An Eavesdropping System Based on Magnetic Side-Channel Signals Leaked by Speakers

QIANRU LIAO, Hong Kong University of Science and Technology(GZ), China
YONGZHI HUANG, Hong Kong University of Science and Technology(GZ), China
YANDAO HUANG, Hong Kong University of Science and Technology(GZ), China
KAISHUN WU, Hong Kong University of Science and Technology(GZ), China

The use of speakers in electronic devices has become widespread, but the security risks associated with micro-speakers, such as earphones, are often overlooked. Many assume that soundproof barriers can prevent sound leakage and protect privacy. This paper presents the prototype MagEar, an eavesdropping system that exploits magnetic side-channel signals leaked by a micro-speaker to restore intelligible human speech. MagEar outperforms some high-precision magnetometers in detecting magnetic fields at the nanotesla level. Even at a distance of 60 cm, it can recover high-quality audio with a 90% similarity to the original audio. Moreover, the MagEar prototype is portable and can be concealed within a headset housing. We have implemented MagEar as a proof-of-concept system and conducted multiple case studies on the eavesdropping of various speaker-embedded devices, including earphones. The recovered speech can be transcribed using automatic speech recognition techniques, even when obstructed by soundproof walls. It is our aspiration that our work can prompt manufacturers to reconsider the security vulnerabilities of speakers.

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

Additional Key Words and Phrases: Eavesdropping, Side Channel attack, Magnetic field, Mobile security, Privacy disclosure

1 INTRODUCTION

Motivation: Recent years have witnessed the expansion of advanced information and communication technologies. We are surrounded by various technological devices offering complete services and enriching 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 Technavio [1] forecasted that the micro-speaker market would reach 15.16 billion units by 2021. 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 countermeasures, new eavesdropping techniques are constantly emerging, infringing on users' privacy and security.

State-of-the-art eavesdropping systems have demonstrated the feasibility of inferring sensitive user information using various non-acoustic sensors in smartphones, such as motion sensors [5, 7, 14, 27, 39] and vibration motors [30]. However, such methods assume that the adversary has gained access to sensor readings by planting malware.

Authors' addresses: Qianru Liao, Hong Kong University of Science and Technology(GZ), Guangzhou, China, qliao551@connect.hkust-gz.edu.cn; Yongzhi Huang, Hong Kong University of Science and Technology(GZ), Guangzhou, China, yhuang849@connect.hkust-gz.edu.cn; Yandao Huang, Hong Kong University of Science and Technology(GZ), Guangzhou, China, yhuangfg@connect.ust.hk; Kaishun Wu, Hong Kong University of Science and Technology(GZ), Guangzhou, China, wuks@ust.hk.

ACM 1550-4859/2023/12-ART
https://doi.org/10.1145/3637063

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

By deploying new policies [3], the equipment provider or another relevant organization can effectively impose restrictions on such hacking. In addition, previous works have concentrated mostly on how to convert a device near a victim into a microphone capable of recording human speech. Only a few works have noted that the speaker can "talk" as well. In ART [36], the authors proposed the concept of wireless vibrometry to eavesdrop on loudspeakers. They successfully demodulated the vibration signal of a loudspeaker from the received signal strength (RSS) readings of received Wi-Fi packets, thereby recovering the sound of a piano and ten distinct digits. However, for small devices, such as headphones, 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 voice from speakers, including the micro-speakers in earphones.

Our approach: In this paper, 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 consumer electronic devices with speakers pose a concern of leaking user private 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, there is an 'intermediate product,' i.e., the changing magnetic field, which a simple coil can capture. An adversary could deduce the original digital audio input by analyzing the variations in the magnetic field. We refer to such a process as a magnetic side-channel attack.

Challengs and Solutions: However, we need to overcome several **challenges** before transforming the above-mentioned high-level concept into a practical eavesdropping system.

1) The magnitude of the magnetic field generated by earphones is relatively low, measuring in at the nanotesla (10^{-9} T) level. Additionally, the inherent attenuation characteristics of the magnetic signal restrict its leakage range. The intensity of the magnetic field is inversely proportional to the third power of the distance, i.e., $O(1/d^3)$. Given the feeble nature of the magnetic field, even a high-precision commercial magnetometer can only detect magnetic fluctuations within a range of 5 cm. Consequently, designing a low-cost device capable of eavesdropping beyond the intimate distance of 35 cm poses a significant challenge.

2) In order to measure the faint magnetic field, enhancing the system's sensitivity is necessary. However, a highly sensitive system will inevitably be more vulnerable to noise interference. Our system's noise originates from two sources. Firstly, circuit noise possesses a significant amount of energy at low frequencies that overlap with speech, making it difficult to separate the effective voice signal from this noise using a simple filter. Secondly, we are constantly surrounded by a multitude of electronic devices that emit electromagnetic radiation. The

received signal is prone to contamination or even being overwhelmed by ambient noise. Hence, the second challenge is to eliminate these two types of noise and obtain a clear audio signal.

3) As a magnetic coil is not intended for recording sound in the same way as a microphone, it is unable to provide an adequate listening experience. The audio derived from magnetic signals will be of inferior quality and distorted. It is therefore important to investigate the distinctions between magnetic and acoustic signals and develop a viable recovery technique, which presents another challenge.

Summary of Experiment Results: We evaluated a total of 15 commercially available electronic devices, comprising 10 earphones and 5 smartphones. These devices were utilized to play audio from an open-source corpus, and we employed MagEar to eavesdrop on them. At a distance of 60 cm, the average MOSNet score and cosine similarity for the smartphones were 2.09 and 0.92, respectively. For the earphones, the corresponding values were 1.83 and 0.89, respectively, at a distance of 50 cm.

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

1) We propose a new side-channel attack scheme for eavesdropping on speaker-embedded devices like earphones. The results reveal the alarming risk of user privacy leakage posed by the majority of COTS speaker-embedded devices.

2) Our designed MagEar system is sensitive enough to measure magnetic fields on the nano tesla scale, surpassing the sensitivity of some high-precision magnetometers.

3) Based on the physical model, we redesign a coil so that its performance is equivalent to a coil with three times the diameter. Additionally, we eliminate noise interference and improve the quality of distorted audio. At a distance of 60cm, the audio quality of the recovered speech is equivalent to that of the original audio sampled at a rate of one-third.

4) We propose a dual-coil mode to compensate for the performance loss due to the angular offset between MagEar and the victim's headphones. Even when the angle between MagEar and the target's device is 90°, we can boost the amplitude of the received signal by three times by rotating two receiver coils.

5) We have conducted a series of experiments under diverse circumstances to investigate the feasibility of MagEar. Our findings indicate that the eavesdropped signal can be recovered as intelligible speech and transcribed with ease through the use of a commercial off-the-shelf automatic speech recognition service.

The remaining sections are organized as follows: Section 2 explains the eavesdropping scenarios and how MagEar works in practice. In Section 3, we show the speaker's working principle as well as the mathematical model describing the relationship between input audio and leaked magnetic signals. Then, we introduce the MagEar design and explain specifically how we determine the various parameters of a receiver coil to measure the weak magnetic field of a speaker in Section 4. Section 5 discusses several methods to enhance human speech features and mitigate noise's effects. In Section 6, we introduce our designed receiver coil and circuit module. In Section 7, we evaluate our system performance under different experimental settings. We discuss related work on speech eavesdropping and magnetic sensing in Section 8. We give the potential countermeasure for magnetic eavesdropping in Section 9 and discuss the limitations of our system in Section 10. Finally, we give the conclusion in Section 11.

2 THREAT MODEL

We make the assumption that the victim is wearing either headphones or earphones, while the attacker employs a receiver coil. By measuring the leaking magnetic field in the vicinity of the victim's headphones and applying signal processing algorithms, the adversary is able to extract the audio content of the victim. The victim may be utilizing headphones to participate in a virtual meeting, conduct a phone conversation, or listen to media. The reconstructed audio can be played directly, identified by a human listener, or transcribed using an existing

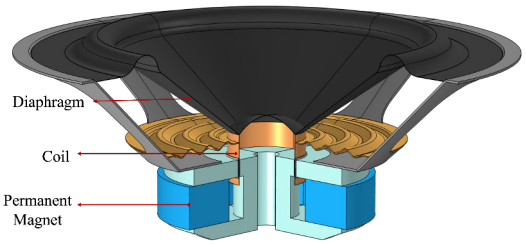


Fig. 2. Simulation model of a dynamic speaker

speech-to-text API. Consequently, the adversary can deduce the target's private information, business details, personal interests and hobbies, and personal information from the recovered audio.

We suppose that both the victim and attacker are in close proximity to each other. For instance, during peak hours on public transportation such as subways or buses, the carriages tend to be crowded, and passengers inevitably come into physical contact with each other. Even for passengers seated, their shoulders are often in contact with those of adjacent passengers. As illustrated in Figure 1, an attacker can conceal a receiver coil inside a headphone shell and eavesdrop on the audio content of adjacent passengers' headphones or earphones. Moreover, in scenarios where an attacker and a victim are seated on the same bench in a waiting room, airport lounge, or park, the audio content played by the victim's earphones is susceptible to leakage.

On the other hand, a receiver coil can serve as an eavesdropping device without requiring human intervention. A malicious coil can be installed on any object that individuals may stand or sit next to. For instance, some passengers wear headphones for extended periods on trains or buses and infrequently check their seats, an adversary may embed coils in railway and bus seats to capture nearby magnetic sounds. Similarly, employees typically use an unoccupied conference room for personal calls or virtual meetings, making a conference room table or chair an ideal location for coils that can eavesdrop on private or business discussions. In addition, individuals may have phone conversations at coffee shops, restaurants, or park benches, potentially exposing themselves to MagEar.

3 BACKGROUND AND PRINCIPLE

The speaker is well known as the core module of earphones, receiving the device's audio signal and producing sound. Although there are different speaker structure designs, dynamic unit speakers are widely used in non-professional headphones. In this section, we will explore the working principle of a speaker and subsequently analyze the relationship between the input audio and the magnetic field within the speaker.

3.1 Working Principle of Speakers

The fundamental component of a speaker is the transducer, which transforms electrical energy into mechanical energy [37]. A dynamic transducer is composed of three essential components: 1) a magnet system, which comprises a permanent magnet and T iron; 2) a vibration system, consisting of a voice coil and a diaphragm; and 3) components that offer structural stability. We designed a simulation model in COMSOL based on the speaker structure [9], as depicted in Figure 2.

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

Due to the constraining effect of the diaphragm's elastic force on the voice coil's motion, it exhibits simple harmonic motion. Nonetheless, the movement of the voice coil alters the magnetic flux of the static magnetic field

that passes through the voice coil. According to Lenz's law, the voice coil impedance hinders the current generated by the magnetic flux fluctuations, thereby introducing resistance to the harmonic motion. As a consequence, the simple harmonic motion is converted into damped motion, which can be characterized by the motion equation provided below [37]:

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

where m represents the total mass of the diaphragm and coil, x denotes the displacement of the diaphragm, and F_{mag} corresponds to the magnetic force. The damping coefficient c and spring constant k depend on the material of the diaphragm. Upon solving this differential equation, we can derive an approximate equation for the displacement [18–20]:

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

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

With respect to the magnetic force, it should be noted that the direction of coil vibration is perpendicular to the magnetic field. Thus, the magnetic force acting on the coil adheres to Fleming's right-hand rule. The magnetic force can be mathematically represented as follows:

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

where B is the total magnetic flux density, I is the audio current that flows through the coil, and l is the length of the coil. Based on Equations (2) and (3), in order to generate sound, the diaphragm must be propelled by sufficient magnetic field changes, which may pose a risk of privacy infringement. However, the momentum of the voice coil and diaphragm necessary for sound transmission may be tiny, resulting in a small magnetic leakage from the earphones. Therefore, is it necessary to be concerned about the risk of magnetic leakage?

3.2 Magnetic Leakage

To answer the question posed in the previous section, it is necessary to quantify the strength of the magnetic field that is emitted by a headset. For this purpose, we generated a 1 kHz sine wave in Python and played it through AirPods. Then, we measured the magnetic field near the AirPods using a TES-1394S electromagnetic field tester with a resolution of 0.001 μT. The angle between the magnetometer and AirPods was 0°. We took measurements at a range of distances away from the AirPods.

As shown in Figure 3, the high-precision magnetometer that we employed was capable of detecting magnetic leakage from a distance of no more than 5 cm from the earphones. Upon fitting the curve obtained by the magnetometer (represented by the dotted line), we can confirm that the earphones' magnetic leakage is minimal.

In accordance with 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 directly proportional to the rate of change of the magnetic flux density[22]:

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

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.

In theory, if we were able to infinitely increase the coil area A and the number of coils N, it would be simple to measure the magnetic signal emitted by an earphone speaker. However, eavesdroppers cannot realistically employ this strategy. 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:

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

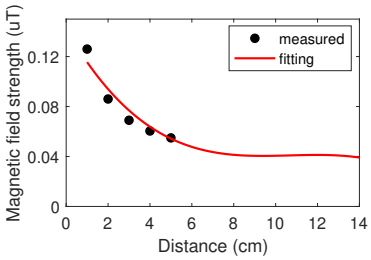


Fig. 3. Electromagnetic leakage of earphones measured at various distances.

This formulation explains why detecting magnetic leakage from other electronic devices is more accessible than from headphones. In general, the operating frequency of commercial electronic equipment surpasses one hundred megahertz (10^8 Hz) or one gigahertz (10^9 Hz) and may exceed 10^{11} Hz in the case of an optical communication device. In contrast, most sounds played through earphones fall within the frequency range of 200 to 2000 Hz, which presents an enormous challenge for detection. It is worth noting that a coil is preferred over an antenna for practical reasons. While an appropriate antenna could potentially yield a greater gain, the size of the antenna required for a reasonable receiving range depends on the signal's wavelength, and even the most miniature monopole antenna needs to be at least 1/4 of the wavelength in size. Therefore, the minimum antenna size necessary for detecting earphone magnetic leakage would be 100 kilometers.

Thus far, eavesdropping on magnetic leakage has appeared unfeasible due to the aforementioned limitations. However, it is indeed possible to obtain and even retrieve the sounds played through earphones from their magnetic leakage. The subsequent section will provide a step-by-step explanation of how we overcome these challenges.

4 MAGEAR DESIGN

4.1 Experimental Setup

In this section, we will detail the process of designing a receiver coil for the purpose of measuring the weak magnetic field leaked by earphones. First, we conduct a theoretical analysis of how the parameters of a coil impact its performance in terms of the passing magnetic flux, induced voltage, sensitivity, signal-to-noise ratio, and resonance effect. Furthermore, we conducted corresponding experiments to validate these theoretical findings. Specifically, we generated 1 kHz sine wave audio using Python and played it through the Apple AirPods (1st generation). At the receiver end, a copper coil was connected to a USB3202 data acquisition board equipped with a 16-bit analog-to-digital converter (ADC) to collect the leaked magnetic field. In order to maximize the magnetic flux flowing that flows through the coil, the angle between the AirPods and the coil was fixed at 0°.

In addition, the receiver coil and headphones are positioned on the table, and the measured distance refers to the horizontal distance between their centers. Subsequently, we recorded the amplitude of the measured signal at 1kHz. As an amplifier will introduce considerable circuit noise at low frequencies, we opted not to connect the coil to the amplifier in this section to enable a more comprehensive investigation of the coil's properties. All experiments were performed in the laboratory.

4.2 Geometric Shape

The magnetic field configuration of an earphone is complex owing to the presence of both a static magnetic field and a dynamic magnetic field generated by a coil. Traditional research is based on ideal magnetic field conditions. Therefore, to optimally measure the magnetic field, we first need to determine the optimal geometric shape of

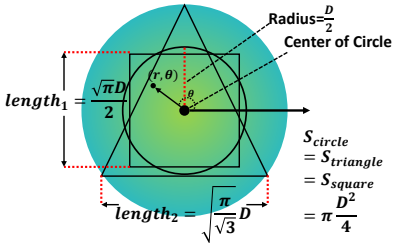


Fig. 4. Magnetic fluxes for receivers of different shapes.

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), if the receiver area and the number of turns are identical, the induced potential will also be equivalent in theory. However, this is not the case in reality because the magnetic flux density across the receiving plane differs. To explain this phenomenon, we need to establish a physical model.

Although the speakers we used in our experiments have different voice coil positions, their coil thickness is only 1 mm, which is significantly less than the experimental distance (at the centimeter level). Since the magnetic field of an N-turn coil is equivalent to N times that of a single coil, each single-turn coil can be viewed as a magnetic dipole. A magnetic dipole is a small closed loop that carries an electric current and is defined as $m = IS$, where m is the magnetic dipole moment, I is the current flowing through the loop, and S is the loop's area. The external magnetic field produced by a magnetic dipole can be expressed as follows [21]:

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

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 depicted in Figure 4, we have designed three receivers, each having an area of $A = \frac{\pi D^2}{4}$ but with distinct shapes: a circle, a square, and an equilateral triangle. We can calculate that the circle's radius is $\frac{D}{2}$ and that the side lengths of the square and triangle are $\frac{\sqrt{\pi}D}{2}$ and $\sqrt{\frac{\pi}{\sqrt{3}}}D$, respectively. We locate the three shapes at the same center and calculate their respective magnetic fluxes. According to formula (7), the magnetic field intensity is greatest near the center. Using the assumption that the magnetic field at the center point is $\frac{1}{R^3}$, we can obtain the following magnetic flux: $\phi \propto \frac{\mu_0 m}{4\pi}\iint \frac{1}{(R_{distance}+r\sin\arctan\frac{r}{R_{distance}})^3} dr d\theta$. 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. All the coils had the same surface area of 7.065 cm^2 and a height of 3 cm. The radius of the circle was 1.5 cm, while the side lengths of the square and triangle were 2.65 cm and 4.04 cm, respectively. The rectangle had dimensions of 3.5 cm in length and 2 cm in width. Then, we recorded the amplitudes at 1 kHz at different distances for each coil. The results, as shown in Figure 5, indicate that the circular coil produces the highest induced voltage.

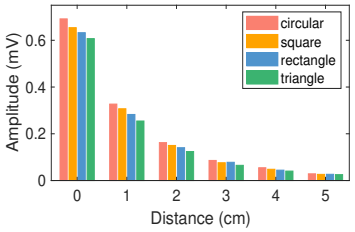


Fig. 5. Induced potentials of differently shaped coils. The results for the circular coil are the best.

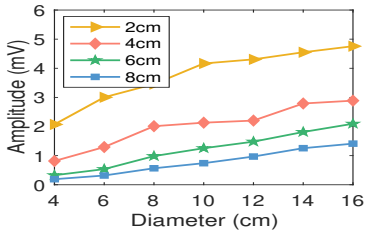


Fig. 6. Coil amplitude vs. diameter at different distances.

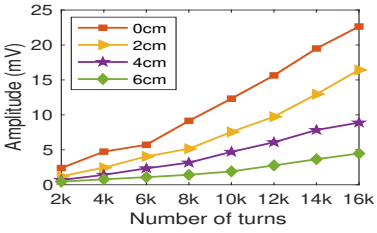


Fig. 7. Coil amplitude vs. number of turns at different distances.

4.3 Geometric Size

In terms of total passing magnetic flux, a circle is the most favorable coil shape based on the theoretical calculations and actual measurements reported in the previous section. In this section, we discuss how to determine the diameter of a coil.

We designed 8 coils with the same number of turns (N=5000) but varying diameters ($D \in \{4, 6, ..., 16\}$). Subsequently, we played a 1 kHz sine wave through AirPods and recorded the amplitudes at 1 kHz for each of the different coils. Figure 6 illustrates the relationships between the diameter and the induced voltage at various distances, which indicates that the amplitude increases as the diameter grows.

The experiment results shown in Figure 6 defy the expectation that the induced voltage should rise proportionally to the square of the diameter D, as indicated by Equation (6). Instead, the voltage growth rate gradually decreases. This is not because Equation (6) is incorrect. To explain this phenomenon, we can 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 $dB \propto (\frac{1}{dD})^3$. Since the diameter must be a non-negative value, it follows that $(D + dD)^2 > D^2 + dD^2$. As a result, it is possible to approximate the increased magnetic flux as $B + dB = B + ((\frac{1}{dD})^3)^2$.

Although a larger diameter will result in a higher induced potential, an eavesdropper may be constrained by the weight of their eavesdropping device. We hypothesize that eavesdroppers will attempt to limit the size of their eavesdropping devices to avoid detection. As the diameter increases, the incremental rise in magnetic flux gradually declines. Therefore, an eavesdropper is likely to select a coil with a high diameter performance, which implies that, for the selected diameter, increasing the coil's unit mass (e.g., by increasing the number of turns) will result in more rapid 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 steady effect with distance than smaller coils. However, given that commercially available earphones do not exceed 8 cm in diameter, we have opted for a diameter of 4 cm, which yields the highest average performance.

We also investigated how the number of turns affects the induced potential. To this end, we designed another 8 coils with varying numbers of turns ($N \in \{2000, 4000, 6000, ..., 16000\}$) and the same diameter (D=8 cm). The experiment was consistent with that of the diameter experiment. Figure 7 reveals that the relationship between amplitude and number of turns is nearly linear.

4.4 Fill or Vacant the Coil's Center

In this section, we explore the coil's central structure based on the selected planar shape and diameter of the coil. Because the complicated coil structure makes it challenging to mathematically express the magnetic flux across the coil, we can only measure it through experiments.

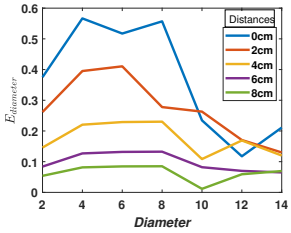


Fig. 8. Coil performance vs. diameter at different distances.

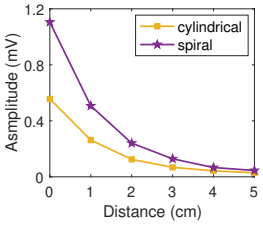


Fig. 9. Induced potential of spiral vs. cylindrical coils.

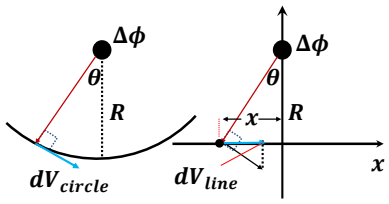


Fig. 10. Induced potential with different wire shapes.

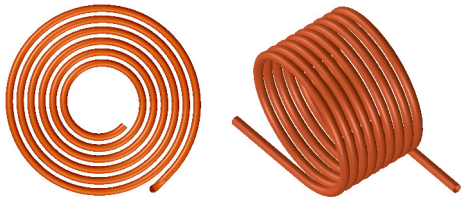


Fig. 11. Spiral coil vs. Unfilled coil.

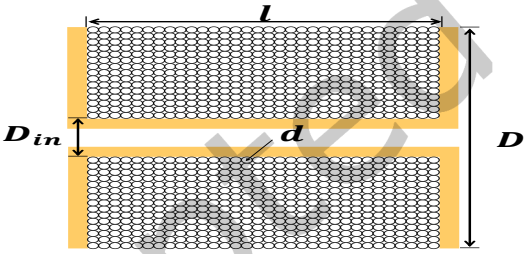


Fig. 12. Cross section of a coil

The magnitude of the induced potential generated is directly proportional to the amount of magnetic flux passing through the coil. Accordingly, in traditional induction coil design, the surface area is enlarged as much as possible to increase the magnetic flux. However, in eavesdropping scenarios, the coil's diameter is restricted, making it challenging to boost the induced potential.

If we fill the middle space of the coil with wires while keeping the overall magnetic flux unchanged, the same magnetic flux will drive more wires. Will the induced potential be bigger in this case than in the unfilled case? To this end, we designed a new coil wound in a spiral shape, as depicted in Figure 11.

To control the variables, the coil was confined to a plane. We compared the experimental results to those presented in Section 4.3 (5000 turns). Given the challenge of measuring the induced potential with just one turn, we compared the result to 1/10 of the value for a 4 cm diameter coil (approximately 0.52), which is the same ratio as the wire length employed in this experiment.

Figure 9 shows that, when the diameter is restricted to 4 cm, the spiral-filled coil could produce an induced potential that is two times larger (about 1.15 at 0 cm) than a normal hollow coil with the same wire length (approximately 0.52).

In experiments, we compared several winding shapes at the center of the coil, such as polygonal (triangular, square, etc.) windings, and observed that their performance was inferior to that of the spiral. This is because the directivity of the induced voltage with a linear wire shape results in attenuation of the induced voltage. As depicted in Figure 10, the induced voltages for 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 wire's arc 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, even When the coil is connected to an op-amp in a shielded room without any magnetic signal an unexpectedly large level of noise would be introduced. This noise is *Johnson–Nyquist noise*, originating mainly from the movement of electric charges inside a circuit. To reduce the impact of signal access on the op-amp, we need to optimize the coil's design. 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 suggests a clearer signal. On the other hand, we can attempt to increase the *sensitivity* so that even a slight change in the magnetic flux could cause a significant change in the induced potential.

As we discuss in the previous section, the optimal design for a coil is a spiral shape. In other words, we need to wrap the wire in a spiral shape to make a cylinder layer by layer, with the cross-section shown in Figure 12. Under the assumption that the center is a hollow circle, we need to calculate the induced voltage accordingly.

However, it is complicated to calculate the induced voltage directly. By approximating each turn of the spiral as a circular coil, we could make the problem simpler. Therefore, the induced voltage is proportional to the total area of these coils. The sum of the areas of all coils in a plane (a vertical row of circles in Figure 12) is $\sum_n \frac{\pi(D_{in}+nd)^2}{4} = \frac{\pi}{4}\sum_n (D_{in}+nd)^2$, where $n = 0, ..., \frac{D-D_{in}}{d}$. To avoid tedious calculations, we use the following conversion method. We assign two coils of different diameters to a group such that their sum equals $D + D_{in}$. 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.

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

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

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

The ***sensitivity*** represents the change in the induced voltage caused by 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 deduce 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 on 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 $V_T = 2\sqrt{k_B T \Delta f R}$, 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:

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

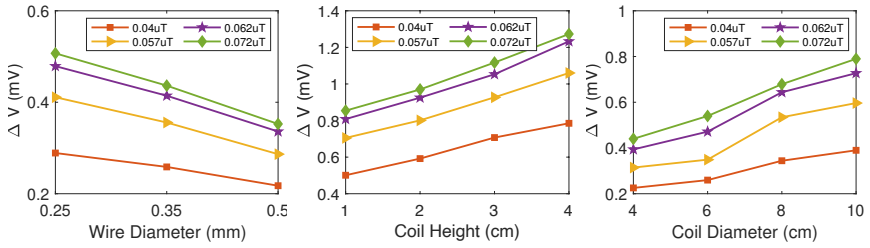


Fig. 13. Relationships between sensitivity and (a) wire diameter, (b) coil height, and (c) coil diameter at different distances

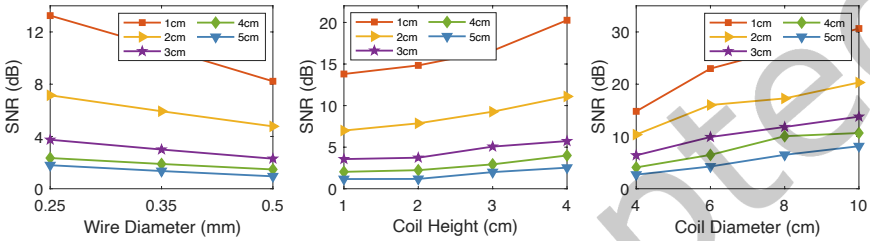


Fig. 14. Relationships between SNR and (a) wire diameter, (b) coil height, and (c) coil diameter at different distances

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:

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

In this SNR model, the first part of the polynomial represents the parameters, while the second part indicates that an increase in the height l of the cylinder enhances the ***SNR***, which is consistent with the experimental results in Figure 14. Second, we obtain the peak value of the ***SNR*** when $\frac{D_{in}}{D} = \frac{1}{2}$. Based on these findings, we chose a wire diameter of 0.25 mm, and a height of 3.5 cm for the coil, which corresponds to the thickness of a whole earphone. Additionally, we set the ratio of the inner and outer diameters to be between $\frac{1}{3}$ and $\frac{1}{2}$.

We conducted experiments to determine if the actual measurements would match the theoretical analysis. Since sensitivity refers to the voltage change induced by a unit of magnetic flux, we studied the sensitivity by observing the variation in the induced voltage as the magnetic flux change. Initially, 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 number of turns (500).

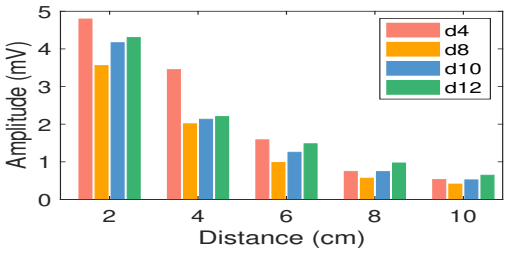


Fig. 15. Optimized coil(D=4cm) vs. larger-diameter coils(D=8,10,12cm).

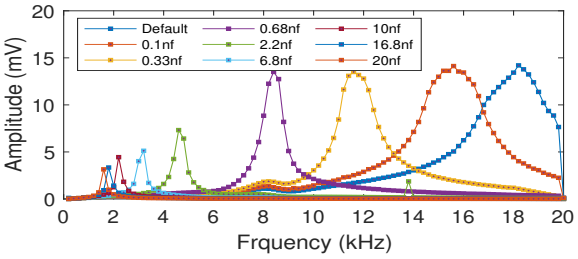


Fig. 16. Frequency response curves when connected in parallel with different capacitors.

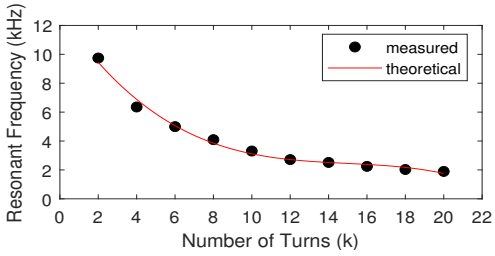


Fig. 17. Peak induced potential and coil turns.

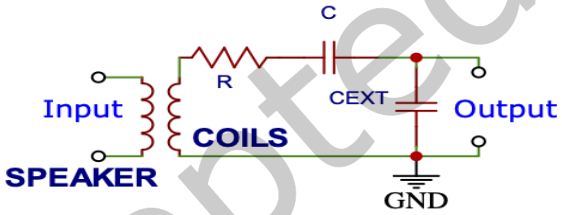


Fig. 18. Equivalent circuit of the coil loaded with an external capacitance and resistance.

Figure 13 (a) and Figure 14 (a) show that a smaller wire diameter results in higher sensitivity and SNR. Next, we designed four coils with varying heights ($l \in \{1, 2, 3, 4\}$), but the same diameter (3 cm) and wire diameter (0.25 mm). Figure 13 (b) and Figure 14 (b) reveal that a greater height leads to higher sensitivity and SNR. To investigate the impact of 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), sensitivity and SNR are directly proportional to the coil diameter. To further improve the inductance of the entire coil, we also inserted an iron core in the inner diameter gap and conducted experiments with the optimized coil. In Figure 15, the label 'd4' denotes the induced voltage of the optimized coil at different distances. We can observe that the optimized coil produces a higher induced voltage than a coil with an unoptimized diameter that is three times larger (labeled 'd12').

In these experiments, we assumed that the eavesdropping signal only consists of a single frequency component; however, this will not be the case 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 amplify the magnetic signal prior to computer processing, the weak magnetic signal is susceptible to being overwhelmed by inevitable circuit noise. As circuit noise is a multifrequency signal that is typically concentrated in the low-frequency range, it is possible to adjust the resonance effect of the coil as a means of weakening the circuit noise and boosting the audio band. To this end, we shall introduce a coil's resonance effect and how we tune its resonant frequency.

The MagEar's receiver is comprised of a coil that serves as an inductance. Its impedance is denoted by $Z_L = j\omega L$, which exhibits a linear increase with the rise in frequency. This suggests that the impedance can offset the amplifier effect caused by frequency. Therefore, we conducted experiments using signals of varying 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 paralleled various capacitors with the coil and plotted the corresponding frequency response curves. The results are shown in Figure 16.

The curve labeled *'Default'* in Figure 16 represents the result of our experiments. In contrast to the theory, the curve does not display a linear or flat trend but instead exhibits a bell-shaped distribution. The inductive reactance of an inductor changes slowly when the frequency is low, resembling a horizontal straight line. However, this observation does not provide an explanation for the rapid rise in induced voltage at the signal range of 14-18 kHz.

We conducted further experiments at frequencies above 18 kHz and observed that the induced voltage decreased rapidly after reaching its peak value. Additionally, we tested coils with different numbers of turns and observed a decrease in the peak frequency as the number of turns increased, as illustrated in Figure 17. This phenomenon can be attributed to the tightly wound nature of the coils, which leads to capacitance between the wires. Consequently, the entire loop becomes a series-connected LC resonant circuit. Our coil exhibited resonance near 18 kHz, resulting in low-frequency magnetic signals with a weak amplitude.

Standard earphones typically operate within the frequency range of 50 Hz to 20 kHz. As it is difficult to predict the specific sound frequencies that will be played by the earphones, increasing the induced voltage signal at the appropriate frequency poses a challenge.

Our findings suggest that connecting a capacitor outside of the coil's circuit can modify the resonance center frequency. This enables us to 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 18, the system transforms into a complex multistage circuit. In this case, the resonance center frequency relies on two parts of the circuit: the RLC circuit in series and the capacitor circuit in parallel. The overall admittance of this circuit can be expressed as follows:

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

where $\omega_0 = 2\pi f$. By setting the overall admittance to $Y_{total} = 0$, we can solve for the root of the equation to determine the center frequency and obtain $f_0 \propto \frac{1}{\sqrt{C_{EXT}}}$. Consequently, as the capacitance C_{EXT} increases, the center frequency will shift to a lower value. Accordingly, as shown in Figure 16, we can utilize variable capacitors to adjust the center frequency, with larger capacitances resulting in a shift towards lower resonance frequencies. Specifically, the resonance frequency is inversely linearly proportional to the capacitance.

Notably, different types of sounds have different frequency ranges. Typically, a small bandwidth is required when MagEar is used to eavesdrop on speech. However, if the target sound is music, a broader bandwidth may reduce 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 of the circuit-sliding varistor. This configuration allows us to adjust the variable resistor to decrease the impedance, thereby reducing the quality factor Q and increasing the bandwidth when a wider frequency range is required. The noise spectrogram in Figure 19 indicates that the circuit noise is concentrated below 400 Hz and persists throughout the audio. Through experiments, we found that tuning the resonance frequency of the coil to a range between 1500 and 2500 Hz can yield satisfactory performance, minimizing the impact on the audio band while effectively attenuating the circuit noise.

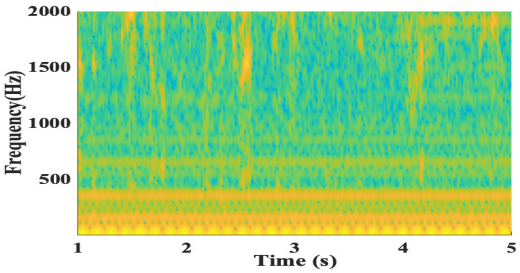


Fig. 19. Noise spectrogram.

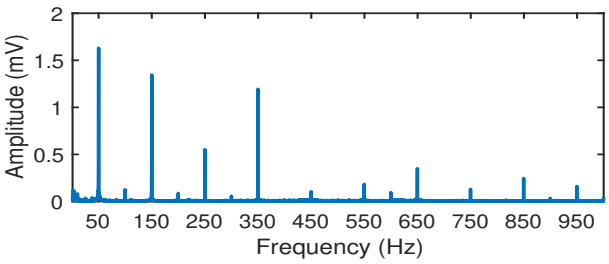


Fig. 20. Noise with 50Hz odd harmonics in the frequency spectrum.

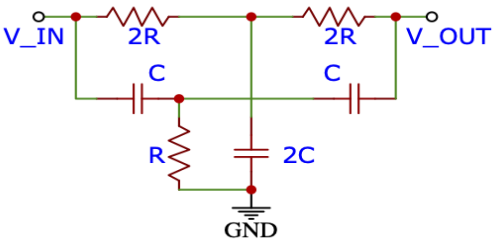


Fig. 21. AC noise removal component.

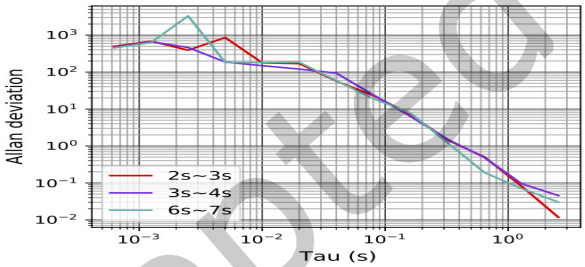


Fig. 22. Allan deviations of noise at different times(seconds).

5.2 Interference Confrontation

After transforming the magnetic signal that leaked from the earphones into sound, we observed a significant amount of noise obscuring the sound. This noise is prevalent both indoors and outdoors, with indoor environments exhibiting more severe noise levels. Figure 20 displays the frequency spectrum of a recorded magnetic signal.

At roughly 50 Hz, there is a sudden increase in amplitude to its maximum level. There are other noise peaks at all odd multiples of 50 Hz. These observations suggest the presence of a noise signal with a 50 Hz fundamental frequency that generates other odd harmonics. Considering that our power supply comes from a stable voltage battery, this component is likely to be hum, a type of noise emanating from the AC cable. This hum diminishes the intelligibility of the recovered audio and is unpleasant to the ear. To mitigate this noise, we employed a notch filter module, as depicted in the circuit diagram in Figure 21. We connected the module to the output signal and employed a three-stage cascade method to eliminate the noise interference from AC power.

Unfortunately, even though we were successful in removing the loudest interference using this method, the noise problem was not entirely resolved. Considering that the circuit noise has a far greater amplitude than the intended magnetic signal and exists throughout the duration of the audio, relying solely on adjusting the coil's resonance effect is inadequate to eradicate this interference. Moreover, electromagnetic radiation from surrounding electronic devices may decrease the system's performance. 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 frequency domain differences. Moreover, we used Allan variance analysis, which was initially proposed to measure the frequency stability of clocks but can also be applied to characterize various types of noise in sensor data. As illustrated in Figure 19, the noise amplitude is negatively correlated with the frequency between 0 and 1000 Hz. The noise slope remains stable ($Amplitude = -10^{-4}f$). As can be seen from the Allan variance results in Figure 22, the slope of the curve on a double logarithmic graph is close to -1, indicating that the signal mainly exhibits white noise across the entire frequency spectrum. The peak between Tau values of 10^{-3} to 10^{-2} indicates the presence of sinusoidal noise between 100 Hz and 10000 Hz.

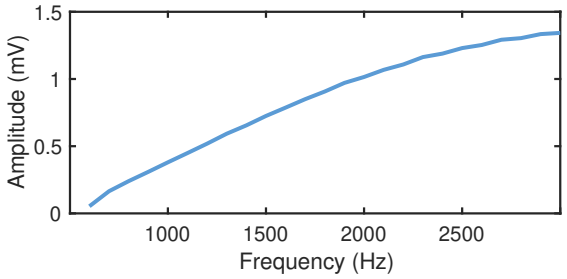


Fig. 23. Induced voltage proportional to the frequency.

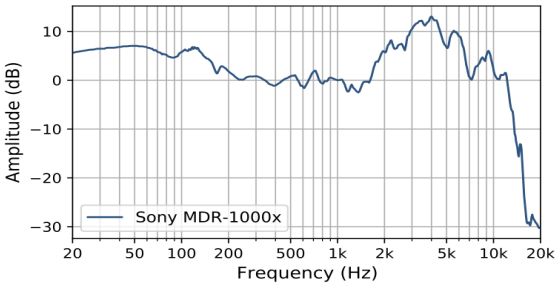


Fig. 24. Frequency response curve of the Sony MDR-1000x.

As this noise is a form of additive noise and is uncorrelated with the desired speech, we can employ spectral subtraction for noise reduction. The spectral subtraction technique involves estimating the noise spectrum during silent periods and subtracting it from the original spectrum to obtain a clean speech signal.

In detail, the calculation for obtaining the clean spectrum by subtracting the noise spectrum from the original 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, some speech information will be eliminated along with the noise, hence diminishing the speech's intelligibility. However, if α is too tiny, considerable noise will remain. In our system, β is set to 0.02 and α is chosen as described in [8]. The denoised time-domain signal can then be obtained by applying the inverse discrete Fourier transform on the power spectrum.

5.3 Frequency Response Equalization

As previously mentioned, we can eliminate the noise using the techniques described earlier. However, the audio recovered by MagEar still exhibits distortion, particularly in the form of sharper sound than the original audio. This deformation has two underlying causes. One is the fact that the magnetic field of the earphones does not correspond to the vibration amplitude, and the other is the damping effect introduced by the resonance of the coil.

Let's begin by discussing the inequality between the magnetic field and the vibration amplitude of the headphones. The magnetic force is utilized to accelerate the earphone coil, and when the coil needs to vibrate rapidly, the acceleration must increase to enable the coil to reach its desired position in less time. For sounds with the same amplitude but different frequencies, higher frequencies require faster movement, resulting in higher acceleration that necessitates a greater magnetic force.

We conducted experiments using headphones to play sounds of equal amplitude but varying frequencies. The results depicted in Figure 23 demonstrate that as the frequency increases, the induced potential also increases. After normalizing this curve, we refer to it as the frequency–magnetic curve $\eta(f)$, which represents the variation in the magnetic field coefficient as a function of frequency.

Another cause of distortion originates from the resonance effect. The damping effect is at its minimum when the signal is at the resonance center frequency, and it increases as the signal deviates from the center frequency, leading to a reduction in the induced voltage. Note that the damping effect can also be altered by adjusting the capacitance (Figure 16).

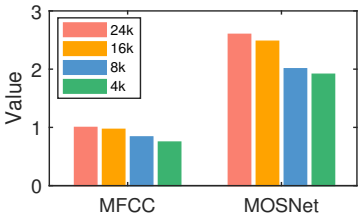


Fig. 25. Audio quality at different downsampling rates.

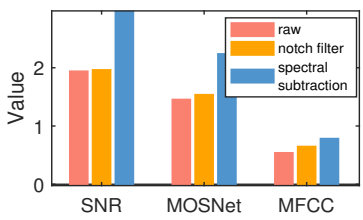


Fig. 26. Improvement in audio quality achieved with denoising algorithms.

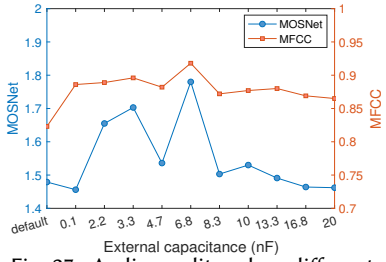


Fig. 27. Audio quality when different capacitances are connected in parallel.

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

The restoration process is divided into the following steps. When we receive eavesdropping signals, we first calculate the resonance frequency f_0 in accordance with the chosen capacitance and obtain the response function $V_B(f_0, C)$. Then, we multiply the signal at all frequencies by $\frac{V_B(f_0)}{V_B(f)}$ to acquire 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 multiplying by the scaling factor, the restored audio is still of poor perceived quality. The reason is that humans are not accustomed to a flat audio response curve. Therefore, to achieve audio restoration, we utilize the Harman curve as a benchmark. The Harman curve is a standard curve utilized by earphone designers to measure earphone sound quality. We use the Harman curve coefficients to readjust the sound again based on the frequency spectrum. After restoration, our ultimate frequency response curve is depicted in Figure 24.

5.4 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. A short-time Fourier transform and a Mel filter bank is applied to estimate the energy distribution characteristics of audio in the frequency domain; specifically, 12 cepstral-domain values are used to measure these characteristics.

To evaluate speech quality, we extract the MFCCs of played audio and recovered speech as their features. We then compute the cosine similarity of two sets of coefficients. Cosine similarity corresponds to the cosine value of the included angle between two vectors. The closer the angle between the two vectors is to 0, the closer their cosines are to 1. Hence, the cosine similarity values are between -1 (opposite) and 1 (same). For the sake of brevity, we will refer to the cosine similarity between the MFCCs of two audio samples as 'MFCC' in the subsequent sections.

MOSNet: Listening tests often use the mean opinion score (MOS) to evaluate audio quality. This score is determined through listener ratings on a 5-point scale, taking into account subjective factors like sound clarity and comfort. 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 [26]. It considers human MOS

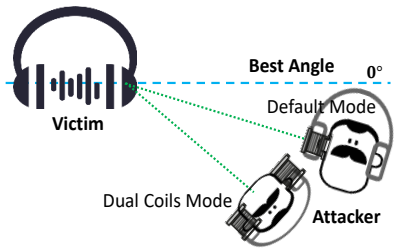


Fig. 28. Two operation modes at different angles.

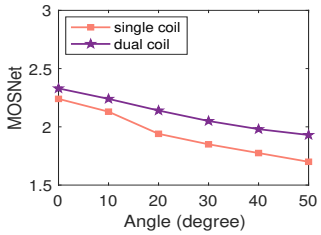


Fig. 29. Performance comparison between a single coil and dual coils at different angles.

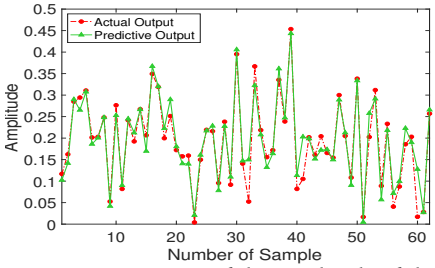


Fig. 30. Comparison of the amplitude of the received magnetic signal and the predictive output of a BPNN.

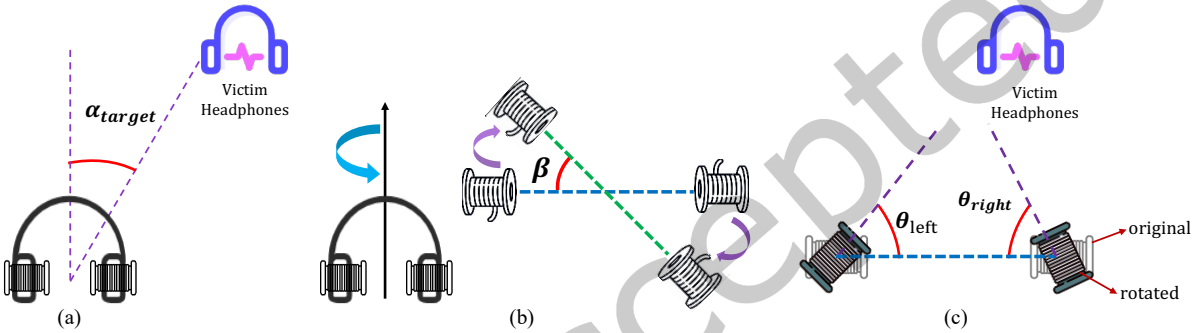


Fig. 31. We define four angles. (a) α_{target}: the angle formed by the center line of the two receiver coils and the target headphones. (b) β_{head}: the degree of an attacker's head rotation. (c) θ_{left} and θ_{right}: the rotation angle of the left coil and right coil.

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 assess the effectiveness of these two metrics, we downsampled an audio recording with a 24 kHz sampling rate to a lower sampling rate and calculated the MFCC and MOSNet results. As shown in Figure 25, as the sampling rate decreases, the MFCC and MOSNet values also decline. The MFCC and MOSNet 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 utilized the MFCC and MOSNet scores to evaluate the performance of various algorithms in enhancing the recovered audio quality. Figure 27 shows the achieved audio quality when a coil is connected in parallel with different capacitors. It can be observed that the relationship between audio quality and external capacitance is not linear. For the designed receiver coil, 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 a receiver coil has a resonance frequency below 2 kHz, we also cannot obtain high-quality audio. 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 speech quality. In addition, Figure 26 demonstrates that applying a notch filter and spectral subtraction significantly improves the audio quality.

Table 1. Comparsion of the default amplitude and the maximum amplitude obtained under the optimal coils' placement as the target headphones' angle varies.

α_{target}	Default amplitude	Optimal amplitude	Optimal angle $[\alpha_{target}, \beta_{head}, \theta_{left}, \theta_{right}]$
0°	0.274	0.5233	[0°, 75°, −60°, −60°]
15°	0.2433	0.4936	[15°, 75°, −60°, −60°]
30°	0.2635	0.4467	[30°, 90°, −60°, −30°]
45°	0.3468	0.4491	[45°, 90°, −30°, −30°]
60°	0.3023	0.4637	[60°, 0°, 60°, 60°]
75°	0.1709	0.4512	[75°, 0°, 60°, 60°]
90°	0.1514	0.4894	[90°, 15°, 60°, 60°]

5.5 Angle Adjustment

5.5.1 Dual Coils Mode. In the previous section, we assumed that the MagEar coils and the victim's earphones were aligned along the same axis. However, in many cases, it's tough to achieve. Because the victim's headphones could be in any position, and the angle between them could be as great as 90°. Unfortunately, an angular offset can cause a reduction in the received magnetic flux, which in turn reduces the quality of the audio obtained.

MagEar employs a dual-coil mode to solve this problem, as shown in Figure 28. In the 'default mode' scenario, an adversary conceals a single coil in a headphone shell. However, in dual-coil mode, the adversary hides one coil on each side of the headphone shell. When the eavesdropping angle is not ideal, this mode enables an attacker using MagEar to rotate to face the victim from the front rather than the side. For example, when the angle between the victim's headphones and the receiver coil is 30 degrees, MagEar's performance degrades compared to the optimal-angle case. In the dual-coil mode, 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 raises the induced voltage of the received signal because two coils form a mutual inductance, thereby enhancing the intensity of the induced voltage. In fact, as shown in Figure 29, the induced voltage is often higher than twice that of a single coil.

5.5.2 Optimal Angle. On the other hand, we can also rotate the angle of two receiver coils to better adapt to the changing victim's position. To be specific, we define four angles to describe the rotation movement of the coils. We assume that the two receiver coils are placed at the same positions on each side of the headphones. First, we refer to the angle formed by the center line of the two receiver coils and the target headphones as target angle α_{target}, as shown in Figure 31(a). Then, we introduce a head rotation angle of the attacker. Imagine a scenario where a victim is standing right in front of us, we could swivel towards the victim to allow as much magnetic flux as possible through the receiver coils. Therefore, we define the degree of head rotation as β_{head}, which also indicates the angle of simultaneous rotation of the internal receiver coils, as displayed in Figure 31(b). In addition to turning the head, we can also control the coils' rotation. We describe the left and right coil's rotation angles as θ_{left} and θ_{right}, respectively.

To investigate how these four angles influence received magnetic signals, we attempted to find a function to describe their respective impacts. However, because there are four variable angles and the relationship among them is not linear, it is challenging to fit a multidimensional nonlinear function with a traditional mathematical model. Thinking of the powerful nonlinear mapping ability of Back Propagation Neural Network (BPNN), we train a model to express the mapping relationship between the four angles and the amplitude of the measured signal. To acquire enough training data, we carried out experiments from different angles. Considering that in real attack

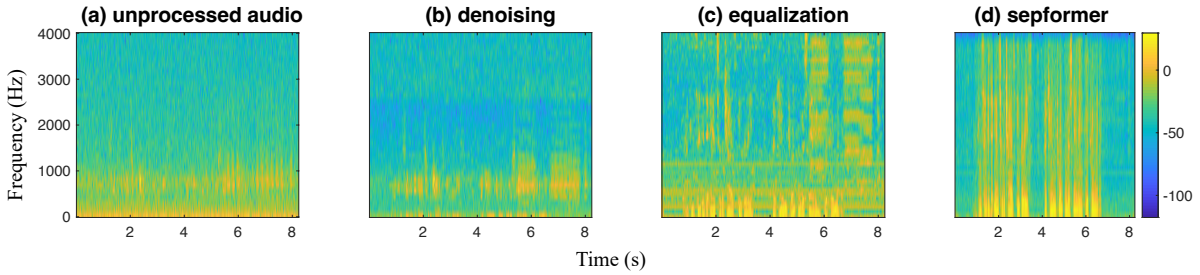


Fig. 32. The spectrograms of (a)unprocessed audio, (b)denoising audio, (c)equalization audio, (d)enhanced audio

scenarios, it is difficult for an attacker to accurately estimate the relative angle to the victim's headset. Therefore, we have selected seven common and well-recognized angles. Specifically, the values of α_{target} and β_{head} were both varied in the range $\{0°, 15°, 30°, 45°, 60°, 75°, 90°\}$. And given that the coils could be rotated in both left and right directions, θ_{left} and θ_{right} were taken as $\{-30°, -60°, 0°, 30°, 60°\}$. Considering that it is hard to precisely calculate the angle of the target victim relative to the attacker and the attacker's head-turning angle in practice, we selected those ordinary integer angles in this experiment. For each group of angles $[\alpha_{target}, \beta_{head}, \theta_{left}, \theta_{right}]$, we recorded its corresponding received signal. Therefore, there are $7 \times 7 \times 5 \times 5$ sets of data in total. In this experiment, we played 1kHz sine wave audio using Apple AirPods, and the left and right coils were connected in series and output through ADC. Then, we recorded the amplitude of the measured magnetic signal at 1kHz.

We divide 95% of 1225 sets of experimental data into the training dataset and use the remaining 5% as the test dataset. Figure 30 shows the predicted amplitude of the test data using BPNN and the actual amplitude collected in the experiment, and the mean square error is approximately 0.001. Table 1 illustrates how the amplitude of the received signal increases by changing the angle of receiver coils when the position of the target's headphones varies ($\alpha_{target} \in \{0°, 15°, 30°, 45°, 60°, 75°, 90°\}$). In this table, 'default amplitude' indicates the amplitude of signals acquired when the two coils are placed in a default way (β_{head}, θ_{left} and θ_{right} are all equal to 0°). 'optimal amplitude' means the amplitude value obtained with the optimal coils' arrangement, i.e. the maximum value., while 'optimal angle' represents the current coils' angle ($[\alpha_{target}, \beta_{head}, \theta_{left}, \theta_{right}]$).

Furthermore, in practice, the angle of the victim's headphones can be arbitrary, i.e., not among the seven fixed angles. By carefully inspecting our calculated optimal angle combinations, as given in Table 1, we find that there is a certain linear variation between different α_{target}. For example, for $\alpha_{target} = 75°$ and $\alpha_{target} = 90°$, their corresponding optimal angles are [75°, 0°, 60°, 60°] and [90°, 15°, 60°, 60°], respectively. The distinction between these two cases lies in the value of β_{head}, which has increased from 0° to 15°. Thus, for an arbitrary alpha between 75° and 90° (e.g., 85°), we can gradually add the value of its β_{head} according to Table 1. Likewise, for any angle in the other ranges, we can refer to the results in Table 1 to dynamically adjust our two coils' placement angles.

5.6 Speech Separation

In our previous section, we employed a denoising technique to preserve the original speech features of the data effectively. This approach is crucial because many speech recognition models heavily rely on data-driven, especially those models integrated into Application Programming Interfaces. By utilizing this method, the speech recognition model can maintain a relatively high accuracy rate even in the presence of noise within the speech data. This positive outcome is attributable to the data-driven approach, which places significant emphasis on speech features and their combinations. However, it is worth noting that simply focusing on data features in the denoising process can lead to a lack of user-friendliness. In speech recognition, unlike machines, humans may not accurately discern when noise closely resembles the target sound. Consequently, relying solely on data-driven denoising methods can result in a subpar listening experience for individuals. As a means of enhancing the listening experience, we will introduce an alternative denoising solution.

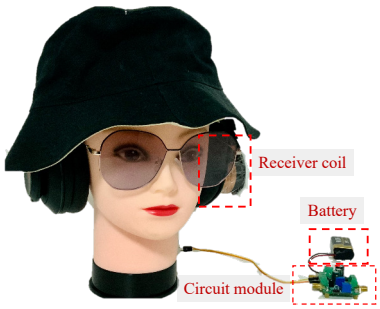


Fig. 33. Diagram of the eavesdropping coil.

We utilized SepFormer [32], a neural network based on the Transformer model, for speech separation. This RNN-free approach enables the SepFormer to effectively capture both short and longterm dependencies by utilizing a multiscale methodology that leverages transformers.

The SepFormer network comprises the encoder and masking network, serving as its key components. The encoder block aims to generate a well-learned representation of the input signal, utilizing transformers to grasp both short and long-term dependencies. By leveraging the parallelization advantages of Transformers, the encoder achieves remarkable performance, even with a downsampling factor of 8, enabling faster processing and reduced memory requirements. On the other hand, the masking network focuses on estimating optimal masks for separating different sources within the mixtures. It utilizes multiple SepFormer blocks, transformer-based neural networks dedicated to learning and estimating the best masks for each source. These blocks are specifically designed to capture the dependencies between the sources and comprehend the long-term structure of the input signal. Additionally, the masking network incorporates a chunking module that breaks down the input signal into smaller portions, thereby reducing the computational complexity of the SepFormer blocks.

It is essential to acknowledge that this approach aims to extract speech relevant to the text, thereby eliminating noise interference and ultimately enhancing the overall quality of the speech. However, it is crucial to recognize that this method may result in the loss of specific speech characteristics. Further details regarding this will be discussed in the subsequent experiments.

Figure 32 shows the spectrogram of the acquired signal processed by our algorithm at a distance of 60cm. We observed that SepFormer effectively removed most of the noise in the signal and enhanced the speech features, resulting in a better listening experience.

6 IMPLEMENTATION

As our receiver, we designed a coil with a diameter of 4 cm and a height of 3.5 cm. In addition, an external capacitor was connected to the coil in parallel to adjust the resonance frequency. To enhance the received signal, the coil was connected to an AD620 amplifier with a gain of 2000×. Then, the coil was connected to a USB3202 data acquisition board, which features a 16-bit ADC to digitize the signal. This ADC supports a sampling rate of up to 250k, covering the entire audible frequency band of human speech. Finally, the ADC was connected to a laptop through a USB port for processing the received signal.

7 EVALUATION

7.1 Experimental Setup

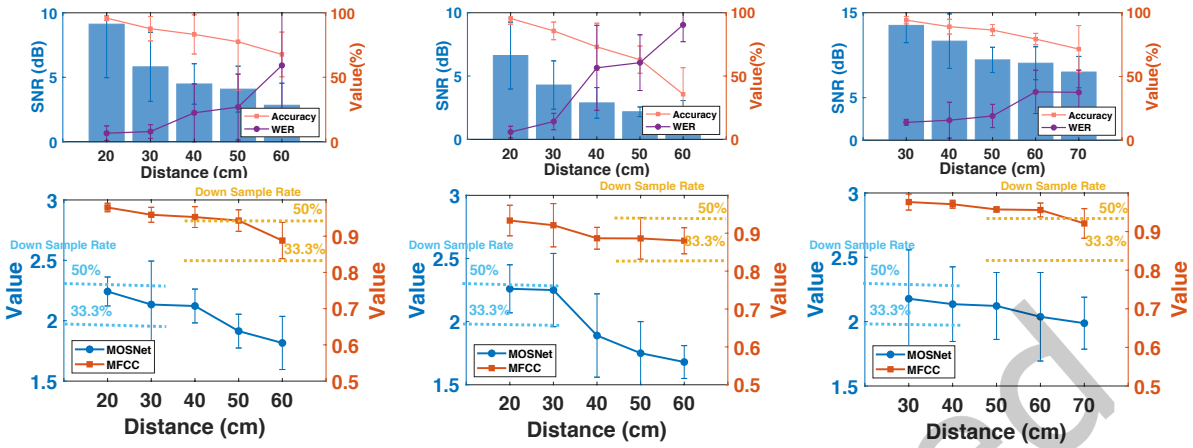
Data: In our evaluation, we used audio from LibriTTS as the eavesdropped content. LibriTTS is a collection of English speech recordings from multiple speakers, which includes around 585 hours of read speech in English at a sampling rate of 24 kHz [38].
Metrics: We used objective (MFCC) and subjective (MOSNet) measures to assess speech quality.
Automatic speech recognition: We conducted additional experiments to evaluate the application of voice recognition on the recovered audio. To transcribe the magnetic sound into text, we utilized the Google speech-to-text (STT) API. The Levenshtein ratio and the word error rate (WER) were employed as speech recognition metrics to evaluate the similarity between the transcribed text and the ground-truth text. 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 a similarity result as 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 earphone's audio loudness was adjusted to 80 dB (80% volume for earphones) and measured using a SMART SENSOR AR844 digital sound level meter. Note that this loudness value is only perceived within the ear canal and is inaudible to anyone outside. The MagEar and the eavesdropped device were positioned at a fixed angle, with the coil oriented towards the speaker to maximize the magnetic flux passing through it.

7.2 **Eavesdropping on Different Speakers**

Eavesdropping on the content of someone's phone calls is one of the most common eavesdropping scenarios. Generally, this can be achieved by implanting malware on the victim's smartphone. If malware implantation is not feasible, a miniature acoustic microphone can be secretly deployed in the vicinity of the victim to record the speech. However, the voice at the remote side of the phone call is typically too small to be recorded in this way. As a supplement, MagEar can eavesdrop on the remote side by recovering sound signals based on the magnetic leakage of the phone's speaker.

In this experiment, we evaluated MagEar's ability to eavesdrop on cellphones and headphones. During the first stage, the subject was instructed to hold five different smartphones in close proximity to their ear. In the second stage, they were asked to wear ten different earphones, each connected to a smartphone. Five tested smartphones including iPhone Xs, OnePlus 7 Pro, and Huawei P30, while ten evaluated headphones including Airpods, Earpods, Sony WH-1000XM3, and ATH-M50x. The distance between the coil and the eavesdropped device ranged 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 during each round, and we collected the magnetic signals for analysis.

Figure 34 shows the recovered speech quality for non-in-ear headphones, earphones, and smartphones at varying distances. The top three subgraphs display the results for SNR, accuracy, and WER, while the bottom subgraphs illustrate the MFCC and MOSNet values. The results indicate that MagEar can effectively restore in-ear and non-in-ear headphone audio with good preservation of sound characteristics within a distance of 60 cm, despite a significant drop in SNR. 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 due to their closed structure and ability to produce sound with lower vibration. Additionally, their piezoelectric systems consume less energy, resulting in smaller magnetic fields than those of non-in-ear headphones. For smartphones, the average ASR accuracy is 95.76% and the average WER is 13.99% at a distance of 30 cm. As the distance increases to 70 cm, the average accuracy decreases to 80.78% and the average WER increases to 37.54%.



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

Fig. 34. Performance in phone call eavesdropping with different distances between the speaker and MagEar.

In addition to recognition accuracy and WER, we also display objective (MFCC) and subjective (MOSNet) measures in the bottom subgraphs in Figure 34 to provide a more comprehensive evaluation of MagEar's performance. To further assess the capabilities of MagEar, we downsampled audio sampled at 24 kHz to several lower sampling rates (i.e., 4 kHz, 8 kHz, 12 kHz, and 16 kHz) and calculated the MFCC and MOSNet scores. The yellow dotted lines in Figure 34 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 MagEar performance metrics exceed those of audio with a 30% downsampling rate and come close to those with a 50% downsampling rate.

7.3 Physical Obstacles

Without loss of generality, we focused our evaluation on a single earphone model (i.e., AirPods) and one smartphone model (i.e., iPhone Xs) in the following experiments. Despite the belief held by most individuals that physical barriers such as walls, soundproof glass, or doors can prevent eavesdropping, our experiment demonstrated the capacity of MagEar to eavesdrop even in the presence of such obstacles. We experimented with concrete walls of three thicknesses (23, 33, and 42 cm), where the eavesdropping distance was equal to 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, with an eavesdropping distance of 40 cm.

The results for the earphones and the smartphone are presented in Figure 35 and Figure 36, respectively. The bar and line graphs represent the MOSNet and MFCC values, respectively. The w/o wall and w/ wall in both figures refer to the case of without a wall and with a wall between the target device and the receiver coil, respectively. It can be seen that the presence of obstacles has a negligible effect on magnetic side-channel eavesdropping, as these barriers possess low conductivity, thereby allowing magnetic signals to pass through with minimal loss.

7.4 Eavesdropping in Different Environments

Next, we evaluated MagEar in diverse environments beyond the laboratory in which the previous experiments were performed. The eavesdropping distance was fixed at 40 cm for AirPods and iPhone Xs. Our test sites comprised indoor and outdoor locations, namely, a park, a coffee shop, a restaurant, and a private residence, each exhibiting distinctive ambient magnetic field strengths. For example, in a coffee shop, many customers may use laptops, tablet PCs, or smartphones for business or entertainment. A coffee shop's light bulbs and power lines

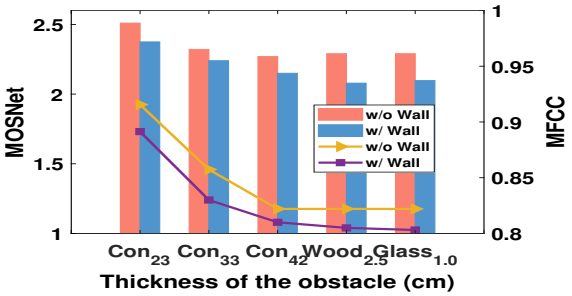


Fig. 35. Results of eavesdropping on AirPods through various obstacles.

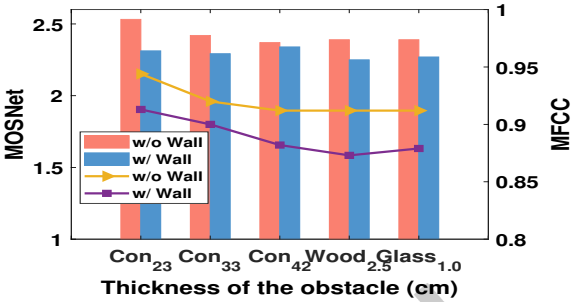


Fig. 36. Results of eavesdropping on an iPhone through various obstacles.

Table 2. Eavesdropping performance in different environments for AirPods and the iPhone Xs (distance=40 cm)

	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

will also emit electromagnetic radiation. The electromagnetic interference emitted from such electronic devices may affect the performance of our system.

In contrast, outdoor environments like parks are typically less congested, with fewer electronic devices in use, resulting in relatively weaker magnetic field strengths. As shown in Table 2, all selected environments exhibit varying degrees of magnetic field interference and noise disruption. We can see that the overall MFCC and MOSNet performance suffers a certain degree of degradation in indoor public scenarios. On the contrary, the park scenario exhibited lower magnetic field interference than the laboratory, leading to higher MFCC and MOSNet scores.

7.5 Enhance Listening Experience

Different from machines, human attackers require an easily distinguishable sound environment. As Section 5.6 shows, we leverage SepFormer[32] for speech separation and enhancing the listening experience of recovered audio. We conducted experiments to investigate the effectiveness of SepFormer in extracting speech content. Specifically, we used Sony WH-1000XM3 headphones to play audio from LibriTTS[38] and placed our receiver coil at various distances from the headphones to capture the magnetic field signals separately. The attack distance ranged from 30cm to 60cm.

Since MOSNet computes subjective human ratings of audio quality, we use MOSNet as a metric for evaluating speech separation performance. Figure 37 shows the improvement of MOSNet values of magnetic signals measured at different distances after being processed by our algorithms. Even when the audio has been denoised and frequency response equalized, applying SepFormer can still significantly increase its MOSNet value. For example, at a distance of 50cm, compared with the unprocessed audio, the MOSNet value increases from 1.593 to 2.023 after applying the denoising and equalization techniques and further rises to 2.514 after applying SepFormer.

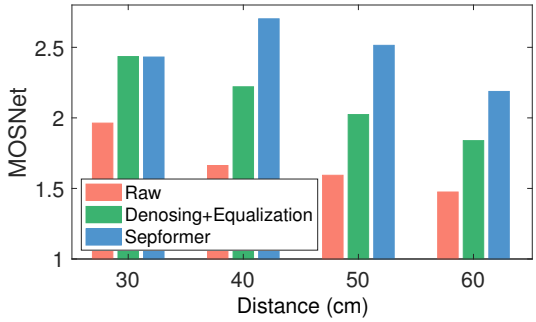
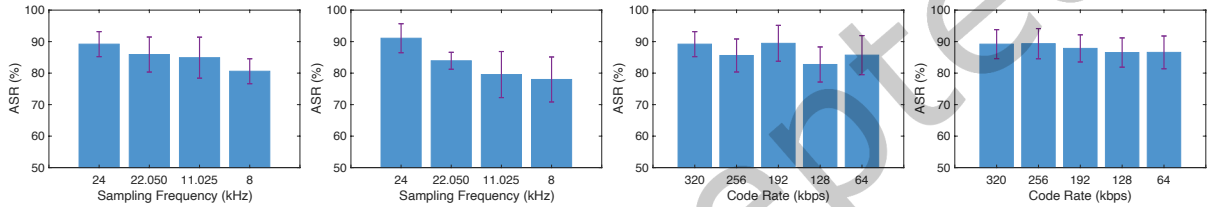


Fig. 37. Improvement in MOSNet values with speech separation.



(a) iPhone Xs (b) Airpods (c) iPhone Xs (d) Airpods

Fig. 38. Speech Recover Performance for different sampling frequency and code rate

7.6 Songs Identification

In this eavesdropping case, we perform song identification to validate the possibility of using magnetic leakage signals to match the audio pattern. Specifically, we use a normal speaker to replay the songs eavesdropped by MagEar and identify the songs using a mobile application—Shazam[33].

Shazam is a music application owned by Apple Inc. It can identify songs given a short sample of the audio track being played. The powerful function is achieved by analyzing the frequency components of a given sound and seeking a matched spectrum fingerprint of over 12 billion songs in the database.

Specifically, we downloaded 100 hit songs in 5 different languages, 20 songs for each language. Note that all of the songs are able to be identified by Shazam under normal playback conditions. We feed Shazam with "magnetic songs" eavesdropped from different earphones, smartphones, and smart speakers. The distance between the coil and playback devices is set to be the critical distance where ASR $\approx$ 60%. The results show that Shazam can identify 96.6% of songs eavesdropped by MagEar.

7.7 Recover different quality of sound source

In this section, we explore how the quality of sound source affects the recovered speech by MagEar. Specifically, we resample the audio sample into different sampling frequencies and code rates. The standard sample rate of telephone communication, AM radio, FM radio is 8000 Hz, 11025 Hz, and 22050 Hz, respectively. The code rate iterates over 64 kbps, 128 kbps, 192 kbps, 256 kbps, and 320 kbps as typical MP3 audio provides. Figure 38 depicts the corresponding results. We can see that the lower sampling frequency leads to a lower ASR accuracy, while code rate is not a critical factor that results in a lower ASR accuracy.

8 RELATED WORK

8.1 Speech Eavesdropping

Over the past few years, many researchers have put forward a range of side-channel attack schemes to effectively intercept human speech by utilizing non-acoustic devices such as accelerometers, gyroscopes, geophones, WiFi, vibration motors, hard drives, cameras, magnetic coils[24], and photodiodes, which raises huge concerns on security risk of leaking personal privacy and confidential information. Existing non-acoustic speech eavesdropping methods can be divided into two main categories, namely, classification-based and recovery-based.

Classification-based: Gyrophone [27] exploits motion sensors for eavesdropping; specifically, it utilizes the gyroscope on a smartphone to record speaker vibrations. The vibration readings reflect the speech information and are utilized to identify the pronunciation of digits. Building on a similar concept, multiple studies have confirmed the feasibility of using accelerometer data for detecting hot words [39] and recognizing isolated words [6, 7]. S Abhishek Anand et al. [5] explored the correlation between speech and motion sensor data and determined that motion sensors can solely detect variations induced by surface vibrations or conductive vibrations, whereas aerial vibrations lack sufficient power to impact motion sensor readings. WiHear [35] can recognize several words based on analysis of channel state information (CSI) patterns that are affected by the movement of the user's mouth. One of the main disadvantages of these systems is that they use machine learning algorithms to classify sound samples, which can only recognize a few words that are stored in the database in advance. On the other hand, MagEar is an external device that does not assume that the attackers obtain access permission to the motion sensor of the victim's device by installing malware. The recovered speech of MagEar is intelligible and can be identified by both humans and machines with high accuracy.

Recovery-based: VibraPhone [30] exploits the reverse electromotive force of a smartphone's vibro-motor to convert it into an acoustic microphone. In [23], Kwong et al. suggested employing a magnetic hard disk as a microphone and demonstrated that they could reconstruct human speech by measuring the deviation of the read/write head from the track's center. Laser/light-based techniques [28, 31, 34] for eavesdropping are prevalently used in espionage activities. A light/laser beam is pointed at an object in the area nearby the victim through a window. The eavesdropper can then recover the speech by analyzing the fluctuations of the reflected beam. A visual microphone [4] captures the micro-vibrations caused by airborne sound pressure from an object (such as an empty snack bag) using a high-speed camera with a frame rate of 2200 FPS to restore human speech in proximity to the object. Lamphone [29] employs a telescope to focus on a light bulb hanging 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 retrieved by analyzing the fluctuations in the photodiode readings. The authors of ART [36] successfully demodulated the vibration signal of a loudspeaker from the received Wi-Fi packet's RSS readings. They were able to recover the sound of a piano and recognize ten digits by analyzing the RSS. In contrast, MagEar takes advantage of electromagnetic leakage of the speaker and recovers the inductive current in a coil to intelligible speech. In addition, MagEar is able to eavesdrop on victims under different environment settings, including through-wall settings.

8.2 Magnetic/Electromagnetic sensing

On the one hand, numerous researchers have proposed interesting sensing applications by leveraging the properties of magnetic and electromagnetic fields. Maghacker [25] is a sensing system that can deduce the contents of handwriting by intercepting the magnetic field emitted from the permanent magnet in a stylus pen. MET [17] tracks the position and orientation of an electric toothbrush based on the unique magnetic field generated by the motor inside the toothbrush to determine whether the user's gestures are correct. Hayashi et al. [15] demonstrated the capability to snoop on the contents displayed on a tablet by capturing the electromagnetic emissions from the display interface. The authors of MagAttack [10] discovered that the various system instructions composing

PC applications are reflected in the CPU consumption of devices and, thus, the corresponding electromagnetic emissions. Hence, it is possible to leverage electromagnetic interference signals radiated from a CPU to infer device activities, i.e., application launching and application execution.

Choi et al. [11, 12] revealed a security vulnerability in Samsung Pay, where they developed magnetic coils to monitor the magnetic fields generated by magnetic secure transmission devices, enabling them to decode one-time payment tokens decoded from the magnetic signals. NFC+ [40] implemented a long-range magnetic field reader with multiple transmission coils based on the resonance effect for a near-field communication network. In [13], the authors proposed a novel approach for intercepting visible light communications through a wall.

9 COUNTERMEASURE

In terms of defensive countermeasures, one potential defense is the implementation of magnetic shielding material within the headphone shell, which can effectively reduce magnetic leakage from headphones. High-permeability metals, such as iron, silicon steel, or permalloy, are commonly utilized for shielding low-frequency magnetic fields. Due to their low magnetic resistance, high-permeability materials concentrate magnetic induction lines within the shielding material and prevent them from passing through any surrounding cavities. To investigate the efficacy 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 was positioned 30 cm apart. After the installation of a 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 declined from 2.19 to 1.88. These results show that high-permeability metals can effectively impede the propagation of magnetic fields and significantly impact system performance. Therefore, we recommend that headphone manufacturers incorporate magnetic shielding materials within their products to mitigate privacy risks arising from magnetic leakage.

10 DISCUSSION

As a proof-of-concept eavesdropping system exploiting magnetic side-channel information, MagEar still has certain limitations that require addressing in the future.

10.1 Eavesdropping range

The primary drawback of MagEar is its eavesdropping distance, which currently stands at approximately 60 cm. Since the source of the magnetic field signal is beyond our control, two main methods can be adopted to extend the eavesdropping distance. The first method is to replace the customized coil with high-end magnetic sensors, such as fluxgate sensors and optically pumped magnetometers, that can measure magnetic fields on the order of pT [16]. Alternatively, advanced noise reduction algorithms or low-noise circuits can be implemented to improve the system's performance.

10.2 Eavesdropping on multiple targets

It is typical that multiple victims are present in front of the eavesdropper. In such scenarios, the eavesdropper may have two kinds of operation purposes. The first one is to eavesdrop on only one victim, and the eavesdropper wants to eliminate the interference of the speaker equipment of the people around. The second one is to eavesdrop on multiple victims simultaneously with relatively high system performance. In our pilot study, the ASR accuracy will suffer a certain degradation if an extra smartphone or headphones are present. The closer the two devices are, the poorer the result we get. If we want to realize the second purpose, multiple coils should be used to form an array. NFC+ has realized the magnetic beamforming using multiple coils for reading multiple tags simultaneously. Therefore, it is possible for MagEar to eavesdrop on multiple targets in the future.

10.3 Mobility

Sensing systems based on RF signals are generally sensitive to dynamic environments (e.g., human mobility). However the directionality of the magnetic field signal is relatively strong, and there will be no multipath effect. In this paper, we can discuss mobility from three aspects. 1) We design MagEar as a lightweight ear-worn eavesdropping device. But the adversary can also eavesdrop with MagEar in his hand, which may cause hands to tremble. In fact, we found that this type of mobility had little effect on system performance during our experiments. 2) In addition, people are walking around during our experiments. The mobility of people around will not affect the system as well. 3) Last but not least, the victim may move during the process of eavesdropping. Since the signal strength will decrease at the same time, it is not feasible for the adversary to compensate the system position for a better-received signal. If the victim's movement is small, we might develop a space-searching scheme to make a calibration.

11 CONCLUSION

In this paper, we propose MagEar, an eavesdropping system that exploits 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, which radiates a magnetic field that can be measured by a magnetic sensor to infer the audio being played. We establish a sound-to-magnetic model and validate the feasibility of our proposal for different types of speaker-embedded electric devices. Our evaluation demonstrates 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 uncovers a significant loophole in the security of commercial off-the-shelf (COTS) speaker-embedded devices, including headphones, smartphones, and smart speakers. We hope our work will encourage manufacturers to reconsider the security vulnerability of speakers.

ACKNOWLEDGMENTS

This research was supported in part by the China NSFC Grant U2001207, Guangdong Provincial Key Lab of Integrated Communication, Sensing and Computation for Ubiquitous Internet of Things, the Project of DEGP (No.2023KCXTD042). 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] Yushi Cheng, Xiaoyu Ji, Wenyuan Xu, Hao Pan, Zhuangdi Zhu, Chuang-Wen You, Yi-Chao Chen, and Lili Qiu. 2019. MagAttack: Guessing Application Launching and Operation via Smartphone. In *Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security* (Auckland, New Zealand) *(Asia CCS '19)*. Association for Computing Machinery, New York, NY, USA, 283–294. https://doi.org/10.1145/3321705.3329817

[11] 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.

[12] 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

[13] 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

[14] 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

[15] 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

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

[17] Hua Huang and Shan Lin. 2020. MET: A Magneto-Inductive Sensing Based Electric Toothbrushing Monitoring System. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking* (London, United Kingdom) *(MobiCom '20)*. Association for Computing Machinery, New York, NY, USA, Article 19, 14 pages. https://doi.org/10.1145/3372224.3380896

[18] 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).

[19] 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.

[20] Yongzhi Huang, Kaixin Chen, Lu Wang, Yinying Dong, Qianyi Huang, and Kaishun Wu. 2021. Lili: liquor quality monitoring based on light signals. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*. 256–268.

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

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

[23] 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)*.

[24] Qianru Liao, Yongzhi Huang, Yandao Huang, Yuheng Zhong, Huitong Jin, and Kaishun Wu. 2022. MagEar: eavesdropping via audio recovery using magnetic side channel. In *Proceedings of the 20th Annual International Conference on Mobile Systems, Applications and Services*. 371–383.

[25] 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.

[26] 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).

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

[28] Ralph P Muscatell. 1984. Laser microphone. US Patent 4,479,265.

[29] 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).

[30] 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

[31] John R Speciale. 2001. Pulsed laser microphone. US Patent 6,301,034.

[32] Cem Subakan, Mirco Ravanelli, Samuele Cornell, Mirko Bronzi, and Jianyuan Zhong. 2021. Attention is all you need in speech separation. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 21–25.

[33] Avery Wang et al. 2003. An industrial strength audio search algorithm.. In *Ismir*, Vol. 2003. Citeseer, 7–13.

[34] Chen-Chia Wang, Sudhir Trivedi, Feng Jin, V Swaminathan, Ponciano Rodriguez, and Narasimha S Prasad. 2009. High sensitivity pulsed laser vibrometer and its application as a laser microphone. *Applied Physics Letters* 94, 5 (2009), 051112.

[35] Guanhua Wang, Yongpan Zou, Zimu Zhou, Kaishun Wu, and Lionel M. Ni. 2014. We Can Hear You with Wi-Fi!. In *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking* (Maui, Hawaii, USA) *(MobiCom '14)*. Association for Computing Machinery, New York, NY, USA, 593–604. https://doi.org/10.1145/2639108.2639112

[36] 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

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

[38] 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).

[39] 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

[40] 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