessional earphones commonly use dynamic unit speakers. In this section, we will discuss the working principle of a speaker before investigating the relationship between the input audio and the magnetic field inside the speaker.

3.1 Working Principle of Speakers

The main speaker module is the transducer, which converts electrical energy into mechanical energy [27]. A dynamic transducer comprises three critical components: 1) a magnet system, including a permanent magnet and T iron; 2) a vibration system, consisting of a voice coil and a diaphragm; and 3) parts providing structural stability. We designed a simulation model in COMSOL based on the speaker structure [9], as shown in Figure 2.

When an audio current flows through the voice coil, the ring structure will generate the Oersted current magnetic effect. A change in the magnetic flux of the voice coil generates a force on the static magnetic field of the magnet system, which, in turn, drives the movement of the diaphragm.

3.2 Mathematical Model of Speakers

Because the diaphragm's elastic force constrains the voice coil's motion, it undergoes simple harmonic motion. However, the motion of the voice coil also changes the magnetic flux of the static magnetic field through the voice coil. Subject to Lenz's law, the impedance of the voice coil hinders the current generated by the magnetic flux changes, causing resistance to the harmonic motion. Finally, the simple harmonic motion is transformed into damped motion described by the following motion equation [27]:

$$m\frac{d^2x}{dt^2} = F_{mag} - c\frac{dx}{dt} - kx \tag{1}$$

where m is the combined mass of the diaphragm and coil, x is the displacement of the diaphragm, and F_{mag} is the magnetic force. The damping coefficient c and the spring constant k depend on the material of the diaphragm. Solving this differential equation, we obtain an approximate expression for the displacement [16, 17]:

$$x \propto F_{mag}\sin(\omega t + \varphi) \tag{2}$$

From Equation (2), we observed that the displacement of the diaphragm is proportional to the magnetic force.

Regarding the magnetic force, the direction of the coil vibration is perpendicular to the magnetic field. Thus, the magnetic force acting on the coil follows the Fleming rule. The magnetic force can be expressed as follows:

$$F_{mag} = BIl \tag{3}$$

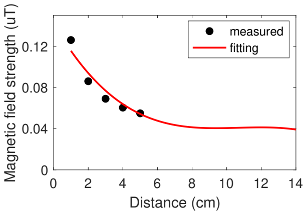


Figure 3: Electromagnetic leakage of earphones measured at various distances.

where B is the total magnetic flux density, I is the audio current flowing through the coil, and l is the length of the coil. Based on Equations (2) and (3), if the diaphragm needs to produce sound, it must be driven by sufficient magnetic field changes, presenting a risk of privacy leakage. However, the momentum of the voice coil and diaphragm that is required when transmitting sound may be tiny, which means that the earphone's magnetic leakage is weak. Therefore, is it necessary to worry about the magnetic leakage risk?

3.3 Magnetic Leakage

To answer the question raised in the previous section, we need to measure the magnitude of the magnetic field leaked by a headset. For this purpose, we generated a 1 kHz sine wave in Python and played it with Airpods. Then, we used an electromagnetic field tester TES-1394S with a resolution of 0.001 μT to measure the magnetic field near the Airpods. The angle between the magnetometer and Airpods was 0°. We recorded measurements at various distances from the Airpods.

As shown in Figure 3, this high-precision magnetometer could only detect magnetic leakage from no more than 5 cm away from the earphones. We fitted the curve of the magnetometer (dotted line), and the result confirms that the magnetic leakage from the earphones is weak.

According to Faraday's law of induction, if we adopt a magnetic coil as our receiver to measure the magnetic field near the speaker, the induced electromotive force in the coil will be proportional to the changing rate of the magnetic flux density[19]:

$$V = -NA\frac{dB}{dt} \tag{4}$$

where V is the induced voltage in the coil, N is the number of turns, A is the cross-sectional area of the coil, and B is the magnetic flux density.

Theoretically, if we could infinitely increase the coil area A and the number of coils N, we could easily measure the magnetic signal leaked by an earphone speaker. However, this approach cannot be feasibly applied by eavesdroppers. Moreover, the change in the magnetic signal generated by an earphone speaker is not linear. If we take the sine function $B(t) = B_m \sin \omega t$ as the leaked magnetic signal, we can obtain the following equation:

$$V = -\omega NAB_m \cos \omega t \tag{5}$$

This formulation explains why the magnetic leakage from other electronic devices is easier to obtain than that from earphones. The working frequency of a commercial electronic device tends to be high; it generally exceeds one hundred megahertz (10^8 Hz) or one gigahertz (10^9 Hz) and may even be greater than 10^{11} Hz in an optical communication device. In contrast, the frequencies of the sounds played by earphones are mostly in the range of 200 to

2000 Hz, which presents an enormous challenge for detection. It should be noted that for practical reasons, we need to use a coil instead of an antenna. A suitable antenna could indeed achieve a more significant gain. However, the antenna size needed for a reasonable receiving range depends on the signal's wavelength, and even the most miniature monopole antenna needs to reach the 1/4-wavelength scale. Therefore, the minimum antenna size needed for receiving earphone magnetic leakage would be 100 kilometers.

For the reasons discussed above, eavesdropping on magnetic leakage has seemed infeasible to date. Nevertheless, it is possible to obtain and even recover the sounds played by earphones from their leakage. The following section will present a step-by-step explanation of how we have overcome this challenge.

4 MAGEAR DESIGN

4.1 Experimental Setup

In this section, we describe how we design a receiver coil to measure the weak magnetic field leaked by earphones. We first theoretically analyze how the parameters of a coil affect various aspects of its performance, such as the passing magnetic flux, induced voltage, sensitivity, signal-to-noise ratio, and resonance effect. In addition, we report corresponding experiments conducted to verify these theoretical results. In these experiments, we generated 1 kHz sine wave audio in Python and played it with Apple Airpods (1st generation). At the receiver end, a copper coil was connected to a USB3202 data acquisition board with a 16-bit analog-to-digital converter (ADC) to collect the leaked magnetic field. The angle between the Airpods and the coil was set to 0° to maximize the passing magnetic flux. In addition, both the receiver coil and the headphones are placed on the table, and the measured distance refers to the horizontal distance between their centers. Then, we recorded the amplitude at 1 kHz in the measured signal. Since an amplifier will introduce high circuit noise at low frequencies, we did not connect the coil to the amplifier in this section in order to better investigate the coil's characteristics. All experiments were performed in the laboratory.

4.2 Geometric Shape

Since an earphone has both a built-in static magnetic field and a dynamic magnetic field driven by a coil, the magnetic field configuration is complex. Traditional research is based on ideal magnetic field conditions. Therefore, to obtain the best magnetic field effect, we first need to determine the optimal geometric shape of the coil. Under the assumption that the area of the receiver shape is $A = \frac{\pi D^2}{4}$, Equation (5) can be rewritten as follows:

$$V_m = \frac{1}{2}\pi^2 f N D^2 B_m \tag{6}$$

where f is the frequency of the measured magnetic field. According to Equation (6), theoretically, as long as the receiver area and the number of turns are the same, the induced potential is the same. However, in reality, this is not the case because the magnetic flux density on the receiving plane varies. To explain this phenomenon, we need to develop a physical model.

Although the positions of the voice coils of the speakers used in our experiments differ, the coil thickness of the speakers is only 1 mm, much smaller than the experimental distance (at the centimeter level). The magnetic field of an N-turn coil is equivalent to N times that of a single coil. Each single-turn coil can be regarded as a magnetic dipole. A magnetic dipole refers to a small closed loop carrying an electric current and can be defined as $m = IS$, where m is the magnetic dipole moment, I is the current flowing through the loop, and S is the area of the loop. The external magnetic field produced by a magnetic dipole can be written as follows [18]:

$$B = \frac{\mu_0 m}{4\pi R_{distance}^3} \tag{7}$$

where B is the magnetic flux density, $R_{distance}$ is the distance in the radial direction, and μ_0 is the vacuum permeability constant.

As shown in Figure 4, we have designed three receivers, each with an area of $A = \frac{\pi D^2}{4}$ but with different shapes: a circle, a square and an equilateral triangle. We can calculate that the circle's radius is $\frac{D}{2}$, and the side lengths of the square and triangle are $\frac{\sqrt{\pi}D}{2}$ and $\sqrt{\frac{\pi}{\sqrt{3}}}D$, respectively. We place the three shapes at the same center position and calculate their individual magnetic fluxes. According to formula (7), the magnetic field intensity is highest at the center point. If we assume that the magnetic field at the center point is $\frac{1}{R^3}$, we can obtain the following magnetic flux: $\phi \propto \frac{\mu_0 m}{4\pi} \iint \frac{1}{(R_{distance}+r \sin \arctan \frac{r}{R_{distance}})^3} drd\theta$. We can easily conclude that the circle has the greatest magnetic flux, nearly 1.01 times that of the square and 1.02 times that of the triangle.

For experimental testing, we manually wound four coils of different shapes: a circle, a square, a rectangle, and an equilateral triangle. They shared the same area of 7.065 cm^2 and the same height of 3 cm. The radius of the circle was 1.5 cm, and the side lengths of the square and triangle were 2.65 cm and 4.04 cm, respectively. The length and width of the rectangle were 3.5 cm and 2 cm, respectively. Then, we recorded their amplitudes at 1 kHz at different distances. Figure 5 shows the results, which indicate that the circular coil produces the maximum induced voltage.

4.3 Geometric Size

From the theoretical results and actual measurements presented in the previous section, we can see that a circle is the most advantageous coil shape in terms of the total magnetic flux. In this section, we investigate how to choose the diameter.

We designed 8 coils with different diameters ($D \in \{4, 6, ..., 16\}$) and the same number of turns (N=5000). Then, we played a 1 kHz sine wave with Airpods and measured the amplitudes at 1 kHz of the different coils. Figure 6 shows the relationships between the diameter and the induced voltage at different distances. The amplitude increases with increasing diameter.

Although Equation (6) indicates that the induced voltage should increase in proportion to the square of the diameter D, the experimental results in Figure 6 contradict this expectation. Instead, the voltage growth rate gradually decreases. This is not because Equation (6) is incorrect. To explain this phenomenon, we simplify the variation process of the induced voltage as follows. When we add a differential diameter dD, the induced voltage changes to $V + dV \propto (D + dD)^2$. However, the magnetic flux on the differential diameter is $dB \propto (\frac{1}{dD})^3$. Because the diameter must be a non-negative number, we can obtain $(D + dD)^2 > D^2 + dD^2$.

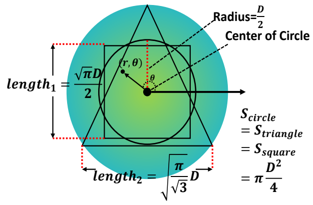

Radius=$\frac{D}{2}$
Center of Circle
(r,θ)
θ
$length_1 = \frac{\sqrt{\pi}D}{2}$
$length_2 = \sqrt{\frac{\pi}{\sqrt{3}}}D$
$S_{circle} = S_{triangle} = S_{square} = \pi\frac{D^2}{4}$

Figure 4: Magnetic fluxes for receivers of different shapes.

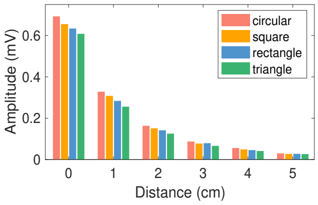


Figure 5: Induced potentials of of differently shaped coils. The results for the circular coil are the best.

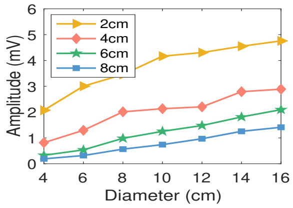


Figure 6: Coil amplitude vs. diameter at different distances.

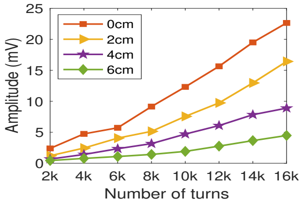


Figure 7: Coil amplitude vs. number of turns at different distances.

Therefore, the increased magnetic flux can be approximated as $B + dB = B + ((\frac{1}{dD})^3)^2 = B + \frac{1}{dD}$.

Although an increase in diameter will increase the induced potential, an eavesdropper cannot carry a device that is too heavy. We believe that eavesdroppers will attempt to limit the size of their eavesdropping devices to better hide their behavior. As the diameter grows, the incremental increase in the magnetic flux gradually weakens. Therefore, an eavesdropper will choose a coil with high diameter performance, which means that for the chosen diameter, increasing the unit mass of the coil (for example, by increasing the number of turns) will result in faster growth in the electrical potential.

We calculate the increase ΔV in the induced potential caused by the coil diameter expansion ΔD, and the corresponding equation can be written as $E_{diameter} = \frac{\Delta V}{\Delta D}$.

As shown in Figure 8, a coil with a diameter of 14 cm has a more stable effect with distance than smaller coils; however, the earphones commercially available on the market do not exceed 8 cm in size. Therefore, we choose a diameter of 4 cm, which offers the highest performance on average.

We also investigated how the number of turns affects the induced potential. Specifically, we designed another 8 coils with different numbers of turns ($N \in \{2000, 4000, 6000, ..., 16000\}$) and the same diameter (D=8 cm). The experimental process was the same as in the diameter experiment. Figure 7 reveals that the relationship of the amplitude with the number of turns presents a nearly linear variation.

4.4 Fill or Vacant the Coil's Center

Based on the selected planar shape and diameter of the coil, in this section, we explore the coil's center structure. Because the complicated coil structure makes it challenging to analytically express the magnetic flux through the coil, we can only measure it through experiments.

The greater the magnetic flux that passes through the coil, the more significant the induced potential that will be generated. This is why in traditional induction coil design, the area is enlarged as much as possible. However, in the eavesdropping scenario, the diameter of the coil is limited. In this case, increasing the induced potential becomes a challenge.

When the total magnetic flux remains the same, if we fill the middle space of the coil with wires, the same magnetic flux will drive more wires. Will the induced potential in this case be greater than that in the unfilled case? We designed a coil to answer this question. As shown in Figure 10, we prepared the new coil by winding the wire in a spiral shape.

To control the variables, we limit the coil in the plane. We compared the experimental results with those in Section 4.3 (5000 turns). Considering the difficulty of collecting the induced potential with one turn, we compared the result with 1/10 of the value for a 4 cm diameter (approximately 0.52), the same ratio as the wire length used in this experiment.

In Figure 9, we can see that when the diameter is restricted to 4 cm, the spiral-filled coil can produce a higher induced potential (approximately 1.15 at 0 cm), two times greater than that of a standard hollow coil with the same wire length (approximately 0.52).

In experiments, we compared various winding shapes at the center of the coil, such as polygonal (triangular, square, etc.) windings, and we found that the effect was not as good as that of the spiral. This is because the directivity of the induced voltage with a linear wire shape will result in attenuation of the induced voltage. As illustrated in Figure 11, the induced voltages in the cases of linear and circular wire shapes can be expressed as follows:

$$\begin{aligned} dV_{circle} &\propto \frac{\theta}{2\pi R^3} d\theta \\ dV_{line} &\propto \frac{1}{(R+|x| \sin\arctan\frac{x}{R})^3} \sin\arctan\frac{x}{R} dx \end{aligned} \tag{8}$$

Through the substitution method, we find that when the arc of the wire is closer to a circle, the induced voltage generated by the magnetic field will be higher.

4.5 Optimizing Performance via Coil Design

Because the electrical signal generated by magnetic leakage is tiny, we need to connect an operational amplifier (op-amp) to amplify it. However, high unexpected noise is introduced when the coil is connected to the op-amp in a shielded room without any magnetic signal. This noise is *Johnson–Nyquist noise*, originating mainly from the movement of electric charges inside the circuit. To reduce the influence of signal access to the op-amp, we need to optimize the design of the coil. Two reference parameters that we need to maximize are the **signal-to-noise ratio (SNR)** and **sensitivity**.

The *SNR* measures the proportion of *Johnson–Nyquist noise* in the signal; a higher value indicates that a clearer signal is obtained. In addition, we can attempt to increase the *sensitivity* to allow even a small change in the magnetic flux to result in a high induced potential.

As determined in the previous section, the best coil design is a spiral shape. In other words, we need to wrap the wire in a spiral shape to form a cylinder layer by layer, with the cross section shown

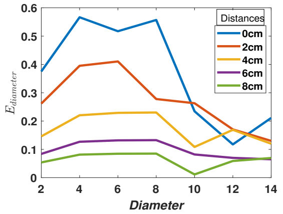


Figure 8: Coil performance vs. diameter at different distances.

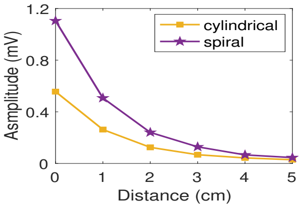


Figure 9: Induced potential of spiral vs. cylindrical coils.

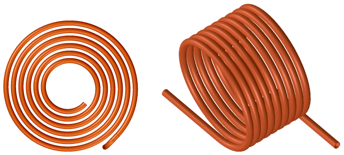


Figure 10: Spiral coil vs. Unfilled coil.

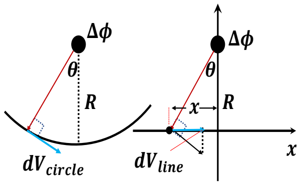


Figure 11: Induced potential with different wire shapes.

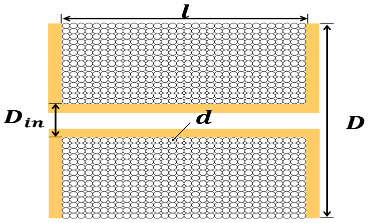


Figure 12: Cross section of a coil

in Figure 12. Under the assumption that the center is a hollow circle, we need to calculate the induced voltage accordingly.

Calculating the induced voltage directly is a complicated problem. We can simplify the problem by approximating each turn of the spiral as a circular coil. Therefore, the induced voltage is proportional to the sum of the areas of these coils. The sum of the areas of all coils in a plane (a vertical row of circles in Figure 12) is $\sum\limits_{n} \frac{\pi(D_{in}+nd)^2}{4} = \frac{\pi}{4}\sum\limits_{n}(D_{in}+nd)^2$, where $n = 0, ..., \frac{D-D_{in}}{d}$. To eliminate tedious calculations, we use the following conversion method. We assign two coils of different diameters to a group such that their sum is equal to $D+D_{in}$. Using Cauchy’s inequality, $\frac{a^2+b^2}{2} \geq (\frac{a+b}{2})^2$, the sum area for each group of coils can then be reduced to the same equation, $\frac{\pi}{4}(D^2+{D_{in}}^2) \geq \frac{\pi}{8}(D+D_{in})^2$. We divide all coils into N groups in this way and substitute Equation (5) for the induced voltage.

$$V_m = Nf\frac{\pi^2}{4}(D+D_{in})^2B_m \tag{9}$$

The number of turns N can be determined from the wire diameter d and the packing factor(a measure of tightness) k:

$$N = \frac{1}{2k}\frac{l}{d}\frac{(D-D_{in})}{d} = \frac{l(D-D_{in})}{2kd^2} \tag{10}$$

The ***sensitivity*** represents the change in the induced voltage due to the magnetic field strength. We combine Equation (9) and Equation (10) to obtain the ***sensitivity*** as follows:

$$Sensitivity = \frac{Vm}{Bm} = \frac{\pi^2}{8}\frac{l(D-D_{in})(D+D_{in})^2}{kd^2}f \tag{11}$$

From this equation, we can see that the induced voltage can be increased by increasing the height l of the cylinder, reducing the diameter d of the wire, or making the winding gaps smaller, that is, minimizing k. We show corresponding experimental results in Figure 13.

On the other hand, we can attempt to maximize $(D-D_{in})(D+D_{in})^2$. Because we have placed a limit the maximum diameter, we can divide both sides of the formula by D^3 to recast it as a function expressed in terms of $\frac{D_{in}}{D}$.

$$y = -(\frac{D_{in}}{D})^3 - (\frac{D_{in}}{D})^2 + \frac{D_{in}}{D} + 1 \tag{12}$$

We can conclude that Equation (12) reaches its maximum when $\frac{D_{in}}{D} = \frac{1}{3}$.

Next, we will calculate the ***SNR***. The *Johnson–Nyquist noise* caused by the coils can be expressed as $V_T = 2\sqrt{k_BT\Delta fR}$, where k_B is Boltzmann’s constant, T is the temperature, and R is the resistance value. The resistance of the coil can be calculated as follows:

$$R = \frac{\rho L}{A_{wire}} \tag{13}$$

where ρ is the resistivity constant of the material, L is the length of the wire, and A_{wire} is the cross-sectional area of the wire, expressed as $A_{wire} = \frac{\pi d^2}{4}$.

The total length L can be obtained by adding the circumferences of the different-diameter coils. All circumferences can be expressed as

$$\begin{aligned} L &= \pi\sum_{n}(D_{in}+nd)\frac{l}{d} \\ &= \pi\frac{(D_{in}+D)(D-D_{in})}{2d}\frac{l}{d} = \frac{\pi(D_{in}+D)(D-D_{in})l}{2d^2} \end{aligned} \tag{14}$$

where $n = 0, ..., \frac{D-D_{in}}{d}$. Using the Gaussian formula, we can obtain a simplified equation. Combining Equation (13) and Equation (14), we obtain

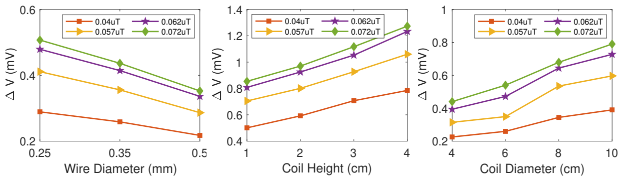
$$R = \frac{2\rho l}{d^4}(D-D_{in})(D+D_{in}) \tag{15}$$

Therefore, the ***SNR*** of the coil can be written as follows:

$$SNR = \frac{V_m}{V_T} = \frac{\pi^2B_mf}{16k\sqrt{2k_BT\rho\Delta f}}\sqrt{l(D-D_{in})(D+D_{in})^3} \tag{16}$$

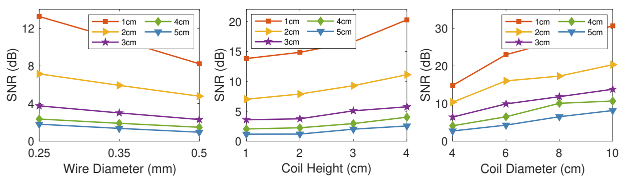
In this SNR model, the first part of the polynomial is the parameters. In the second part, we know that an increase in the height l of the cylinder enhances the ***SNR***, consistent with the experimental results in Figure 14. Second, we obtain the peak value of the ***SNR*** when $\frac{D_{in}}{D} = \frac{1}{2}$. Therefore, we choose a 0.25 mm wire diameter and use a whole-earphone thickness (for ease of camouflage) of 3.5 cm as the height of the coil. Moreover, we set the ratio of the inner and outer diameters to between $\frac{1}{3}$ and $\frac{1}{2}$.

We conducted experiments to determine whether the actual measurements would follow the theoretical analysis. Since the sensitivity refers to the voltage change induced by a unit of magnetic flux, we measured the variation in the induced voltage under different changes in the magnetic flux to study the sensitivity. First, we designed three coils with wire diameters of 0.25 mm, 0.35 mm and 0.5 mm. They had the same diameter (3 cm) and the same number of turns (500). Figure 13 (a) and Figure 14 (a) show that a coil with a smaller wire diameter has a higher sensitivity and SNR. We also designed four coils with different heights ($l \in \{1, 2, 3, 4\}$), the same diameter (3 cm), and the same wire diameter (0.25 mm). Figure 13 (b) and Figure 14 (b) similarly reveal that a coil with a greater height will achieve a higher sensitivity and SNR. To investigate



(a) (b) (c)

Figure 13: Relationships between sensitivity and (a) wire diameter, (b) coil height, and (c) coil diameter at different distances



(a) (b) (c)

Figure 14: Relationships between SNR and (a) wire diameter, (b) coil height, and (c) coil diameter at different distances

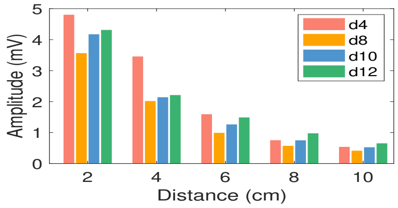


Figure 15: Optimized coil(D=4cm) vs. larger-diameter coils(D=8,10,12cm).

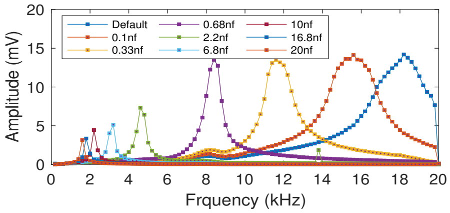


Figure 16: Frequency response curves when connected in parallel with different capacitors.

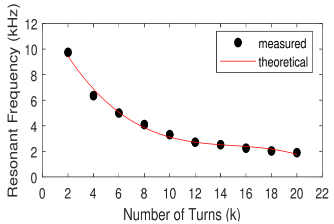


Figure 17: Peak induced potential and coil turns.

the influence of the coil diameter, we used 8 coils with different diameters ($D \in \{4, 6, 8, 10\}$). The results are shown in Figure 13 (c) and Figure 14 (c). In accordance with Equation (11) and Equation (16), the sensitivity and SNR are proportional to the coil diameter. We also added an iron core in the gap of the inner diameter to further increase the inductance of the entire coil and performed experiments with the optimized coil. In Figure 15, the label 'd4' indicates the induced voltage with the optimized coil at different distances. We can observe that the optimized coil produces an induced voltage that is higher than that of a coil with an unoptimized diameter that is larger by a factor of three (labeled 'd12').

In these experiments, we assumed the frequency to be a fixed value; however, this will not be true in practice. Therefore, in the next section, we will introduce the impact of frequency.

5 FEATURE ENHANCEMENT

5.1 Frequency Band Adjustment

Since the coil needs to be connected to an amplifier and ADC to further amplify the magnetic signal to be processed by a computer, the weak magnetic signal can easily be drowned out by inevitable circuit noise. Considering that the circuit noise is not a single-frequency signal but is composed of multiple frequencies concentrated in the low-frequency range, we can adjust the resonance effect of the coil to weaken the circuit noise and boost the audio band. To this end, we will introduce the resonance effect of a coil and how we can tune the resonant frequency.

The receiver of MagEar is a coil, which also acts as an inductance. Its impedance is expressed as $Z_L = j\omega L$, which will increase linearly as the frequency increases. This seems to indicate that the impedance can offset the amplifier effect caused by frequency. Therefore, we tested signals of different frequencies.

In this experiment, we generated audio with frequencies ranging from 200 Hz to 20 kHz with a step size of 200 Hz in Python and played it with Airpods. We used a coil with a diameter of 4 cm as a receiver. Then, we connected different capacitors in parallel with the coil and plotted the frequency response curves. The results are shown in Figure 16.

The curve labeled *'Default'* in Figure 16 is the result of our experiments. We find that, in contrast to the theory, the curve does not rise in a straight line, nor is it flat; instead, it is a bell-shaped curve. In fact, the inductive reactance of an inductor changes slowly when the frequency is low, close to a horizontal straight line. However, this does not explain why the induced voltage rises faster when the signal reaches 14–18 kHz.

Upon experimenting with higher frequencies (>18 kHz), we found that the induced voltage will fall at an accelerated rate after reaching the peak value. We also tested coils with different numbers of turns. As shown in Figure 17, an increase in the number of turns of the coil will reduce the peak frequency. This phenomenon mainly arises from the tight winding of the coils, as a tightly wound coil will cause capacitance between the wires. Consequently, the entire loop becomes a series-connected LC resonant circuit. The resonance peak of our coil is located near 18 kHz, which results in lower-frequency magnetic signals having a weak amplitude.

The frequencies of the sounds played by general earphones are between 50 Hz and 20 kHz. Since we cannot predict the sound frequencies that will be played by the earphones in advance, increasing the induced voltage signal at the proper frequency is a challenge.

However, we have found that if we connect a capacitor outside the circuit of the coil, the center frequency of the resonance will change. In this way, we can obtain a typical LC resonant circuit with a center frequency of $f_0 = \frac{1}{2\pi\sqrt{LC}}$.

When we connect a capacitor in parallel, as shown in Figure 16, the system becomes a complex multistage circuit. In this case, the resonance center frequency depends on two parts of the circuit: the RLC circuit in series and the capacitor circuit in parallel with it. The total admittance of this circuit is

$$\begin{aligned} Y_{total} &= Y_{RLC} + Y_C \\ &= -j\frac{\omega_0 C}{(\omega_0^2 L_{coil} C + \omega_0 CR - 1)} + j\omega_0 C_{EXT} \end{aligned} \tag{17}$$

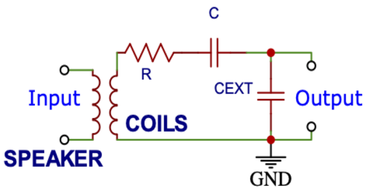


Figure 18: Equivalent circuit of the coil loaded with an external capacitance and resistance.

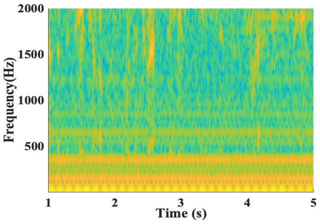


Figure 19: Noise spectrogram.

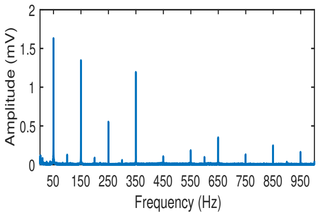


Figure 20: Noise with 50Hz odd harmonics in the frequency spectrum.

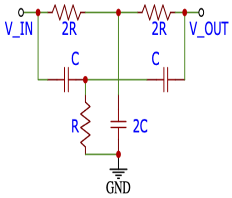


Figure 21: AC noise removal component.

where $\omega_0 = 2\pi f$. If we set the admittance to $Y_{total} = 0$, we can solve for the root of the equation for the center frequency and obtain $f_0 \propto \frac{1}{\sqrt{C_{EXT}}}$. Therefore, as the capacitance C_{EXT} increases, the center frequency will shift to a lower value. Accordingly, as shown in Figure 16, we can use variable capacitors to adjust the center frequency. A larger capacitance is easily able to shift the resonance frequency toward lower values. Specifically, the resonance frequency is inversely linearly proportional to the capacitance.

Notably, different types of sounds have different frequency ranges. When MagEar is to be used to eavesdrop on speech, a small bandwidth will often be required. However, if the target sound is music, then a broader bandwidth may reduce the degree of sound distortion. The resonance frequency f_0 and the quality factor Q determine the bandwidth $BW = \frac{f_0}{Q}$. Accordingly, we can obtain Q=$\frac{R_0}{2\pi f_0 L}$, where R_0 is the resistance and L is the self-inductance.

Specifically, we tried connecting a variable resistor outside the circuit-sliding varistor. With this configuration, when we need a wider bandwidth, we can adjust the variable resistor to decrease the impedance, which will reduce the quality factor Q and increase the bandwidth. Figure 19 shows the noise spectrogram. We can observe that the circuit noise is concentrated below 400 Hz and persists throughout the duration of the audio. Through experiments, we have found that tuning the resonance frequency of the coil to the range of 1500~2500 Hz can yield good performance, minimizing the impact on the audio band while attenuating the circuit noise.

5.2 Interference Confrontation

When we recorded the magnetic signal leaked from the earphones and converted it into sound, we found that the sound was obscured by considerable noise. Such noise will exist even outdoors and tends to be more serious indoors. We present the frequency spectrum of a recorded magnetic signal in Figure 20.

We can observe a sharp increase to the maximum amplitude at approximately 50 Hz. There are also other noise peaks at all odd multiples of 50 Hz. These findings indicate that there is a noise signal with a fundamental frequency of 50 Hz that causes other odd harmonics. Considering that our power supply comes from a battery with a stable voltage, this component may be noise originating from the AC cable, known as hum. Such hum reduces the intelligibility of the recovered audio and is unpleasant to listen to. To mitigate this noise, we created a notch filter module. Figure 21 shows its circuit diagram. We connected it to the output signal and used a three-stage cascade method to eliminate the noise interference from AC power.

Although were able to eliminate the loudest interference in this way, unfortunately, the noise problem was not completely overcome. Considering that the circuit noise is much larger in amplitude than the target magnetic signal and exists throughout the audio duration, merely adjusting the resonance effect of the coil is insufficient to eliminate this interference. In addition, electromagnetic radiation from ambient electronic devices can cause the system performance to degrade. To remove circuit noise and ambient electromagnetic interference, we recorded the magnetic noise during a period of silence and established a 1 s sliding window to compare the differences in the frequency domain. Moreover, we used the Allan variance for analysis. The Allan variance was initially proposed to measure frequency stability in clocks. It can be used to characterize different types of noise in sensor data. As shown in Figure 19, we found that the noise amplitude is negatively correlated with frequency in the range of 0–1000 Hz. The noise slope remains stable ($Amplitude = -10^{-4}f$). The Allan variance results in Figure 22 show that the signal mainly exhibits white noise throughout the whole frequency band, as the slope of the curve on a double logarithmic graph is close to -1. The peak between Tau values of 10^{-3} to 10^{-2} indicates the presence of sinusoidal noise between 100 Hz and 10000 Hz.

Considering that this noise is a type of additive noise and is uncorrelated with the target speech, we can employ spectral subtraction for noise reduction. The spectral subtraction procedure consists of estimating the noise spectrum during periods of silence and then subtracting this from the original spectrum to obtain a clean speech signal.

In detail, the calculation of the power spectrum by subtracting the noise spectrum from the original spectrum to obtain the clean spectrum is expressed as follows [8]:

$$|X(\omega)|^2 = \begin{cases} |Y(\omega)|^2 - \alpha|D(\omega)|^2 & |Y(\omega)|^2 > (\alpha+\beta)|D(\omega)|^2 \\ \beta|D(\omega)|^2 & else \end{cases} \tag{18}$$

where $Y(\omega)$ is the speech signal, $X(\omega)$ is the clean speech signal, $D(\omega)$ is the additive noise signal, α is the subtraction factor and β is the spectral floor parameter. The parameter α controls the amount of speech distortion. If α is too large, then some speech information will be eliminated along with the noise, reducing the intelligibility of the speech. However, if α is too small, then considerable noise will remain. In our system, we set β = 0.02 and select the value of α as described in [8].

Then, by applying the inverse discrete Fourier transform to the power spectrum, the denoised time-domain signal can be acquired.

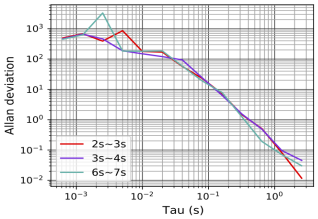


Figure 22: Allan deviations of noise at different times(seconds).

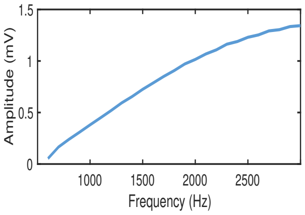


Figure 23: Induced voltage proportional to the frequency.

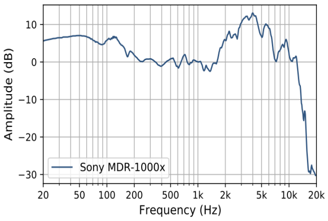


Figure 24: Frequency response curve of the Sony MDR-1000x.

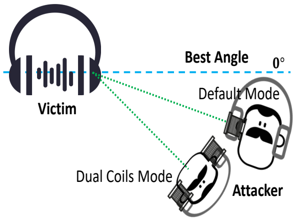


Figure 25: Two operation modes at different angles.

5.3 Frequency Response Equalization

As described in the previous section, we can achieve noise elimination, but the audio collected by MagEar is still distorted. Specifically, the sound is sharper than the original audio. This deformation has two causes. One is the fact that the magnetic field of the earphones is not equivalent to the vibration amplitude, and the other is the damping effect introduced by the resonance of the coil.

First, let us discuss the inequality between the earphones' magnetic field and the vibration amplitude. Since the magnetic force is used to provide the acceleration of the earphone coil, when the coil needs to vibrate quickly, the acceleration will increase to drive the coil to reach the target position in a shorter time. For sounds of the same amplitude at different frequencies, faster movement is required for higher frequencies, and the higher acceleration will require more magnetic force.

We conducted experiments using the earphones to play the sounds of the same amplitude at different frequencies. As shown in Figure 23, the higher the frequency is, the greater the induced potential. After normalizing this curve, we refer to it as the frequency–magnetic curve $\eta(f)$, which represents how the coefficient of the magnetic field changes as a function of frequency.

Another change originates from the resonance effect. The damping is the smallest at the center frequency of the resonance and gradually increases as the signal offset from the center frequency increases, causing the induced voltage to decrease accordingly. Note that the damping will also change if we change the capacitance (Figure 16).

We first used different capacitors to record the response function $V_B(f,C)$ to the magnetic field signal, where C is the capacitance value corresponding to the curve. The normalized curve shows the induced voltage of the coil at different frequencies when the same magnetic field is used. To restore the audio to its original nature, we need to amplify it based on the frequency spectrum.

The restoration process is divided into the following steps. When we receive a signal, we first obtain the resonance frequency f_0 in accordance with the chosen capacitance and obtain the response function $V_B(f_0,C)$. We multiply the signal at all frequencies by $\frac{V_B(f_0)}{V_B(f)}$ to obtain the actual magnetic field from the received signal. In the next step, we use the frequency–magnetic curve $\eta(f)$ for scaling. We choose the resonance frequency f_0 as the reference for scaling and calculate the scaling factor $\frac{\eta(f_0)}{\eta(f)}$ at different frequency points. This approach is applied because we have found that the signal at f_0 is the least distorted, and the distortion will be aggravated if we use a different frequency point as the reference.

However, after multiplication by the scaling factor, the perceptual quality of the restored audio is still poor. The reason is that the audio response curve that humans are accustomed to is not flat. Therefore, we use the Harman curve as a reference for audio restoration. The Harman curve is the ideal curve used by earphone designers to measure the sound quality of earphones. We use the Harman curve coefficients to adjust the sound again based on the frequency spectrum. Our final frequency response curve after restoration is shown in Figure 24.

5.4 Angle Adjustment

To obtain the best audio quality, we should ensure that the coils of the MagEar system and the victim's earphones are on the same axis. However, in many cases, this is not easy to achieve. Unfortunately, an angular offset will reduce the received magnetic flux, which will reduce the quality of the obtained audio.

A dual-coil mode is used in MagEar to solve this problem, as shown in Figure 25. In Figure 25, 'default mode' refers to the case in which the adversary places only one coil in the headphone shell, whereas in dual-coil mode, the adversary hides one coil on each side of the headphone shell. This allows an attacker using MagEar to choose to rotate to face the victim from the front instead of from the side when the angle is not ideal. For example, when the angle between the victim's headphone and the receiver coil is 30 degrees, MagEar's performance decreases compared to the best-angle case. In the dual-coil mode, when the attacker is facing the victim, because the two coils are located on either side of the head, they are at the same angle with respect to the target headphone. These two coils are connected in series to enhance the received signal and thus compensate for the performance loss due to the angular offset. This method effectively increases the induced voltage of the received signal because the two coils form a mutual inductance, thereby enhancing the intensity of the induced voltage. In fact, the induced voltage is often higher than twice that of a single coil.

5.5 Speech Quality Metrics

To evaluate the performance of our algorithm, we use both objective (MFCCs) and subjective (MOSNet) measures to assess speech quality.

MFCCs: Mel-scale frequency cepstral coefficients (MFCCs) are among the most common and effective sound features for evaluation. They are calculated by using the short-time Fourier transform and a Mel filter bank to identify the energy distribution characteristics of audio in the frequency domain; specifically, 12 cepstral-domain values are used to measure these characteristics.

For our speech quality evaluation, we extract the MFCCs of both the played audio and the recovered speech as their features. Then, we calculate the cosine similarity of the two sets of coefficients. The

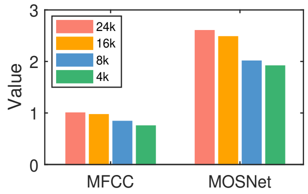


Figure 26: Audio quality at different downsampling rates.

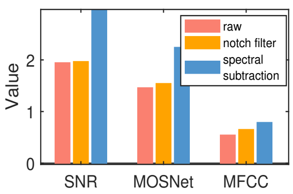


Figure 27: Improvement in audio quality achieved with denoising algorithms.

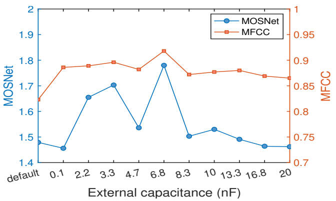


Figure 28: Audio quality when different capacitances are connected in parallel.

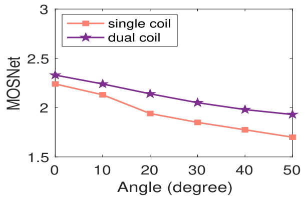


Figure 29: Performance comparison between a single coil and dual coils at different angles.

cosine similarity is the cosine value of the included angle between two vectors. The smaller the included angle between two vectors is, the closer their cosine value is to 1. Hence, the cosine similarity values are between -1 (opposite) and 1 (same). In later sections, for brevity, we will use 'MFCC' to represent the cosine similarity between the MFCCs of two audio samples.

MOSNet: The mean opinion score (MOS), obtained by asking listeners to use a 5-point scale to rate the tested audio quality, is widely used in listening tests. It is judged based on subjective factors such as the comfort and clarity of the sound. MOSNet is a recent advancement in automatic quality evaluation that was built on a convolutional neural network–bidirectional long short-term memory (CNN-BLSTM) model [22]. It considers human MOS judgments as the ground truth and predicts MOS ratings with a fairly close correlation to human subjective ratings. The calculation of MOSNet yields a score ranging from 1 (very poor) to 5 (very excellent). To verify the effectiveness of these two metrics, we resampled an audio recording with a 24 kHz sampling rate to a lower sampling rate and then calculated the MFCC and MOSNet results. As shown in Figure 26, with the decrease in the sampling rate, the values of MFCC and MOSNet decreased. The values for audio samples at different downsampled rates (4k, 8k, 16k and original–24k) were as follows: MFCC (0.74, 0.83, 0.95 and 1) and MOSNet (1.9, 2.0, 2.4 and 2.6).

Then, we used the MFCC and MOSNet scores to evaluate the performance of various algorithms in improving the recovered audio quality. Figure 28 shows the audio quality achieved when connecting a coil in parallel with different capacitors. We can observe that the audio quality does not have a linear relationship with capacitance. The audio quality reaches its maximum value when the external capacitance is 6.8 nF and the resonance frequency is approximately 3200 Hz. When the external capacitance is small, the coil will have a high resonance frequency that cannot amplify speech signals. In contrast, when the coil has a resonance frequency below 2 kHz, we also cannot obtain good audio quality. The reason is that the circuit noise has high energy at low frequencies, as shown in Figure 19. In these frequency bands, the noise can also be amplified, leading to poor audio quality. In addition, Figure 27 shows that the audio quality can be greatly improved by applying a notch filter and spectral subtraction.

6 IMPLEMENTATION

As our receiver, we designed a coil with a diameter of 4 cm and a height of 3.5 cm. We also added an external capacitor in parallel with the coil to tune the resonance frequency. To further amplify the received signal, the coil was connected to an AD620 amplifier, which has a gain of 2000×. Then, the coil was connected to a USB3202 data acquisition board with a 16-bit ADC to digitize the signal; this ADC supports a sampling rate of up to a 250k, covering the whole audible frequency band of human speech. The ADC was connected to a laptop through a USB port for received signal processing.

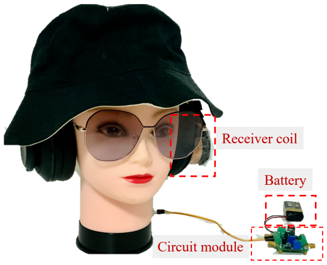


Figure 30: Diagram of the eavesdropping coil.

7 EVALUATION

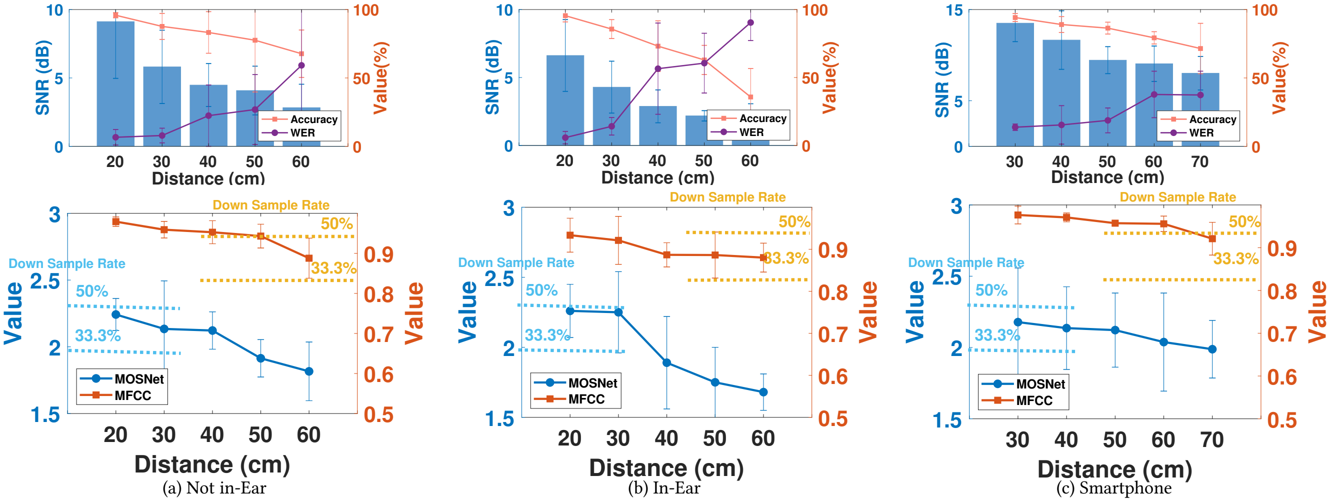
7.1 Experimental Setup

Data: We used audio from LibriTTS as the eavesdropped content in our evaluation. LibriTTS is a multispeaker English corpus consisting of approximately 585 hours of read English speech at a 24 kHz sampling rate [28].

Metrics: We used objective (MFCC) and subjective (MOSNet) measures to assess speech quality.

Automatic speech recognition: We also tested the application of voice recognition on the restored audio in further experiments. We used the Google speech-to-text (STT) API to transcribe the recovered magnetic sound and obtain a text string. To evaluate the similarity between the transcribed and ground-truth texts, we adopted the Levenshtein ratio and the word error rate (WER) as speech recognition metrics. The Levenshtein distance is the minimum number of edits (i.e., substitutions, deletions or insertions) required to transform one string into another, and the Levenshtein ratio is calculated using the equation $1 - \frac{Levenshteindistance_{a,b}}{max(len_a, len_b)}$. We refer to the similarity result as the automatic speech recognition (ASR) accuracy. The WER is a standard metric for speech recognition that is calculated as $WER = \frac{S+D+I}{N}$, where S, D, and I are the numbers of substitutions, deletions and insertions, respectively, and N is the number of words in the reference text.

Other: The audio loudness of the earphones was set to 80 dB (80% volume for earphones), as measured by a SMART SENSOR AR844 digital sound level meter. Note that this loudness value is only observed in the ear canal, and others outside cannot hear the played audio. The angle between MagEar and the eavesdropped speaker-embedded device was fixed. We tried to make the coil face



(a) Not in-Ear (b) In-Ear (c) Smartphone

Figure 31: Performance in phone call eavesdropping with different distances between the speaker and MagEar.

the speaker as much as possible to maximize the magnetic flux through the coil.

7.2 Eavesdropping on Different Speakers

Eavesdropping on the content of someone's phone calls is one of the most common eavesdropping scenarios. When the attacker cannot plant malware, MagEar can be used for eavesdropping by recovering sound signals based on the magnetic leakage of a phone's speaker at a social distance.

In this experiment, we evaluated MagEar for eavesdropping on both smartphones and earphones. The victim was asked to hold five different smartphones next to his ear for the first session and to wear ten different earphones (connected to a smartphone) for the second session. The distance between the coil and the eavesdropped device was varied from 20 cm to 60 cm for earphones and from 30 cm to 70 cm for smartphones. All the audio samples from the LibriTTS corpus were played once in each round, and we collected the magnetic signals for testing.

Figure 31 shows the recovered quality for speech played by non-in-ear headphones, earphones, and smartphones at different distances. The top three subgraphs display the results for the SNR, accuracy, and WER, while the bottom subgraphs illustrate the MFCC and MOSNet values. We can see that within a distance of 60 cm, MagEar can recover in-ear and non-in-ear headphone audio and achieve good preservation of the sound characteristics, and the SNR even drops sharply. Specifically, the average automatic speech recognition (ASR) accuracy for headphones is 98.73% at 20 cm and 74.31% at 60 cm. The recognition accuracy is 47.71% when the distance is 60 cm for in-ear earphones. Notably, in-ear headphones show a faster decline in the MOSNet score because they benefit from their closed structure and can produce sound with lower vibration. Their piezoelectric systems consume less energy, so their magnetic fields are smaller than those of non-in-ear headphones. For smartphones, the average ASR accuracy is 95.76% and the average WER is 13.99% at 30 cm. When the distance increases to 70 cm, the average accuracy decreases to 80.78% and the average WER increases to 37.54%. Since the recognition accuracy and WER are both used to evaluate the transcription results, we also display both objective (MFCC) and subjective (MOSNet) measures in the bottom subgraphs in Figure 31. To gain a more intuitive understanding of the performance capabilities of MagEar, we also resampled an original audio sample at 24 kHz to several lower sampling rates (i.e., 4 kHz, 8 kHz, 12 kHz, and 16 kHz) and then calculated the MFCC and MOSNet scores. The yellow dotted lines in Figure 31 represent the values when the original audio sampling rate is downsampled to 50% and 30% of the original, while the blue dotted lines represent the values of MOSNet. We can observe that all performance metrics of MagEar exceed those for the audio at the 30% downsampling rate and approach those for the 50% downsampling rate.

7.3 Physical Obstacles

Without loss of generality, we evaluated only one earphone model (i.e., Airpods) and one smartphone model (i.e., iPhone Xs) in the following experiments. Often, victims believe that physical obstacles (such as walls, soundproof glass, or doors) can prevent eavesdropping. In this experiment, we demonstrated MagEar's eavesdropping ability when there is an obstacle between MagEar and the victim device. We experimented with concrete walls of three thicknesses (23, 33, and 42 cm), where the eavesdropping distance was the same as the wall thickness. We also used a 2.5 cm thick wooden board and a 1.0 cm thick sheet of tempered glass as obstacles, both with an eavesdropping distance of 40 cm. The results for the earphones and the smartphone are presented in Figure 32 and Figure 33, respectively. The bar and line graphs represent the MOSNet and MFCC values, respectively. The w/o wall and w/all in 32 and Figure 33 refer to the case of without wall and with wall between the target device and the receiver coil, respectively. It can be seen that obstacles have a negligible effect on magnetic side-channel eavesdropping. The reason is that the conductivity of these obstacles is low, and magnetic signals pass through them without significant loss.

7.4 Eavesdropping in Different Environments

Next, we evaluated MagEar in different environments beyond the laboratory in which the previous experiments were performed. The eavesdropping distance was set to 40 cm for the Airpods and the iPhone Xs. The test sites we selected included both indoor scenes and open areas: a park, a coffee shop, a restaurant, and a private home. The essential differences among the five testing

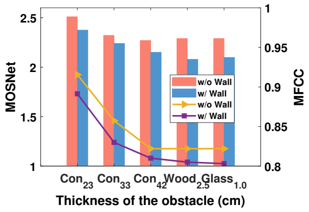


Figure 32: Results of eavesdropping on Airpods through various obstacles.

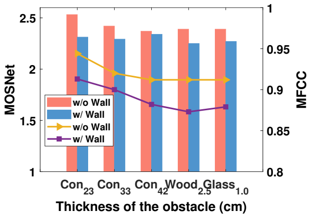


Figure 33: Results of eavesdropping on an iPhone through various obstacles.

	Airpods		Iphone	
Scenarios	MFCC	MOSNet	MFCC	MOSNet
Laboratory (49 dB)	0.85	1.86	0.88	1.92
Home (44 dB)	0.867	1.993	0.907	2.109
Park (52 dB)	0.89	1.928	0.95	2.2
Restaurant (66 dB)	0.82	1.52	0.84	1.78
Coffee Shop (60 dB)	0.83	1.6	0.87	1.98

Table 1: Eavesdropping performance in different environments for Airpods and the iPhone Xs (distance=40 cm)

environments lie in the ambient magnetic field strengths. For example, in a coffee shop, many customers may be using laptops, tablet PCs, or smartphones for work or entertainment. The light bulbs and power lines in a coffee shop will also produce electromagnetic radiation. The electromagnetic interference emitted from such electronic devices may affect the performance of our system. In contrast, outdoor environments such as parks are generally relatively empty, and fewer electronic devices are used, resulting in a comparatively weak magnetic field strength. As shown in Table 1, all of the selected environments have varying degrees of magnetic field interference and noise interference. We can see that the overall MFCC and MOSNet performance suffers a certain degree of degradation in indoor public scenarios. In contrast, in the park scenario, the magnetic field interference is lower than that in the laboratory, resulting in higher MFCC and MOSNet values.

8 RELATED WORK

In recent years, many researchers have proposed various side-channel attack schemes to successfully eavesdrop on human speech based on nonacoustic devices. Gyrophone [23] uses motion sensors for eavesdropping; specifically, it utilizes a smartphone's gyroscope to capture the acoustic vibrations from the speaker. The vibration readings reflect the speech information and are used to recognize digit pronunciations. Based on a similar principle, several studies have validated the feasibility of using accelerometer data for hot word detection [29] and isolated word recognition [6, 7]. S Abhishek Anand et al. [5] investigated the relationship between speech and motion sensor data. They concluded that motion sensors can only capture the variations caused by surface vibrations or conductive vibrations, whereas aerial vibrations are not sufficiently powerful to affect motion sensor readings. VibraPhone [25] exploits the reverse electromotive force of the vibro-motor in a smartphone to convert the vibro-motor into an acoustic microphone. In [20], Kwong et al. considered using a magnetic hard drive as a microphone; they demonstrated the ability to restore human speech by measuring the offset of the read/write head from the center of the track. A visual microphone [4] extracts the microvibrations caused by airborne sound pressure from an object (e.g., an empty bag of snacks) using a 2200 FPS high-speed camera to recover human speech near the object. Lamphone [24] utilizes a telescope to focus on a hanging bulb in the victim's room and a photodiode to record the bulb's vibrations caused by the air pressure of sound waves. The victim's speech is then recovered based on the fluctuations in the photodiode readings. The authors of ART [26] successfully demodulated a loudspeaker's vibration signal from the RSS readings of received Wi-Fi packets. They could recover the sound of a piano and recognize ten digits by analyzing the RSS.

On the one hand, researchers have proposed many interesting sensing applications by leveraging the properties of magnetic and electromagnetic fields. Maghacker [21] is a sensing system that can infer handwriting contents by eavesdropping on the magnetic field emitted from the permanent magnet in a stylus pen. [14] demonstrated the ability to snoop on the content of a tablet display by capturing the electromagnetic emanations from the display interface. Choi et al. [10, 11] reported a security vulnerability in Samsung Pay. They designed magnetic coils to monitor the magnetic fields generated by magnetic secure transmission devices and could thus obtain one-time payment tokens decoded from the magnetic signals. NFC+ [30] implements a long-range magnetic field reader with multiple transmission coils based on the resonance effect for a near-field communication network. In [12], the authors proposed a novel scheme for eavesdropping on visible light communications through a wall.

9 DISCUSSION AND DEFENSE

As a proof-of-concept eavesdropping system exploiting magnetic side-channel information, MagEar still has some limitations to be addressed in the future. The eavesdropping distance is the main drawback of MagEar. MagEar can currently eavesdrop on earphones from no more than approximately 60 cm away. The source of the magnetic field signal is beyond our control. Therefore, if we wish to extend the eavesdropping distance, two main methods remain. The first method is to use high-end magnetic sensors in place of the customized coil. For example, fluxgate sensors and optically pumped magnetometers can measure magnetic fields on the order of pT [15]. The second method is to focus on more advanced noise reduction algorithms or low-noise circuits.

Regarding defensive countermeasures, one potential defense is to use magnetic shielding material inside the headphone shell, which may dampen magnetic leakage from headphones. Shielding for low-frequency magnetic fields usually consists of high-permeability metals, such as iron, silicon steel, or permalloy. High-permeability materials have a low magnetic resistance, so the magnetic induction lines are strongly concentrated in the shielding material and cannot pass through a cavity surrounded by the shielding material. To investigate the effectiveness of shielding, we placed a permalloy plate with a thickness of 0.8 mm between the target earphones and the receiver coil of MagEar, which were separated by a distance of 30 cm. After the introduction of the permalloy plate, the ASR accuracy dropped from 98.37% to 71.95%, the SNR decreased from 4.16 to 1.70, and the MOSNet score decreased from 2.19 to 1.88. These results show that high-permeability metals can effectively impede the propagation of the magnetic field and have a great impact on the system performance. Therefore, we suggest that headphone manufacturers use magnetic shielding materials inside headphone shells to prevent privacy risks due to magnetic leakage.

10 CONCLUSION

In this paper, we propose MagEar, an eavesdropping system that utilizes magnetic signals leaked by a speaker to recover intelligible human speech. The key observation is that the diaphragm of the speaker is driven by a varying magnetic force; therefore, the radiated magnetic field can be measured by a magnetic sensor to infer what the speaker is playing. We establish a sound-to-magnetic model and validate the feasibility of our proposal for different types of speaker-embedded electric devices. Our evaluation shows that such magnetic side-channel attacks pose a risk of privacy leakage to the majority of speaker-embedded devices. Even when the victim's speaker is blocked by soundproof walls, MagEar can achieve a high speech recognition accuracy.

Our exploration of magnetic side-channel attacks reveals that the majority of COTS speaker-embedded devices (i.e., headphones, smartphones and smart speakers) have a significant loophole that presents a risk of invisibly leaking a victim's private information. We hope that our work can push manufacturers to rethink this security vulnerability of speakers.

ACKNOWLEDGMENTS

This research was supported in part by the China NSFC Grant (U2001207, 61872248), Guangdong NSF 2017A030312008, Shenzhen Science and Technology Foundation (No. ZDSYS20190902092853047, R2020A045), the Project of DEGP (No.2019KCXTD005, 2021ZDZX1068), the Guangdong "Pearl River Talent Recruitment Program" under Grant 2019ZT08X603. Kaishun Wu is the corresponding author.

REFERENCES

[1] 2017. Global Micro Speaker Market Driven by the Rise in Adoption of E-commerce Platforms: Technavio. https://www.businesswire.com/news/home/20170224005162/en/Global-Micro-Speaker-Market-Driven-by-the-Rise-in-Adoption-of-E-commerce-Platforms-Technavio. Last accessed March 25, 2021.

[2] 2017. The market for MEMS microphones and ECMs, micro-speakers and audio ICs will be worth 20B in 2022. https://www.i-micronews.com/products/acoustic-mems-and-audio-solutions-2017/. Last accessed March 25, 2021.

[3] 2019. Apple to Limit Accelerometer and Gyroscope Access in Safari on iOS 12.2 for Privacy Reasons. https://www.macrumors.com/2019/02/04/ios-12-2-safari-motion-orientation-access-toggle/. Last accessed March 25, 2021.

[4] Abe, Davis, Michael, Rubinstein, Neal, Wadhwa, Gautham, J., Mysore, Fredo, Durand, William, T., and Freeman. 2014. The visual microphone: passive recovery of sound from video. *Acm Transactions on Graphics Proceedings of Acm Siggraph* (2014).

[5] S Abhishek Anand and Nitesh Saxena. 2018. Speechless: Analyzing the threat to speech privacy from smartphone motion sensors. In *2018 IEEE Symposium on Security and Privacy (SP)*. IEEE, 1000–1017.

[6] S Abhishek Anand, Chen Wang, Jian Liu, Nitesh Saxena, and Yingying Chen. 2020. Motion Sensor-based Privacy Attack on Smartphones. arXiv:1907.05972 [cs.CR]

[7] Zhongjie Ba, Tianhang Zheng, Xinyu Zhang, Zhan Qin, Baochun Li, Xue Liu, and Kui Ren. 2020. Learning-based practical smartphone eavesdropping with built-in accelerometer. In *Proceedings of the Network and Distributed Systems Security (NDSS) Symposium*. 23–26.

[8] Michael Berouti, Richard Schwartz, and John Makhoul. 1979. Enhancement of speech corrupted by acoustic noise. In *ICASSP'79. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 4. IEEE, 208–211.

[9] John Borwick. 2012. *Loudspeaker and headphone handbook*. CRC Press.

[10] Daeseon Choi and Younho Lee. 2016. Eavesdropping One-Time Tokens over Magnetic Secure Transmission in Samsung Pay. In *Proceedings of the 10th USENIX Conference on Offensive Technologies* (Austin, TX) *(WOOT'16)*. USENIX Association, USA, 52–58.

[11] Daeseon Choi and Younho Lee. 2018. Eavesdropping of Magnetic Secure Transmission Signals and Its Security Implications for a Mobile Payment Protocol. *IEEE Access* 6 (07 2018), 1–1. https://doi.org/10.1109/ACCESS.2018.2859447

[12] Minhao Cui, Yuda Feng, Qing Wang, and Jie Xiong. 2020. *Sniffing Visible Light Communication through Walls*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3372224.3419187

[13] Jun Han, Albert Jin Chung, and Patrick Tague. 2017. Pitchln: Eavesdropping via Intelligible Speech Reconstruction Using Non-Acoustic Sensor Fusion. In *Proceedings of the 16th ACM/IEEE International Conference on Information Processing in Sensor Networks* (Pittsburgh, Pennsylvania) *(IPSN '17)*. Association for Computing Machinery, New York, NY, USA, 181–192. https://doi.org/10.1145/3055031.3055088

[14] Yuichi Hayashi, Naofumi Homma, Mamoru Miura, Takafumi Aoki, and Hideaki Sone. 2014. A Threat for Tablet PCs in Public Space: Remote Visualization of Screen Images Using EM Emanation. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security* (Scottsdale, Arizona, USA) *(CCS '14)*. Association for Computing Machinery, New York, NY, USA, 954–965. https://doi.org/10.1145/2660267.2660292

[15] Maurice Hott, Peter A Hoeher, and Sebastian F Reinecke. 2019. Magnetic communication using high-sensitivity magnetic field detectors. *Sensors* 19, 15 (2019), 3415.

[16] Yongzhi Huang, Kaixin Chen, Yandao Huang, Lu Wang, and Kaishun Wu. 2021. A Portable and Convenient System for Unknown Liquid Identification with Smartphone Vibration. *IEEE Transactions on Mobile Computing* (2021).

[17] Yongzhi Huang, Kaixin Chen, Yandao Huang, Lu Wang, and Kaishun Wu. 2021. Vi-liquid: unknown liquid identification with your smartphone vibration.. In *MobiCom*. 174–187.

[18] Nathan Ida. 2015. *Engineering electromagnetics*. Springer.

[19] Valeriy Korepanov and Vira Pronenko. 2010. Induction magnetometers-design peculiarities. *Sensors & Transducers* 120, 9 (2010), 92.

[20] Andrew Kwong, Wenyuan Xu, and Kevin Fu. 2019. Hard Drive of Hearing: Disks that Eavesdrop with a Synthesized Microphone. In *2019 IEEE Symposium on Security and Privacy (SP)*.

[21] Yihao Liu, Kai Huang, Xingzhe Song, Boyuan Yang, and Wei Gao. 2020. MagHacker: eavesdropping on stylus pen writing via magnetic sensing from commodity mobile devices. In *Proceedings of the 18th International Conference on Mobile Systems, Applications, and Services*. 148–160.

[22] Chen-Chou Lo, Szu-Wei Fu, Wen-Chin Huang, Xin Wang, Junichi Yamagishi, Yu Tsao, and Hsin-Min Wang. 2019. MOSNet: Deep learning based objective assessment for voice conversion. *arXiv preprint arXiv:1904.08352* (2019).

[23] Yan Michalevsky, Dan Boneh, and Gabi Nakibly. 2014. Gyrophone: Recognizing Speech From Gyroscope Signals. In *23rd USENIX Security Symposium*.

[24] Ben Nassi, Yaron Pirutin, Adi Shamir, Yuval Elovici, and Boris Zadov. 2020. Lamphone: Real-Time Passive Sound Recovery from Light Bulb Vibrations. *BlackHat USA* (2020).

[25] Nirupam Roy and Romit Roy Choudhury. 2016. Listening through a Vibration Motor. In *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services* (Singapore, Singapore) *(MobiSys '16)*. Association for Computing Machinery, New York, NY, USA, 57–69. https://doi.org/10.1145/2906388.2906415

[26] Teng Wei, Shu Wang, Anfu Zhou, and Xinyu Zhang. 2015. Acoustic Eavesdropping through Wireless Vibrometry. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking* (Paris, France) *(MobiCom '15)*. Association for Computing Machinery, New York, NY, USA, 130–141. https://doi.org/10.1145/2789168.2790119

[27] Jerry Whitaker and Blair Benson. 2001. *Standard handbook of audio engineering*. McGraw-Hill Education.

[28] Heiga Zen, Viet Dang, Rob Clark, Yu Zhang, Ron J Weiss, Ye Jia, Zhifeng Chen, and Yonghui Wu. 2019. LibriTTS: A corpus derived from LibriSpeech for text-to-speech. *arXiv preprint arXiv:1904.02882* (2019).

[29] Li Zhang, Parth H. Pathak, Muchen Wu, Yixin Zhao, and Prasant Mohapatra. 2015. AccelWord: Energy Efficient Hotword Detection through Accelerometer. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services* (Florence, Italy) *(MobiSys '15)*. Association for Computing Machinery, New York, NY, USA, 301–315. https://doi.org/10.1145/2742647.2742658

[30] Renjie Zhao, Purui Wang, Yunfei Ma, Pengyu Zhang, Hongqiang Harry Liu, Xianshang Lin, Xinyu Zhang, Chenren Xu, and Ming Zhang. 2020. NFC+: Breaking NFC Networking Limits through Resonance Engineering. In *Proceedings of the Annual Conference of the ACM Special Interest Group on Data Communication on the Applications, Technologies, Architectures, and Protocols for Computer Communication* (Virtual Event, USA) *(SIGCOMM '20)*. Association for Computing Machinery, New York, NY, USA, 694–707. https://doi.org/10.1145/3387514.3406219