Spoofing Eavesdroppers with Audio Misinformation

Zhambyl Shaikhanov
University of Maryland, College Park
zhambyl@umd.edu

Mahmoud Al-Madi
Rice University
mha6@rice.edu

Hou-Tong Chen
Los Alamos National Laboratory
chenht@lanl.gov

Chun-Chieh Chang
Los Alamos National Laboratory
cchang@lanl.gov

Sadhvikas Addamane
Sandia National Laboratories
saddama@sandia.gov

Daniel M. Mittleman
Brown University
daniel_mittleman@brown.edu

Edward Knightly
Rice University
knightly@rice.edu

***Abstract*—Wireless eavesdropping on phone conversations has become a major security and safety concern, especially with advancements toward 5G and beyond featuring higher frequencies and higher sensing resolution. As demonstrated recently, attackers can remotely detect even micron-scale acoustic vibrations emanating from a smartphone's earpiece via off-the-shelf millimeter-wave radar for audio information eavesdropping, all without the victim ever noticing. Here, we present a new architecture, MiSINFO, that not only thwarts such attacks but also enables the victim to counter-attack by spoofing of eavesdroppers with audio misinformation. With emerging attacks targeting the physical medium, i.e., acoustic signals, which cannot be protected by digital encryption and are the weakest segment of the communication chain, MiSINFO aims to systematically modify the eavesdroppers' fundamental sensing observations, concealing native signals while encoding alternate synthetic data. MiSINFO incorporates a low-profile, reconfigurable metasurface and double-inference principles to dynamically generate artificial audio-vibration signatures, injecting deceptive misinformation. We design, implement, and experimentally evaluate MiSINFO. Our results reveal that eavesdroppers detect none of the original words emitted by the speaker, while the injected misinformation is reconstructed with a low average word error rate of 2.29%.**

Our work represents the first such eavesdropping countermeasure which not only prevents attackers from accurately decoding the true signal but also uses a false signal to fool them into believing that they have succeeded. This approach transforms defensive measures from merely reactive to proactively deceptive, giving the defender an advantage and the capability to delude attackers into trusting false information.

1. Introduction

With over 6.9 billion smartphone users worldwide [1], sensitive information, including financial records, personal identifications, healthcare details, and classified business and government information, is communicated daily over the phone. However, the confidentiality of such audio information can be readily compromised with the emergence of inexpensive yet highly advanced high-frequency radar systems [2], [3], [4], [5], [6], [7]. Namely, attackers are empowered with new micron-resolution wireless sensing capabilities, allowing them to remotely acquire even minuscule byproduct information about the physical world with extreme accuracy, all without the victim ever noticing. That is, attackers can now remotely detect tiny vibrations of a phone's earpiece or speaker and reconstruct the underlying audio information, all without needing to install any malware on the smartphone or monitor voice traffic [2], [3].

In fact, the adoption of millimeter-wave frequencies (30 GHz - 300 GHz) in 5G and beyond marks a significant shift in wireless, with an order of magnitude increase in available bandwidth and a decrease in wavelengths that enable unprecedented wireless sensing capabilities, facilitating not only emerging applications such as interactive virtual reality systems and self-driving vehicles [8], [9], [10] but also bringing forth new security challenges of malicious side-channel information sensing with fine-grained resolution [4], [5], [6], [7]. Specifically, the eavesdropper, Eve, can direct a millimeter-wave radar system towards the victim, Alice, who is talking on her phone, to detect few-micron scale vibration patterns. With these tiny displacements being strongly correlated with the generated sound waves [2], [3], Eve recovers audio conversation, intercepting sensitive information. Unfortunately, this analog micro-information cannot be encrypted or protected by existing methods, as Eve deliberately exploits the physical propagation of sound to steal information during interactions between users and the analog world, in contrast to directly targeting sensitive data in the highly protected digital domain. In fact, due to the inherent physical medium, the acoustic signal is the weakest link in the communication chain of a phone call, making it very challenging to protect with existing methods.

In this paper, we present **MiSINFO**, a new class of proactive defensive architecture for audio misinformation security. We develop a novel eavesdropping countermeasure that not only prevents attackers from succeeding but also injects a false signal into their wireless sensing observations, enabling strategic misinformation capabilities for defenders. Unlike traditional approaches that obscure malicious sensing

signals, such as by randomizing or jamming, we design a radically stronger counter-method that makes attackers oblivious to the countermeasures while also spoofing [1] them with false information. This approach provides significant strategic and security advantages across various scenarios. For instance, attackers could be misled into intercepting fake business contracts and negotiations (corporate), false bank account details and social security numbers (civilian), or incorrect coordinates and mission details (military). In this work, we make the following three contributions.

First, we develop the key principles of MiSINFO. We study the foundations and implications of the attack and subsequently exploit these features to develop the basis of the proposed defensive mechanism. Specifically, we find that the overall attack is built upon a double inference process: Eve infers targeted acoustic signals from physical vibrations (due to the mechanical coupling between the smartphone speaker/earpiece and the phone case) while also inferring these micron-level vibrations from radar wireless signals. MiSINFO exploits Eve's double inference process and secretly modifies the characteristics of the wireless signal interacting with vibrations, thereby enabling subsequent alteration of all observations in the inference chain. Moreover, MiSINFO integrates on-phone dynamic metasurface - a surface structure with programmable electromagnetic (EM) properties - to proactively induce EM changes to Eve's radar chirp transmissions. As such, Alice can potentially emulate any targeted vibration pattern, in essence, taking control over Eve's fundamental sensing observations. We also propose several key schemes to manipulate Eve's sensing, including scrambling the phone conversation that Eve wishes to extract, as well as spoofing Eve with legitimate yet false audio information, completely misleading Eve.

Second, we realize MiSINFO by designing an experimental testbed with a dynamic programmable metasurface and investigating its key characteristics. In particular, we design a metasurface comprising an array of metallic sub-wavelength-sized split-ring resonators (SRRs) lithographically deposited on a GaAs substrate. The SRR array is constructed to form a Schottky contact with the GaAs, with applied reverse bias majorly changing the carrier density in the depletion region underneath the metallic elements, which modifies the split-gap conductivity and, thus, the electromagnetic response of the surface. Our prototyped metasurface, an integral part of the MiSINFO, comprises 18 programmable columns that are dynamically controlled by applying a tunable voltage in the range of -10 V to 0 V. To demonstrate the proposed architecture, we first characterize the metasurface by studying the phase response across different switching frequencies. We discover that it exhibits a low-pass response, providing Alice with a broader range of phase control at slower switching frequencies, while the range decreases at higher rates. Moreover, we establish the mapping between time-varying voltage control signals and phase responses, which is later employed to implement different defense strategies. Our results reveal that Alice can achieve very high-resolution, sub-micron control over displacement by operating in the non-saturated control signal region, enabling Alice to accurately modify Eve's sensing observations.

Third, we implement our proposed defense strategies and experimentally evaluate the system. To do so, we first create a set of different audio samples using deep learning-based algorithms that analyze human voice patterns and generate speech from text. Demonstrating the attack with Eve employing a Texas Instruments millimeter-wave radar and Alice implementing MiSINFO and yet initially in the off state (i.e., no control signals are applied to activate the metasurface), we show how Eve intercepts audio information emanating from Alice's smartphone speaker using the time-varying phase of the reflected radar signal and accurately recognizes all words with automatic speech recognition algorithms. However, as a strong baseline strategy, we demonstrate how Alice can reconfigure the metasurface with random voltage signals at audio sampling rates to scramble the phase of the signal reaching Eve, thereby disrupting the integrity of the audio information Eve aims to extract. Furthermore, we experimentally demonstrate how Alice generates and applies a temporal sequence of control signals that mimic the signature of legitimate and yet misleading information, thereby spoofing Eve with misinformation. We showcase this by injecting audio misinformation with purposefully altered sensitive details such as different personal identification, bank account numbers, and social security numbers, with all of the false information being accurately decoded by Eve as legitimate data. Finally, we conduct a large-scale experiment with more than a thousand words of audio. We demonstrate that MiSINFO not only thwarts such a devastating attack but also effectively spoofs the attacker, with Eve detecting none of the original information while accurately decoding misinformation with an average word error rate of 2.29%.

The remainder of this paper is organized as follows: Section 2 reviews related work, Section 3 describes the threat model, and Section 4 presents the MiSINFO design. Section 5 introduces the fabrication and implementation, while Section 6 describes the experimental results. Finally, Section 7 provides a discussion, and Section 8 concludes the paper.

2. Related Work

Eavesdropping on Smartphones. Given the ubiquity of smartphones and the vast amount of sensitive information communicated via these devices, attackers have been exploiting different methods, including internal smartphone sensors such as accelerometers and gyroscopes, sensitive to acoustic signals, to eavesdrop on the phone speech [11], [12], [13], [14], [15]. While devastating, such techniques often require the installation of malware on phones to acquire pre-collected data and necessitate training for specific smartphone user-victims for accurate speech reconstruction. With the recent advancement towards 5G and beyond, the

1. In the paper, "spoofing" refers to deceiving the attacker with an alternate audio inference vs. the true one emitting from the speaker.

attackers have been shown to exploit high-frequency technologies such as millimeter-wave radars to remotely eavesdrop on phone conversations [2], [3], utilizing side-channel vibration information [4], [5], [6], [7], [16]. As a result, these attacks are devastating as well as extremely difficult to detect, as attackers do not require physical access to the phone to install malware, nor do they need to train user-specific models to accurately reconstruct intercepted speech. Unlike prior work, in this paper, we present the first countermeasure against such audio-vibration eavesdropping, which not only thwarts the attacks but also spoofs the attackers with misinformation. Our short, 2-page report-paper [17] outlines a similar roadmap, but it neither has a full system design nor evaluation.

Metasurface Wireless Security. Metasurfaces are two-dimensional structures capable of manipulating electromagnetic waves in controlled ways, enabling various applications, including in wireless communication, sensing, and security [18], [19], [20], [21], [22], [23], [24], [25], [26]. Recently, switchable metasurfaces operating in the millimeter-wave spectral ranges have become a focus of considerable research [27]. A common goal of this body of research has been to enhance the switching speed of such devices, for example by reducing the area of the surface in order to minimize its RC time constant [28], [29]. This scaling can enable applications in data modulation [30] or high-speed wavefront manipulation [31]. However, in this paper, we exploit the unique properties of switchable metasurfaces and realize the first smartphone misinformation security capability. In fact, the metasurface in our design requires relatively low-bandwidth (∼kHz) operation, and can therefore exploit a large (∼cm^2) surface area. Given the size and power constraints, our approach could readily be integrated with commercial systems using methods analogous to those employed for ultra–compact cameras [32], [33] or versatile antenna [34], [35].

3. Threat Model

In this section, we first discuss the attack scenario and then Eve's principles for audio-vibration eavesdropping.

3.1. Audio-induced Vibrations on Phones

We consider a general attack scenario where Alice communicates confidential information while speaking on her smartphone. Meanwhile, an eavesdropper, Eve, aims to intercept that sensitive phone conversation. Typical to many cases, we consider that Eve does not have access to Alice's smartphone, for instance, to install malware and secretly control the device, nor can she physically stand next to Alice to overhear the conversation, thereby exposing herself.

In contrast, Eve exploits side-channel information to carry out the audio eavesdropping attack. Specifically, she takes advantage of the mechanical coupling between the smartphone speaker/earpiece and the phone case, which induces a vibration pattern $d(t)$ that is strongly correlated with the emitted sound waves from the device. That is,

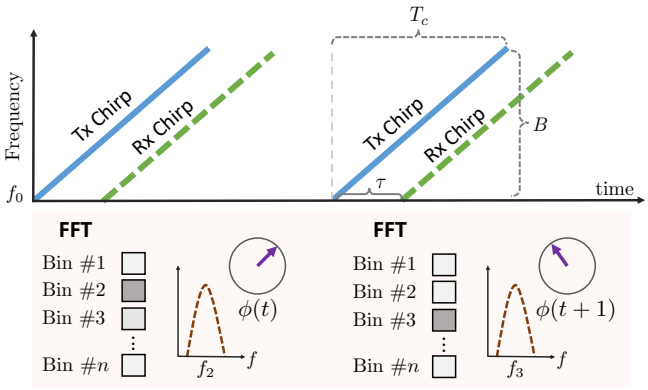


Figure 1: The remote attacker's audio-vibration sensing principles via FMCW millimeter-wave radar chirps.

$a(t)$ represents the signal driving Alice's speaker, which triggers the speaker to vibrate and generate audio waves that Alice hears. Then, the audio-induced subtle physical displacements $\eta(f_a)$ on the speaker diaphragm [36] as a function of acoustic frequency f_a can be represented as

$$\eta(f_a) = \frac{e_g}{2\pi f_r B l Q_{es}} |\gamma(f_a)| \tag{1}$$

in which e_g indicates the voltage at the speaker's terminals, B is the magnetic flux density, l is the voice coil length, Q_{es} denotes electrical damping, f_r is the speaker's resonance frequency, and $\gamma(f_a)$ is a dimensionless frequency response function represented as

$$\gamma(f_a) = 1/[1 - \frac{f_a}{f_r^2} + j\frac{1}{Q_{ts}} \cdot \frac{f_a}{f_r}] \tag{2}$$

in which Q_{ts} indicates the total damping effect, including the electrical damping Q_{es} and the mechanical damping Q_{ms}. Given that the vibration exhibits typical harmonic motion, the sound-induced time-varying displacement $d(t)$ can then be expressed as $d(t) = \eta(f_a)\cos(2\pi f_a t)$. Eve targets to infer $a(t)$ via sensing these vibrations. For that, Eve employs commercially available portable radar, directing the transmission beam towards Alice's smartphone from a distance. This allows her to remotely sense the vibration patterns and recover the underlying audio information. Exemplary demonstrations of the threat model we consider here include [2], [3].

3.2. Eve with COTS Millimeter-wave Radar

Eve repurposes a frequency-modulated continuous wave (FMCW) millimeter-wave radar, commonly employed in automotive and industrial applications, into a vibration-sensing system for the attack. She transmits a series of FMCW signals, known as chirps, directed toward Alice's smartphone. With the physical displacements encoded onto the chirp signals, she processes the reflected signals to recover the vibrations and, subsequently, audio information.

As depicted in Figure 1, the frequency of the chirp signal linearly increases with time, commencing from f_0

and spanning a bandwidth of B within a duration of a chirp T_c. Then, transmitted and received chirps at time t can be shown as

$$\begin{aligned} S_{\text{Tx}}(t) &= A_{Tx}\cos[\ 2\pi f_{Tx}(t)t + \phi_{Tx}] \\ S_{\text{Rx}}(t) &= A_{Rx}\cos[\ 2\pi f_{Rx}(t)t + \phi_{Rx}] \end{aligned} \tag{3}$$

where A_{Tx} and A_{Rx} indicate the amplitude of transmitted and the received chirps, respectively, $f_{Tx}(t)$ and $f_{Rx}(t)$ designate the frequencies of transmitted and the received chirps at time t, while ϕ_{Tx} and ϕ_{Rx} are the phases of transmitted and the received chirps, respectively. With Eve's radar r distance away from Alice's smartphone, the reflected chirp frequency at the receiver $f_{Rx}(t) = f_{Tx}(t-\tau)$ with a round trip delay $\tau = 2r/c$ where c is the speed of the radar signal and $f_{Tx}(t) = f_0 + kt$ with k indicating the slope of the chirp signal.

Further combining the transmitted and received signals at the radar transceiver by a mixer and applying a low-pass filter to focus on the frequency difference of the two carrier signals, Eve generates a beat signal $S_b(t)$. Such beat signal can be formulated as

$$S_b(t) = \frac{A_{Tx}A_{Rx}}{2}\cos[\ 2\pi(f_{Tx} - f_{Rx})t + (\phi_{Tx} - \phi_{Rx})] \tag{4}$$

with corresponding beat frequency and phase information.

Traditionally, a range FFT operation is performed on $S_b(t)$ to extract beat frequency f_b information and range the object. In particular, the distance between the radar and the object can be expressed as

$$r = \frac{f_b}{k}\frac{c}{2}. \tag{5}$$

However, f_b is insufficient to recover the micron-level vibration patterns. That is, the ranging resolution of the radar is limited to $r_{res} = \frac{c}{2B}$ and primarily governed by the bandwidth B swept by the chirp [37]. For instance, with a typical millimeter-wave radar of a chirp bandwidth 4 GHz, the ranging resolution corresponds to only several centimeters, several orders below targeted resolutions.

3.3. Micron-scale Audio-vibration Sensing

In addition to the beat frequency f_b component, the beat signal S_b also contains phase information as formulated in Equation 4. Eve exploits the beat phase details to sense such minuscule vibration patterns.

In particular, Eve performs Doppler FFT operation on beat signal to detect phase changes caused by the speaker vibrations $d(t)$. Such a relation can be formulated as

$$\phi(t) = \frac{4\pi d(t)}{\lambda} \tag{6}$$

in which λ denotes the free-space wavelength of the employed radar signal. Although the phase of chirp signals is susceptible to noise in the radar hardware implementation, Eve uses standard data processing techniques to readily address such challenges.

Eve addresses common issues arising from hardware-induced phase discontinuities such as frame reset artifacts via polynomial interpolation to ensure smooth phase transitions across frames. Additionally, phase wrapping can distort signal interpretation, but Eve corrects this using phase unwrapping techniques to reconstruct the true phase evolution over time. Furthermore, radar systems often introduce unwanted low-frequency noise that can obscure fine-grained phase variations crucial for accurate signal extraction. To counteract this, Eve applies bandpass filtering to isolate the relevant frequency components, enhancing the clarity and reliability of the recovered phase information. These techniques, as demonstrated in prior work [16], enable Eve to reconstruct phase-modulated signals with high precision, underscoring the practicality of such attacks despite hardware imperfections. Additionally, more sophisticated Eves have been shown to utilize machine learning algorithms to better recover phase information from noise, e.g., as demonstrated in [3].

Importantly, due to the high frequency and smaller wavelength of the employed radar signals, the phase changes enable Eve to dynamically track minuscule vibrations, achieving a resolution better than $\lambda/100$. For instance, using a typical millimeter-wave radar, she can detect a phase change as small as approximately 1°, corresponding to an acoustic-vibration displacement of about 0.005 mm for a radar frequency of 77 GHz. Such fine-grained resolution sensing allows the attackers to accurately detect audio-vibration patterns and recover targeted audio information $a(t)$ with high precision, as demonstrated in prior works and experimentally evaluated in Section 6 in this paper.

4. MiSINFO Architecture

In this section, we first discuss the principles of the proposed defense mechanism and present MiSINFO architecture model that enables spoofing of the attackers with audio misinformation. Next, we describe the design of our system, including the emulation of speaker-scale vibrations and their reconfiguration at audio sampling rates. Finally, we present new strategies for generating audio misinformation as well as random acoustic signatures.

4.1. Principles and System Model

The key principles of our proposed defense mechanism stem from several insights about the attack and its implications. These observations provide the foundational basis for the design of our MiSINFO architecture.

When Eve intercepts the phone conversation, it is important to note that she does not directly access the source of the information—the actual acoustic signals emanating from Alice's smartphone. Instead, she infers it from the sound-induced mechanical displacement on the smartphone's earpiece and speakers, as discussed in Section 3.3. Moreover, Eve does not visually see or otherwise physically feel these induced mechanical vibrations either. In fact, doing so would be infeasible because a) such vibrations occur on a tiny

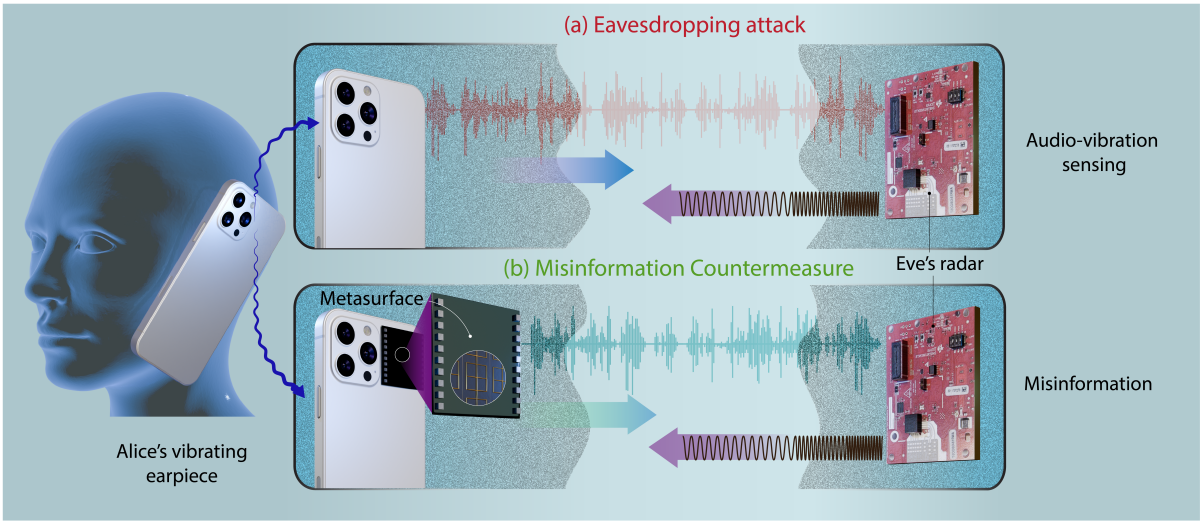


Figure 2: Schematic of the proposed MiSINFO architecture: (a) Malicious Eve directs a radar signal toward the victim, Alice, who is talking on her phone, to sense micron-scale vibrations and recover sensitive audio conversations, as demonstrated in prior works. (b) We develop and demonstrate the first system that enables Alice to dynamically and systematically modify Eve's malicious sensing observations, inducing targeted false signatures to spoof Eve's sensing observations and thereby mislead Eve with audio misinformation.

scale, at the micron level, making them extremely difficult for her to perceive. In addition to that, b) Eve carries out a remote attack to avoid detection, potentially from tens of meters away from Alice's smartphone, as shown in the prior works, which would even further complicate the challenge presented by a). In contrast, the attacker employs wireless radar signals to also infer these minuscule vibrations. Then, the overall attack is built upon a double inference process, i.e., inferring acoustic signals from mechanical vibrations and inferring vibrations from wireless signals, to realize the secret eavesdropping attack on the phone call.

In the MiSINFO design, we leverage the attacker's inference processes to develop the defense mechanism. Specifically, we propose to modify the characteristics of the wireless signal interacting with vibrations, thereby subsequently altering all observations in the inference chain. That is, a modified wireless signature would imply a different vibration pattern than the original one, which in turn would correspond to different audio information when decoded by Eve. Moreover, we propose to strategically induce these changes so that Alice can inject her targeted legitimate yet false information, which we refer to as misinformation, to Eve. This enables Alice to not only protect her sensitive data, but to also convey fake, misleading data to Eve.

We illustrate the schematic of our proposed architecture in Figure 2. In Figure 2(a), we first show how Eve remotely senses vibration patterns reflecting off Alice's smartphone and reconstructs audio data from that sensory information. With the proposed MiSINFO to thwart the attack, Alice enables it during her phone conversation and controllably alters Eve's fundamental sensing observations as depicted in Figure 2(b). We introduce new sensing security principles that enable Alice to generate and inject audio misinformation $m(t)$, selected by Alice. Thus, Alice can vibrate her speaker with the true $a(t)$ such that she has a live interactive audio conversation while simultaneously conveying misinformation to Eve, effectively deceiving Eve into believing she has succeeded.

4.2. Emulating Vibrations with a Metasurface

As discussed in Section 3 and formulated in Equation (6), Eve utilizes the phase information $\phi(t)$ of the millimeter-wave radar chirps to extract the audio-vibration patterns. Our MiSINFO system enables Alice to proactively alter these phase changes, in essence taking control over Eve's fundamental sensing observations. For that, Alice employs a smartphone-mounted dynamic metasurface with a dynamically programmable electromagnetic response and then emulates her targeted vibration signatures.

Metasurfaces are 2D structures composed of an array of subwavelength metallic structures. They are engineered to manipulate electromagnetic waves in controllable ways, even surpassing the capabilities found in nature [38]. They have facilitated a multitude of applications across various domains and frequencies [38], [39], [40]. With the advancement towards higher frequencies, such as millimeter-wave in 5G and beyond, metasurfaces are shown to be particularly valuable, especially for compact devices like smartphones. They facilitate the miniaturization and versatility of antenna systems and other components on phones by manipulating electromagnetic waves at subwavelength scale [33], [34]. In

the MiSINFO design, we exploit an on-phone metasurface to realize novel security capability. We discuss our smartphone-mounted metasurface design in Section 5.1.

In order to emulate acoustic vibrations and deceive Eve, Alice applies a control voltage signal $\vec{V}$ to the metasurface device, dynamically reconfiguring it. In response, the metasurface induces targeted phase changes at the surface interface. However, before controllably reconfiguring it, Alice needs to first characterize her metasurface to understand its excited electromagnetic properties. To realize it, she constructs a mapping defines as

$$f(\vec{V}_i) = \phi_i, \text{ for } i = 1, 2, ..., n \tag{7}$$

which establishes the correspondence between each applied control signal $\vec{V}_i$ and the resulting phase response ϕ_i. Alice then employs pre-characterized metasurface with the mapping information to generate audio signatures discussed in Section 4.3. We also present the details on the characterization of our metasurface in Section 6.1.

4.3. Programming Acoustic Signatures

Having characterized her metasurface, Alice can spoof the attacker, Eve, with potentially any audio misinformation $m(t)$, as each audio signal has a unique phase response ϕ^m when observed via a radar. To achieve this, she can determine the metasurface control signal $\vec{V}^*(t)$ based on the pre-characterized properties as

$$\vec{V}^*(t) = \underset{\vec{V}(t)\in\vee}{\text{argmin}} \; |f(\vec{V}(t)) - \phi^m(t)| \tag{8}$$

in which $\vee$ denotes the set of feasible input control signals. In response to the applied control signal, the on-phone metasurface can then manipulate the electromagnetic waves of Eve's chirp signals to produce the targeted phase response. Yet, considering Alice with a transmissive on-phone metasurface (as described in Section 5.1), Eve's radar chirp signals will propagate through the metasurface while also interacting with the physical vibrations on the phone. Then, Eve's observation $\phi^*(t)$ will effectively be the cumulative impact of both phenomena, along with some noise ϵ

$$\phi^*(t) = \phi^a(t) + \phi^m(t) + \epsilon. \tag{9}$$

Then Eve's perceived vibration pattern can be expressed as

$$d^*(t) = \frac{\phi^*(t)\lambda}{4\pi}. \tag{10}$$

Evident from Equations (9) and (10), the true vibration pattern from the smartphone speaker could be revealed, with Eve reconstructing audio with both the actual information $a(t)$ and the injected misinformation $m(t)$ overlaid on each other. However, we note that the dynamic phase shifts programmably induced on the radar electromagnetic wave by the metasurface can be notably larger (and different) than that induced by the mechanical vibration of the phone. Consequently, this artificially injected signal overwhelms the actual speech signal that Eve seeks to detect, with Eve perceiving and reconstructing an alternate audio stream. We experimentally demonstrate the high efficacy of Alice's ability to spoof Eve in Section 6.

Furthermore, we highlight that Alice has a large design space for constructing audio misinformation $m(t)$ in the MiSINFO design. For instance, as a strong baseline strategy, she could generate random audio. That is, Alice can induce a random time-varying voltage signal at the metasurface interface to randomize the phase of the signal reaching Eve, thereby disrupting the integrity of the audio information Eve wishes to extract. We demonstrate this strategy in Section 6.2.

Beyond obscuring her conversation, Alice can implement our proposed novel defensive security strategy to actively spoof the eavesdropper with false audio information. In particular, she can generate a temporal sequence of metasurface control signals to imitate a vibrational pattern which is not random, but which instead corresponds to a speech signal that is different from the one that Alice is actually speaking. Because the phase shift induced on the radar electromagnetic wave by the metasurface can be larger than that induced by the mechanical vibration of the phone, this artificially injected signal can overwhelm the actual speech signal. As a result, Eve perceives and reconstructs an alternate audio stream. We demonstrate this strategy in Section 6.3.

Finally, we remark that with MiSINFO, Alice can continue to have a normal conversation (with Bob) without altering the original $a(t)$ and instead modifies only $m(t)$ to spoof the attacker. Then, there are a few ways for Alice to generate $m(t)$. First, she can pre-record a false message, with Eve decoding it as legitimate one, as we demonstrated in this paper. Second, Alice can dynamically select from a set of pre-recorded messages, which essentially extends the approach in the first method. Thirdly, she can employ a dynamic AI engine to learn and change sensitive information, e.g., dates, numbers, and names, in real-time. Such an approach is beyond the scope of this paper but viable with a strong Alice.

4.4. Reconfiguring at Audio Sampling Rates

Another essential aspect of the MiSINFO is the switching frequency of the metasurface, i.e., how quickly Alice should reconfigure the on-phone structure with corresponding control signals to generate audio misinformation that Eve can decode effectively. Such a design choice is largely governed by the key features of human speech as well as Eve's ability to recover and process audio vibration signals.

Human speech is a complex auditory system, with the lungs producing periodic air pressure that passes through the vocal cords and later through the throat, mouth, and nasal cavity, modulated throughout each of those steps to form fine-tuned speech. The frequency content of the speech then depends on the sounds being produced, and the voiced and unvoiced sounds are the main two categories [41].

Voiced sounds are mainly characterized by the vibration of the vocal cords, resulting in patterns of compression

and rarefaction in the air. However, unvoiced sounds are created without significant vibration of the vocal cords and instead, formed by the passage of air through the vocal tract, causing turbulence or frication at specific points of articulation. All vowel sounds and certain consonants such as 'b', 'd', and 'g' are voiced, while unvoiced sounds are associated with consonants like 's', 'p', 'k', 'sh', and 'th'. In voiced sounds, the energy is typically concentrated in lower frequencies, whereas in unvoiced sounds, energy is more evenly distributed across frequency bands. In general, the bandwidth of 80 Hz to 1.7 kHz contains key components of human speech: 100% of vowels' fundamental frequencies, 62.5% of consonants' fundamental frequencies, and 68.8% of vowels' second harmonics [2], [42], [43].

Yet, with Eve reconstructing these audio signals from the vibration observations, the achievable highest frequency content depends on her radar's sampling rate. According to the Nyquist theorem, the sampling frequency of Eve's radar chirps must be at least twice the maximum frequency she aims to recover in the signal. In addition to the sampling rate, the SNR of her received signal is equally important to Eve to recover the speech with high accuracy. The SNR of her received signal can be expressed as [37]

$$SNR = \frac{\lambda^2 G_{Tx} G_{Rx} \alpha}{(4\pi)^3 r^4 F} \tag{11}$$

in which G_{Tx} and G_{Rx} are the transmitter and receiver gains of the employed radar, respectively, α represents the radar hardware configuration coefficient, and F denotes the noise floor of the sensor.

Considering the aforementioned factors, Alice should clearly avoid configuring the metasurface at low frequencies, i.e., below a kHz, as it would poorly emulate the audio vibrations and negatively impact the clarity of her injected misinformation data. On the other hand, typical metasurfaces can switch far beyond the kHz scale. However, if possible, Eve should also avoid extreme frequencies, such as MHz scale or even hundreds of kHz, as this might introduce unnecessary redundancy to the data (potentially exploitable by a sophisticated Eve with very high SNR and an unusually high sampling rate). In this paper, we demonstrate the high efficacy of the proposed defense mechanism with a few kHz switching rates, as shown in Section 6.

5. Fabrication and Implementation

In this section, we first discuss the design and fabrication of our metasurface. Next, we discuss the metric for quantifying the injected misinformation to Eve, along with the automatic speech recognition system for detecting the words in the recovered audio. Then, we describe our experimental setup, including the millimeter-wave radar system.

5.1. Programmable Metasurface Design

We realize MiSINFO by designing an experimental testbed with a compact 2.5 cm $\times$ 2 cm reconfigurable metasurface. Our metasurface, show in Figure 3, comprises an array of metallic split-ring resonators (SRRs). They are lithographically deposited on a $2\mu m$ thick doped Gallium Arsenide (GaAs) epilayer grown on a semi-insulating GaAs substrate. The period, SRR size, gap, and width are 210 μm, 160 μm, 2 μm, and 6 μm, respectively. The metallic SRR array is designed to form a Schottky contact with the GaAs, such that a reverse bias applied to the array produces a significant change in the carrier density in the depletion region underneath the metallic elements. This enables the modification of the split-gap conductivity and, thereby, also the electromagnetic response of the surface [44]. That is, the switching mechanism of our metasurface stems from the charging/discharging of an effective capacitor formed by the Schottky contact. The surface area of the device determines the effective capacitance and, thus, the RC time constant. Thus, a larger surface area corresponds to a slower response time.

Our metasurface prototype is transmissive and comprises 18 programmable columns that are dynamically controlled by applying DC voltage bias via a waveform generator. Designed for millimeter-wave frequencies (with strong resonant response at a center frequency of 150 GHz with around 20 GHz bandwidth), it manifests a phase response over a much broader frequency range [45], [46], extending to frequencies below the $77 - 81$ GHz frequency band of the commonly employed radar system. By varying the bias applied to the metasurface, we can control the phase imposed on the radar signal that interacts with it, thus modifying Eve's observations (which are based on analysis of the phase of the returned radar signature).

To randomize the eavesdropper's observation as demonstrated in Section 6.2, we scale the white noise signal to an amplitude of 6.2 and center it approximately 3.1, as per the range of our phase modulation. We then map these values onto the corresponding voltage values to four decimal places in accordance with a polynomial best-fit line shown in Figure 6. The resulting voltage waveform, ranging from -7.1 V to 0 V and centered around -3.55 V, is then dynamically configured through the metasurface at a

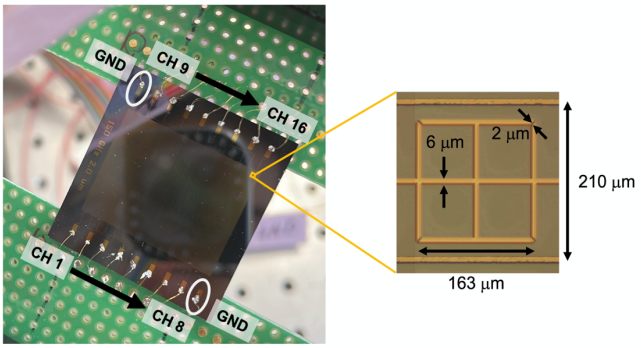


Figure 3: Custom-designed MiSINFO dynamic programmable metasurface, comprising metallic split-ring resonators (SRRs) on GaAs substrate.

frequency of 4 kHz. Similarly, we construct the misinformation waveform by scaling the targeted misinformation audio signal according to our phase range, subsequently mapping it onto the corresponding voltage values, and controlling voltage signals of the metasurface over time at the audio sampling rate.

5.2. Misinformation Metric

To demonstrate the principles of MiSINFO, we create various audio samples in the paper using a deep learning-based system. We leverage it to analyze human voice patterns and converts text into speech using synthetic voices [47]. Subsequently, we employ Amazon Transcribe's machine learning algorithms to process Eve's reconstructed audio and obtain transcribed text [48], which we analyze further.

To evaluate the efficacy of misinformation injection, we quantify the ratio of errors in Eve's transcript to the total words in the misinformation audio. Specifically, we compute the word error rate (WER) defined as:

$$\text{WER} = \frac{N^{\text{Substitutions}} + N^{\text{Insertions}} + N^{\text{Deletions}}}{N^{\text{Total Words in Misinformation}}}, \qquad (12)$$

in which $N^{\text{Substitutions}}$ indicates the instances where the transcribed word differs from the word in the reference text, $N^{\text{Deletions}}$ represents null transcription results for a word in the reference text, and $N^{\text{Insertions}}$ is additional transcribed words without corresponding words in the reference text. Note that in the WER calculation, misinformation words are not in general connected to the words emitting from the speaker; rather, WER quantifies Alice's effectiveness in injecting misinformation words. Overall, a lower WER indicates improved accuracy for Eve in recognizing misinformation speech and thus quantifies Alice's effectiveness in misleading the attacker, which we demonstrate in Section 6.2-6.4.

5.3. Experimental Testbed and Setup

We prototype and experimentally demonstrate the principles of our proposed system by positioning a smartphone 0.5 meters away from the radar, directly in front of the radar's receiver and transmitter antenna array. Our metasurface is placed in front of the phone, as close as physically possible within the constraints of the frame and wires of the surface.

Experiments are conducted using a commercially available radar system (Texas Instruments AWR1843BOOST) radar operating in the frequency range of 77 GHz to 81 GHz as shown in Figure 4. It has a horizontal 3 dB-beamwidth of approximately ± 28 degrees and an elevation 3 dB-beamwidth of approximately ± 14 degrees. The radar is physically connected to a real-time data capture adaptor (Texas Instruments DCA1000EVM), which transmits the data to a computing unit for further processing and analysis. Eve transmits chirps in frames, each consisting of 128 chirps, at a sampling rate of 10 kHz, and analyzes

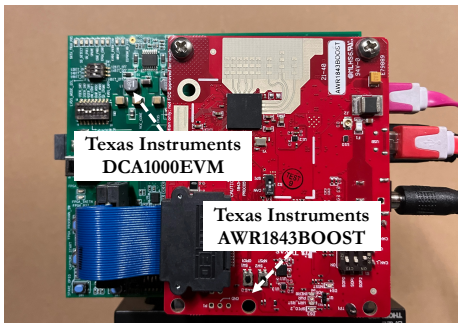


Figure 4: Commercially available low-cost portable millimeter-wave radar employed by the eavesdropper.

the received chirps as they reflect off the smartphone. The originally transmitted chirps are mixed with the received ones at the radar transceiver to produce a beat signal.

Once raw ADC data from the radar are collected and converted to an IQ array, the Fourier transform of the beat signal is performed to identify the beat signal frequency. Then, the phase of that frequency component is extracted to compute the vibration displacement at a given instant. Due to the radar's frame-based chirp transmission and the frame reset process, the radar hardware causes artifacts in the vibration measurements, manifesting as random spikes every 128 readings. To address this, cubic interpolation is performed to replace the artifact at the beginning of each frame. Moreover, the phase resets at the start of each frame, causing undesired phase drift. To correct this drift, polynomial regression is applied. This regression subtracts the hardware-induced phase shift from the original signal, effectively detrending the phase drift over time. Finally, a bandpass filter is applied to eliminate low-frequency noise and harmonic artifactsand then the processed vibrations waveform is converted into an audio waveform.

6. Experimental Evaluation

In this section, we perform a set of experiments to demonstrate the proposed MiSINFO system. First, we conduct experiments to characterize Alice's on-phone metasurface, focusing on both metasurface switching frequency and control signal to phase shift mapping. Next, we show MiSINFO strategy to scramble Alice's conversation, which Eve wishes to extract. Then, we demonstrate how Alice actively spoofs Eve with false audio information. Finally, we perform large-scale experiments with over a thousand words generated by the MiSINFO.

6.1. Alice's Metasurface Characterization

To implement the proposed defense mechanism, Alice must first characterize her on-phone metasurface, as described in Section 4. Here, we demonstrate how Alice acquires the metasurface phase response across different switching frequencies and establishes the mapping between control signals and phase responses. With this information,

she can then dynamically reconfigure the structure to emulate acoustic-vibration patterns.

Metasurface switching frequency: Different metasurface designs and hardware will produce distinct electromagnetic responses across switching frequencies. Employing the metasurface described in Section 5.1, we first conduct experiments investigating the phase response of the metasurface across different switching speeds. For that, we utilize the experimental setup detailed in Section 5.3, positioning Eve's smartphone-metasurface and Alice's radar at a distance of half a meter. We then apply a square-wave control signal (with a peak-to-peak voltage of 10 V and an offset of -5 V) to the phone-metasurface at varying switching frequencies, while keeping the phone silent, i.e., no ongoing phone conversation during this experiment. Then we analyze Eve's chirp signals to extract the resulting phase shifts.

We present the results in Figure 5. The x-axis represents the switching frequency of the metasurface, starting from 1 Hz and increasing to 5 kHz, while the y-axis displays the corresponding phase shifts in degrees. We also analyze the temporal response at exemplary frequencies, depicted in insets on both sides of the figure. The inset on the left-hand side provides a detailed view of the results at a switching frequency of 100 Hz, whereas the inset on the right-hand side zooms in on the results at 4 kHz.

Importantly, notice that the phase shift decreases with increasing switching frequency, and this pattern is nonlinear. This phenomenon arises due to the collective RC response of the split-ring resonator elements described in Section 5.1. In particular, rapid voltage transitions lead to shorter charging and discharging duration for the capacitor, providing a smaller achievable phase range. Such response suggests Alice has a broader range of phase control at slower switching frequencies, exceeding 10° below 100 Hz, while the range decreases at higher switching rates. We select a 4 kHz switching frequency that allows both sufficient phase

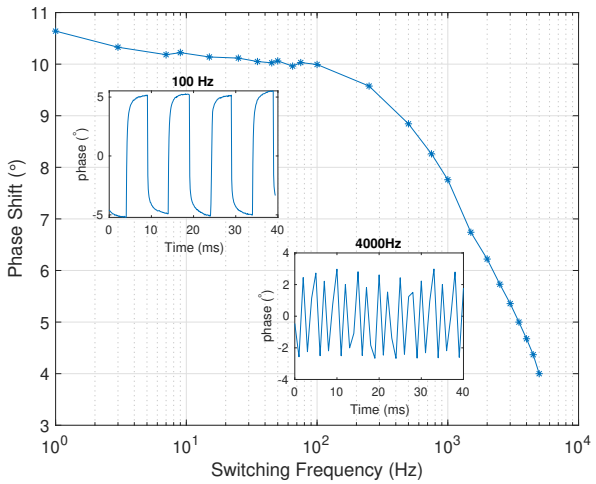


Figure 5: Measured phase change by Eve as a function of the switching speed of the signal applied by Alice to the metasurface.

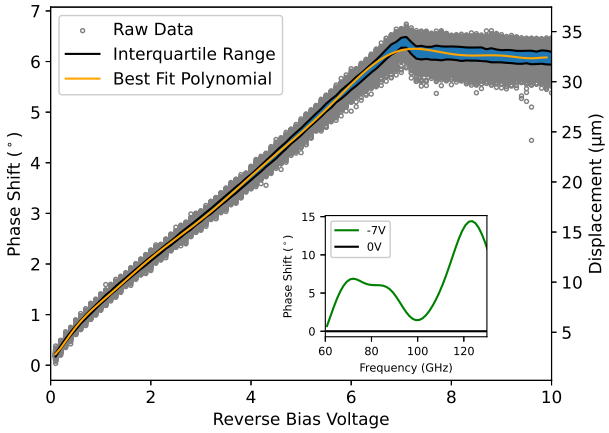


Figure 6: Characterization of Alice's metasurface through measured control signal response mapping.

range and bandwidth to mimic the richness and comprehensibility of voiced speech as demonstrated in Sections 6.2 and 6.3.

Control signal vs. phase shift mapping: Then, Alice needs to characterize the mapping between her control signal and the corresponding phase response to generate a targeted vibration pattern. To study it, we adopt the previously discussed setup and conduct experiments in which we vary the magnitude of the applied control voltages on the metasurface from -10 V to 0 V, maintaining a fixed switching frequency.

The results are presented in Figure 6, with the x-axis showing the applied reverse bias voltage, from 0 V to 10 V in steps of 0.1 V. Induced phase shifts are displayed on the left-hand of the y-axis while the right-hand axis shows the size of this phase shift converted to effective equivalent displacement, as per Equation (6). The yellow curve in the figure represents the best polynomial fit, the grey dots depict the recorded phase shift readings, and the blue region indicates the interquartile range. The inset shows the measured phase shift induced by the metasurface across a wide frequency range, spanning the $77-81$ GHz band of the radar system.

Focusing on the key yellow curve, note that it is nearly linear for approximately the first three quarters and then levels off. This indicates that Alice can induce a unique phase shift and displacement by operating within the range of control signals from -7.1 V to 0 V. However, we also find that at larger reverse bias values, the phase shift no longer changes. It is due to saturation of the growth of the induced depletion region at the SRR Schottky contacts [44]. As such, the -10 V to -7 V control signals region is considered less useful for Alice due to the many-to-one mapping.

Furthermore, the results reveal that Alice has extremely fine control over the displacement resolution, down to submicrometer scales, as illustrated on the right-hand axis in Figure 6. As such, Alice can accurately reproduce even complex vibration waveforms while also operating at an audio sampling rate. Together with this characterization of

the bias-induced phase shift at 77 GHz, we also measure the corresponding amplitude shift. However, since Eve employs phase demodulation as discussed in Section 4.3, this data is less relevant to this security situation.

Finding: The on-phone metasurface exhibits a low-pass response, providing Alice with a broader range of phase control at slower switching frequencies while the range decreases at higher rates. Despite having only a few degrees of phase range, Alice can achieve very high-resolution, sub-micron control over displacement by operating in the non-saturated control signal region and configuring the metasurface at an audio sampling rate of 4 *kHz.*

6.2. Generating Random Information

Thus far, we have demonstrated how Alice characterizes her on-phone metasurface. Next, we study Alice's strategies to counter the eavesdropping attack. First, we show how Alice scrambles Eve's sensing measurements, inducing a random time-varying voltage signal at the metasurface interface to randomize the phase of the signal reaching Eve and thus disrupting the integrity of the audio information she wishes to extract.

Adopting the previous experimental setup, we conduct experiments by configuring a control signal to a random voltage between -7.1 V and 0 V (uniformly distributed) at a rate of 4 kHz. Simultaneously, we play a 12 sec exemplary audio signal through the phone's earpiece speaker with the phrase "*Hi, I'm James Kelly. Sure, my bank account is 1, 2, 3, 5, 6, 7, and my social security number is 8, 9, 10.*" The audio (see **Audio 1**) is recorded at a high 48 kHz sampling rate and mimics a sensitive phone conversation with personal identification and numbers. We refer to this audio signal as the source audio information.

The results are depicted in Figure 7, and the source audio waveform is shown in Figure 7(a). As a baseline, we consider the scenario where Alice's metasurface is turned off, meaning no control signals are applied to activate the structure. Then, remote Eve aims to intercept the information using the time-varying phase of the reflected radar signal. Figure 7(b) illustrates Eve's observation in this baseline scenario (see **Audio 2**), shown as the measured phase shift in each time bin, converted to vibration displacement using Equation (6). The text above each snippet of the audio waveform indicates the corresponding word from the Amazon automated speech recognition audio-to-text transcription, as described in Section 5.2. Figure 7(c) displays Eve's measurements with the proposed defense mechanism implemented

First, we discover that Eve is indeed capable of replicating the audio information pattern by observing vibration patterns (see Audio 2). Although she samples displacements at a much lower chirp rate than the original audio source rate, resulting in differences in waveform shapes between Figure 7(a) and Figure 7(b), the subsequent audio-to-text transcription shows that Eve can accurately recognize every single word from the eavesdropped audio information, emphasizing the severity of the threat.

However, the results reveals that our design enables Alice to thwart the attack using randomly switching metasurface control signals. In particular, Figure 7(c) shows that the random phase shifts induced by the metasurface produce an arbitrary vibration displacement pattern which completely overwhelms the (smaller) signal containing the actual audio information. Consequently, Eve obtains only a noisy random waveform (see **Audio 3**), several times larger in amplitude and induced at the audio sampling rate. With such a defense mechanism in place, Eve fails to recover any audio information in the attack.

Finding: Eve can indeed remotely sense micron-scale mechanical displacements to intercept audio information, accurately recovering all the words in the eavesdropped phone call. However, with our proposed defense mechanism, in which Alice configures the metasurface with random voltage signals to scramble radar phase measurements, Eve fails to recover any of the words and observes only noise.

6.3. Generating Audio Misinformation

Until now, we have shown that MiSINFO can thwart the attack by obscuring Alice's conversation. Furthermore, here we demonstrate how MiSINFO enables actively spoofing the eavesdropper with false audio information. Unlike previously, Alice here generates a temporal sequence of metasurface control signals $\vec{V}_m(t)$ that can imitate a vibrational pattern corresponding to misinformation $m(t)$ containing different sensitive data from what Alice is actually speaking. This not only prevents Eve from accurately decoding the true signal but also injects a false acoustic signal.

Building on the previous experimental setup, we first create exemplary audio misinformation (see **Audio 4**) with the phrase "*Hi, I'm John Wick. Sure, my bank account is 7, 7, 3, 8, 0, 1, and my social security number is 3, 5, 9.*" This is similar to the audio information signal discussed above (see **Audio 1**), but with the name and sensitive numbers altered. We then create this waveform utilizing the control signal characterization depicted in Figure 6 to establish the mapping between the metasurface activation voltages and their associated excited properties. We then translate the temporal audio signal's amplitude information into corresponding displacement and phase data, generating a sequence of reverse-bias voltage control signals. The active metasurface, responding to these control signals, generates a radar response that yields falsified data observed and reconstructed by the eavesdropper.

Figure 8(a) illustrates Eve's sensing observation when the phone remains silent (with no audio on the speaker), but the on-phone metasurface is activated with misinformation. The results reveal that Eve infers an artificial vibration signature corresponding to the misinformation despite the absence of any mechanical sound wave vibrations from the phone. As shown in the transcript text in Figure 8(a), the word error rate, defined in Section 5.2, is zero in this experiment. That is, Eve accurately intercepted and decoded all of the misinformation (see **Audio 5**), even though the phone was silent and no information was communicated.

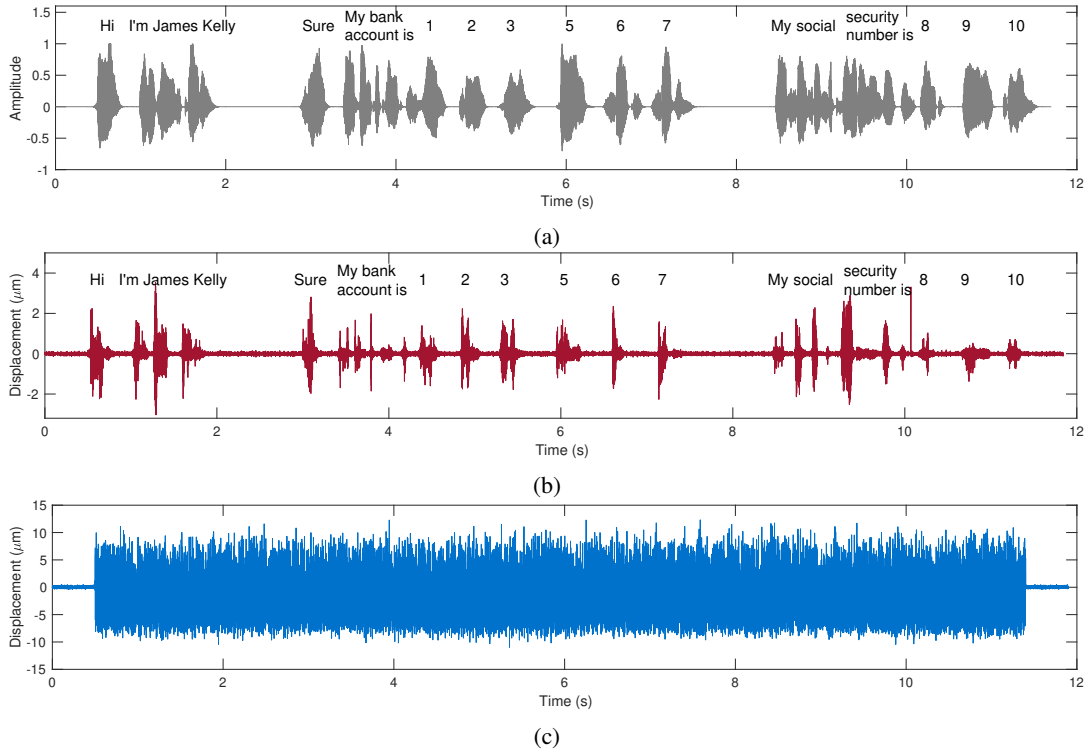


Figure 7: (a) An exemplary phone conversation audio incorporating personal identification, bank account details, and a social security number. (b) Remote Eve accurately measures the temporal mechanical displacement using radar signals, allowing her to recover all words in the eavesdropped data. (c) With the proposed defense mechanism, Eve completely fails to recover any words, observing only noise.

Next, we conduct an experiment in which Alice activates the misinformation security capability while actively speaking, such that the true audio information is simultaneously played on the phone speaker while the metasurface is being programmed to inject misinformation. Additionally, she intentionally introduces small-scale additive white noise, totaling 10% of the maximum amplitude, into the misinformation signature. This serves to further obscure legitimate information, which has an amplitude several times lower. Figure 8(b)-(c) depicts Eve's measurements, both temporal and spectral responses.

Importantly, we discover that the eavesdropper detects none of the original emitted words from the speaker, as shown in the reconstructed audio transcript in Figure 8(b) while intercepting only the misinformation injected by Alice (see **Audio 6**). Specifically, Eve falsely detects the name as "John Wick" instead of "James Kelly", while the original bank account numbers "1, 2, 3, 5, 6, 7" and the social security number "8, 9, 10" is misleading intercepted as "7, 7, 3, 8, 0, 1" and "3, 5, 9", respectively.

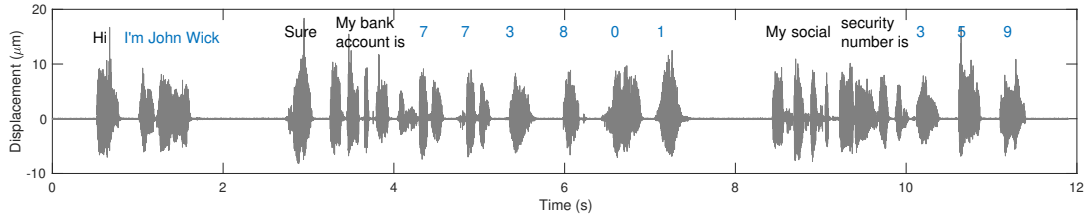
Finding: MiSINFO enables Alice to spoof Eve with false information, purposefully altering sensitive details such as personal identification, bank account numbers, and social security numbers. Alice can also inject a false acoustic signal even in the absence of a phone conversation.

6.4. More than a Thousand Words

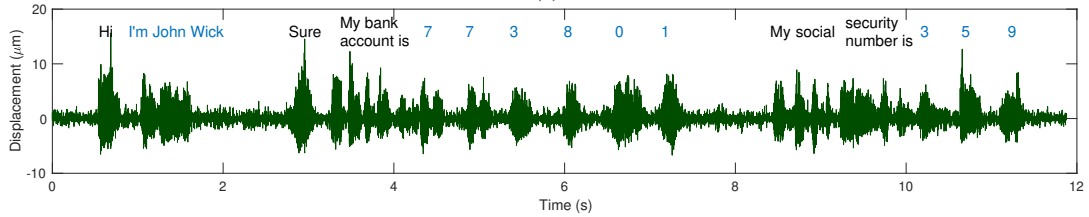
Here, we investigate Eve's performance in decoding MiSINFO audio vibration patterns through a large-scale experiment comprising over a thousand words of exemplary audio.

Adopting the setup from the previous experiment, we record the first 1006 words from the Declaration of Independence. Next, we convert the amplitude information of the temporal audio signal into displacement and phase data and generate a sequence of reverse-bias voltage control signals, as described in Section 4. As Eve activates the MiSINFO, Eve's sensing observations get modulated accordingly. We evaluate the performance by analyzing Eve's word error rate as she intercepts misleading audio data. We present the results in Figure 9, with Figure 9(a) depicting the ten most commonly recognized words by Eve, while Figure 9(b) illustrates the incorrectly recognized words.

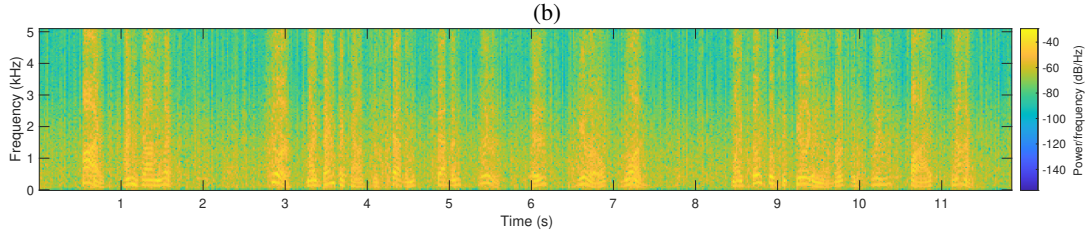
We discover that Eve incorrectly decodes only 23 words out of 1006. Namely, the number of substitution words is 15, deletion words 7, and insertion words 1. That is, Alice's injected misinformation is reconstructed by Eve with an extremely low word error rate of 2.29%. , Among a small group of incorrectly decoded words, we also observe a pattern of many voiceless consonants such as 'f', 's', 'sh', and 'th', for example, "sufferable", "pressing", "shewn", and "therein". As such, these words contain some signal



(a)

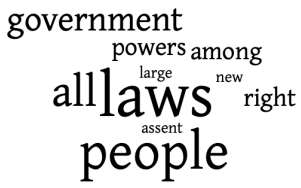


(b)

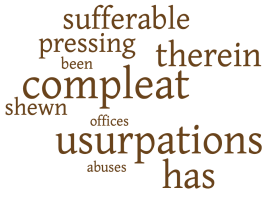


(c)

Figure 8: (a) MiSINFO enables spoofing the eavesdropper with audio misinformation, even when there is no audio on the speaker. (b) Eve observes superimposed information and misinformation patterns, recognizing none of the original words while recovering all misinformation words. (c) The spectrogram of the audio in (b) shows that most of the human speech energy is indeed below several kHz.



(a) Correctly recognized



(b) Incorrectly recognized

Figure 9: Top ten words (a) correctly and (b) incorrectly decoded by Eve in a large-scale experiment with over a thousand words from the Declaration of Independence audio information.

energy in higher frequencies, above 5 kHz, as detailed in Section 4.4. Missing these details with the radar's 10 kHz chirp sampling rate may occasionally lead the audio-to-text transcription algorithm to mistakenly associate them with other similar and common words. For instance, "shewn" is transcribed as "shown", and "pressing" is transcribed as "press". However, we highlight that the majority of the signal energy in human speech lies below 4 kHz, and this is also illustrated by the yellow color in Figure 8(c) spectrogram heatmap. Therefore, MiSINFO effectively enables the spoofing of the eavesdropper with legitimate yet false information.

Finding: The large-scale experiment with over a thousand words demonstrates that Alice can highly accurately inject audio misinformation to Eve, achieving an average word error rate of 2.29%

7. Discussion

In this paper, we present the MiSINFO architecture, which introduces a new foundation for sensing security and offers significant strategic advantages over conventional countermeasures. Traditionally, malicious sensing signals are either obscured, e.g., by randomizing the responses, or jammed, e.g., [49], [50]. However, the key challenge is that traditional countermeasures can quickly alert attackers to the presence of a countermeasure, prompting attackers to adapt and employ alternative audio eavesdropping techniques, such as microphones, laser-microphones [51], visual-microphones [52], or even malware [53]; although each method has its own advantages and disadvantages - for instance, using a microphone to recover audio is particularly challenging due to the low sound pressure levels emitted from smartphone earpieces (as opposed to loudspeakers) and the presence of acoustic noise in the environment.

Millimeter-wave radars: An attacker using a millimeter-wave radar can tap into various forms of

physical information leakage, including eavesdropping on audio by detecting tiny vibrations on surfaces such as windows and walls [54]. Unlike traditional listening devices such as microphones, which rely on air pressure changes and can be blocked by walls or other barriers, millimeter-wave radars work remotely and can extract sound information by analyzing Doppler shifts or microscopic surface movements. This makes them a stealthy and contactless alternative for audio surveillance. The risk becomes even greater with advanced radar architectures—narrow-beam phased arrays and high-resolution sensing allow attackers to focus precisely on their targets while filtering out background noise. As these technologies improve, the potential for remote audio reconstruction grows, raising serious security concerns and emphasizing the need for effective countermeasures against unauthorized millimeter-wave-based surveillance.

To address this, our architecture not only hides private acoustic signals but also injects a misleading alternative signal, potentially rendering the countermeasure unnoticed by attackers. Such a proactive defensive capability to mislead provides major strategic and security advantages across civilian, corporate, and military scenarios. As such, our proposed security approach transforms defensive measures from merely reactive to proactively deceptive, giving the defender an advantage.

On-phone metasurfaces: Metasurfaces are quickly transitioning into commercially viable products, with a growing presence in consumer electronics. This shift is driven by their ability to replace traditional, bulkier sensors, offering a more compact and lightweight alternative. Notable examples of this trend include the integration of metasurfaces as optical elements in high-end devices such as the Galaxy S23 Ultra and Google Pixel 8 Pro, where they serve to enhance functionality while reducing the overall size and weight of the devices [55], [56]. Unlike conventional sensors, metasurfaces can perform complex tasks such as manipulating light and electromagnetic waves with high precision, all within a thin, planar form factor. This development not only highlights the potential of metasurfaces to revolutionize consumer devices, but also strengthens the case for their integration into our on-phone metasurface security architecture, ensuring that such technologies are both feasible and scalable for practical, real-world applications.

We also note that MiSINFO can be activated on demand, such as during sensitive conversations, ensuring minimized power consumption when not in use. In fact, the power consumption of our metasurface architecture is low (below several mW) due to the reverse voltage bias design, which results in minimal current flow. Additionally, we emphasize that our metasurface is not intended to cover the phone's millimeter-wave antennas, if present, so the misinformation signal is directed away from the phone. Any interference would arise from Eve's radar signal, which has been previously explored, e.g., [57]. This design makes our proposed architecture both efficient and practical for continuous or frequent use.

Counter-counter attacks: Wireless security is an arms race, and a strong Eve could design sophisticated techniques to counter the proposed defensive countermeasure. For example, signal processing methods such as blind source separation [58] could be used to isolate audio misinformation from legitimate information. Adaptive filtering [59] could aid in identifying remnants of the actual audio vibrations by leveraging statistical properties and exploiting stochastic perturbations. Additionally, deep learning-based denoising techniques, such as generative adversarial networks (GANs) and deep autoencoders, could help reconstruct clean signals from obfuscated content [60].

Similarly, on-phone metasurfaces could introduce non-idealities that may be detectable. Specifically, resonance meta-atoms on the structure are likely to create frequency artifacts that are distinguishable from the natural spectral characteristics [44]. By analyzing these artifacts, an adversary could infer the presence of a defense mechanism. Then, as a potential counter-counterattack, Eve could randomize her radar signals, making it more difficult for Alice to construct the correct spoofing phase, thus requiring an even more advanced Alice than considered here. These topics present promising directions for future research.

Current limitations and future work: In this paper, we developed the concept, designed the system with a dynamic on-phone metasurface, and experimentally demonstrated the security mechanism. However, the considered threat model is demonstrative, yet constrained, with experimental evaluations limited to modest half-meter distances and fixed smartphone positions. Nevertheless, we emphasize that MiSINFO is designed to modulate the phase of the incoming radar beam, regardless of whether the chirps propagate over the air for half a meter or tens of meters. Additionally, different designs of the on-phone metasurface and angular responses impact the control signal mapping. In general, considering a stronger threat model, extending the operational range to tens of meters, and exploring performance under varying smartphone positions, motions, and user-holding conditions are promising directions for future research.

8. Conclusion

In this paper, we develop, design, and experimentally evaluate MiSINFO, a novel architecture for audio misinformation security. Our work is the first eavesdropping countermeasure system that not only prevents Eve from succeeding but also injects a false signal into her observation to fool her into believing that she has. Our experimental results reveal Eve detects none of the original words emitted from the smartphone speaker, and reconstructs the injected misinformation almost perfectly, so that she cannot tell that she has been spoofed.

Acknowledgment

This research was supported by Army Research Office DURIP grant W911NF-23-1-0340, Intel, Cisco, and by NSF grants 2402783, 2211618, 2148132, 1955075, 2402781,

2211616, and 1954780. This work was performed, in part, at the Center for Integrated Nanotechnologies, an Office of Science User Facility operated for the U.S. Department of Energy (DOE) Office of Science by Los Alamos National Laboratory (Contract 89233218CNA000001) and Sandia National Laboratories (Contract DE-NA-0003525).

References

[1] Number of smartphone users worldwide. https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/. Accessed: February 20, 2024.

[2] Chao Wang, Feng Lin, Tiantian Liu, Kaidi Zheng, Zhibo Wang, Zhengxiong Li, Ming-Chun Huang, Wenyao Xu, and Kui Ren. mmEve: eavesdropping on smartphone's earpiece via COTS mmWave device. In *Proceedings of the 28th Annual International Conference on Mobile Computing And Networking*, pages 338–351, 2022.

[3] Suryoday Basak and Mahanth Gowda. mmspy: Spying phone calls using mmWave radars. In *2022 IEEE Symposium on Security and Privacy (SP)*, pages 1211–1228. IEEE, 2022.

[4] Chao Wang, Feng Lin, Zhongjie Ba, Fan Zhang, Wenyao Xu, and Kui Ren. Wavesdropper: Through-wall word detection of human speech via commercial mmwave devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 6(2):1–26, 2022.

[5] Yiwen Feng, Kai Zhang, Chuyu Wang, Lei Xie, Jingyi Ning, and Shijia Chen. mmeavesdropper: Signal augmentation-based directional eavesdropping with mmWave radar. In *IEEE INFOCOM 2023-IEEE Conference on Computer Communications*, pages 1–10. IEEE, 2023.

[6] Pengfei Hu, Yifan Ma, Panneer Selvam Santhalingam, Parth H Pathak, and Xiuzhen Cheng. Milliear: Millimeter-wave acoustic eavesdropping with unconstrained vocabulary. In *IEEE INFOCOM 2022-IEEE Conference on Computer Communications*, pages 11–20. IEEE, 2022.

[7] Chao Wang, Feng Lin, Tiantian Liu, Ziwei Liu, Yijie Shen, Zhongjie Ba, Li Lu, Wenyao Xu, and Kui Ren. mmphone: Acoustic eavesdropping on loudspeakers via mmwave-characterized piezoelectric effect. In *IEEE INFOCOM 2022-IEEE Conference on Computer Communications*, pages 820–829. IEEE, 2022.

[8] Mohammed S Elbamby, Cristina Perfecto, Mehdi Bennis, and Klaus Doppler. Toward low-latency and ultra-reliable virtual reality. *IEEE Network*, 32(2):78–84, 2018.

[9] Shuping Dang, Osama Amin, Basem Shihada, and Mohamed-Slim Alouini. What should 6g be? *Nature Electronics*, 3(1):20–29, 2020.

[10] Jaime Lien, Nicholas Gillian, M Emre Karagozler, Patrick Amihood, Carsten Schwesig, Erik Olson, Hakim Raja, and Ivan Poupyrev. Soli: Ubiquitous gesture sensing with millimeter wave radar. *ACM Transactions on Graphics (TOG)*, 35(4):1–19, 2016.

[11] Nirupam Roy and Romit Roy Choudhury. Listening through a vibration motor. In *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*, pages 57–69, 2016.

[12] Yan Michalevsky, Dan Boneh, and Gabi Nakibly. Gyrophone: Recognizing speech from gyroscope signals. In *23rd USENIX Security Symposium (USENIX Security 14)*, pages 1053–1067, 2014.

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

[14] S Abhishek Anand, Chen Wang, Jian Liu, Nitesh Saxena, and Yingying Chen. Spearphone: a lightweight speech privacy exploit via accelerometer-sensed reverberations from smartphone loudspeakers. In *Proceedings of the 14th ACM Conference on Security and Privacy in Wireless and Mobile Networks*, pages 288–299, 2021.

[15] Zhongjie Ba, Tianhang Zheng, Xinyu Zhang, Zhan Qin, Baochun Li, Xue Liu, and Kui Ren. Learning-based practical smartphone eavesdropping with built-in accelerometer. In *NDSS*, volume 2020, pages 1–18, 2020.

[16] Pengfei Hu, Wenhao Li, Riccardo Spolaor, and Xiuzhen Cheng. mmecho: A mmwave-based acoustic eavesdropping method. In *Proceedings of the ACM Turing Award Celebration Conference-China 2023*, pages 138–140, 2023.

[17] Zhambyl Shaikhanov, Mahmoud Al-Madi, Hou-Tong Chen, Chun-Chieh Chang, Sadhvikas Addamane, Daniel M Mittleman, and Edward W Knightly. Audio misinformation encoding via an on-phone sub-terahertz metasurface. *Optica*, 11(8):1113–1114, 2024.

[18] Hanting Zhao, Ya Shuang, Menglin Wei, Tie Jun Cui, Philipp del Hougne, and Lianlin Li. Metasurface-assisted massive backscatter wireless communication with commodity wi-fi signals. *Nature communications*, 11(1):3926, 2020.

[19] Zhambyl Shaikhanov, Sherif Badran, Hichem Guerboukha, Josep Jornet, Daniel Mittleman, and Edward Knightly. Metafly: Wireless backhaul interception via aerial wavefront manipulation. In *2024 IEEE Symposium on Security and Privacy (SP)*, pages 151–151. IEEE Computer Society, 2024.

[20] Fahid Hassan, Zhambyl Shaikhanov, Hichem Guerboukha, Daniel M Mittleman, Kaushik Sengupta, and Edward W Knightly. Rmdm: Using random meta-atoms to send directional misinformation to eavesdroppers. In *2023 IEEE Conference on Communications and Network Security (CNS)*, pages 1–9. IEEE, 2023.

[21] Zhambyl Shaikhanov, Fahid Hassan, Hichem Guerboukha, Daniel Mittleman, and Edward Knightly. Metasurface-in-the-middle attack: from theory to experiment. In *Proceedings of the 15th ACM Conference on Security and Privacy in Wireless and Mobile Networks*, pages 257–267, 2022.

[22] Geng-Bo Wu, Jun Yan Dai, Kam Man Shum, Ka Fai Chan, Qiang Cheng, Tie Jun Cui, and Chi Hou Chan. A universal metasurface antenna to manipulate all fundamental characteristics of electromagnetic waves. *Nature Communications*, 14(1):5155, 2023.

[23] Zhambyl Shaikhanov, Sherif Badran, Josep M Jornet, Daniel M Mittleman, and Edward W Knightly. Remotely positioned metasurface-drone attack. In *Proceedings of the 24th International Workshop on Mobile Computing Systems and Applications*, pages 110–116, 2023.

[24] Xingyu Chen, Zhengxiong Li, Biacheng Chen, Yi Zhu, Chris Xiaoxuan Lu, Zhengyu Peng, Feng Lin, Wenyao Xu, Kui Ren, and Chunming Qiao. Metawave: Attacking mmwave sensing with metamaterial-enhanced tags. In *The 30th Network and Distributed System Security (NDSS) Symposium 2023*, pages 1–17. The Internet Society, 2023.

[25] Zhambyl Shaikhanov, Fahid Hassan, Sherif Badran, Hichem Guerboukha, Josep Miquel Jornet, Daniel M Mittleman, and Edward Knightly. Metafly: Aerial" metasurface-in-the-middle" attacks on wireless backhaul links. *GetMobile: Mobile Computing and Communications*, 28(4):5–11, 2025.

[26] Rohith Reddy Vennam, Ish Kumar Jain, Kshitiz Bansal, Joshua Orozco, Puja Shukla, Aanjhan Ranganathan, and Dinesh Bharadia. mmspoof: Resilient spoofing of automotive millimeter-wave radars using reflect array. In *2023 IEEE Symposium on Security and Privacy (SP)*, pages 1807–1821. IEEE, 2023.

[27] Hongxin Zeng, Sen Gong, Lan Wang, Tianchi Zhou, Yaxin Zhang, Feng Lan, Xuan Cong, Luyang Wang, Tianyang Song, YunCheng Zhao, Ziqiang Yang, and Daniel M. Mittleman. A review of terahertz phase modulation from free space to guided wave integrated devices. *Nanophotonics*, 11:415–437, 2022.

[28] Hou-Tong Chen, Sabarni Palit, Talmage Tyler, Christopher M. Bingham, Joshua M. O. Zide, John F. O'Hara, David R. Smith, Arthur C. Gossard, Richard D. Averitt, Willie J. Padilla, Nan M. Jokerst, and Antoinette J. Taylor. Hybrid metamaterials enable fast electrical modulation of freely propagating terahertz waves. *Applied Physics Letters*, 93:091117, 2008.

[29] Yaxin Zhang, Kesen Ding, Hongxin Zeng, Wei Kou, Tianchi Zhou, Hongji Zhou, Sen Gong, Ting Zhang, Lan Wang, Shixiong Liang, Feng Lan, Yazhou Dong, Zhihong Feng, Yubin Gong, Ziqiang Yang, and Daniel M. Mittleman. Ultrafast modulation of terahertz radiation using on-chip dual-layer near-field coupling. *Optica*, 9:1268–1275, 2022.

[30] Y. Zhang, S. Qiao, S. Liang, Z. Wu, Z. Yang, Z. Feng, H. Sun, Y. Zhou, L. Sun, Z. Chen, X. Zou, B. Zhang, J. Hu, S. Li, Q. Chen, L. Li, G. Xu, Y. Zhao, , and S. Liu. Gbps terahertz external modulator based on a composite metamaterial with a double-channel heterostructure. *Nano Letters*, 15:3501–3506, 2015.

[31] Feng Lan, Luyang Wang, Hongxin Zeng, Yaxin Zhang, Shixiong Liang, Tianyang Song, Wenxin Liu, Pinaki Mazumder, Ziqiang Yang, and Daniel M. Mittleman. Real-time programmable metasurface for terahertz multifunctional wave front engineering. *Light: Science & Applications*, 12:191, 2023.

[32] Ethan Tseng, Shane Colburn, James Whitehead, Luocheng Huang, Seung-Hwan Baek, Arka Majumdar, and Felix Heide. Neural nano-optics for high-quality thin lens imaging. *Nature Communications*, 12:6493, 2021.

[33] Klint Finley. A new lens technology is primed to jump-start phone cameras, 2021. Wired.

[34] Jaebaek Jung, Woojun Lee, Gyuha Lee, Songcheol Hong, and Jungsuek Oh. Ultra-thinned metasurface-embedded smartphone antenna-in-package for millimeter wave 5G/6G coverage enhancement. *IEEE Transactions on Antennas and Propagation*, 2023.

[35] Jaebaek Jung and Jungsuek Oh. 3D tempered glass-covered metasurface antenna-in-package enabling reliable 5G/6G smartphone beam coverage. In *2022 International Symposium on Antennas and Propagation (ISAP)*, pages 193–194. IEEE, 2022.

[36] Alessio Izzo, Ludovico Ausiello, Carmine Clemente, and John J Soraghan. Loudspeaker analysis: A radar based approach. *IEEE Sensors Journal*, 20(3):1223–1237, 2019.

[37] S. RAO. Introduction to mmwave sensing: Fmcw radars. TI mmWave Training Series, 2017.

[38] Wenshan Cai, Uday K Chettiar, Alexander V Kildishev, and Vladimir M Shalaev. Optical cloaking with metamaterials. *Nature photonics*, 1(4):224–227, 2007.

[39] Gun-Yeal Lee, Jong-Young Hong, SoonHyoung Hwang, Seokil Moon, Hyeokjung Kang, Sohee Jeon, Hwi Kim, Jun-Ho Jeong, and Byoungho Lee. Metasurface eyepiece for augmented reality. *Nature communications*, 9(1):1–10, 2018.

[40] Badreddine Assouar, Bin Liang, Ying Wu, Yong Li, Jian-Chun Cheng, and Yun Jing. Acoustic metasurfaces. *Nature Reviews Materials*, 3(12):460–472, 2018.

[41] Norman R French and John C Steinberg. Factors governing the intelligibility of speech sounds. *The journal of the Acoustical society of America*, 19(1):90–119, 1947.

[42] R. J. Baken. *Clinical Measurement of Speech and Voice*. Taylor and Francis Ltd, London, 2 edition, 2000.

[43] James L Fitch and Anthony Holbrook. Modal vocal fundamental frequency of young adults. *Archives of Otolaryngology*, 92(4):379–382, 1970.

[44] Hou-Tong Chen, Willie J. Padilla, Joshua M. O. Zide, Arthur C. Gossard, Antoinette J. Taylor, and Richard D. Averitt. Active terahertz metamaterial devices. *Nature*, 444:597–600, 2006.

[45] Hou-Tong Chen, Willie J Padilla, Michael J Cich, Abul K Azad, Richard D Averitt, and Antoinette J Taylor. A metamaterial solid-state terahertz phase modulator. *Nature Photonics*, 3:148–151, 2009.

[46] Nicholas Karl, Kimberly Reichel, Hou-Tong Chen, Antoinette J. Taylor, Igal Brener, Alexander Benz, John L. Reno, Rajind Mendis, and Daniel M. Mittleman. An electrically driven terahertz metamaterial diffraction modulator with over 20 dB of dynamic range. *Applied Physics Letters*, 104:091115, 2014.

[47] Speechify.com. Artificial intelligence voice generator for customizable human-like speech from text. https://www.speechify.com/, Accessed 2023. Online.

[48] Amazon Web Services. Amazon transcribe. https://aws.amazon.com/transcribe/, Accessed 2023. Online.

[49] Ming Gao, Yike Chen, Yajie Liu, Jie Xiong, Jinsong Han, and Kui Ren. Cancelling speech signals for speech privacy protection against microphone eavesdropping. In *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*, pages 1–16, 2023.

[50] Ke Sun, Chen Chen, and Xinyu Zhang. " alexa, stop spying on me!" speech privacy protection against voice assistants. In *Proceedings of the 18th conference on embedded networked sensor systems*, pages 298–311, 2020.

[51] Miles J. Anderson. Laser microphone, 1984.

[52] Abe Davis, Michael Rubinstein, Neal Wadhwa, Gautham J Mysore, Fredo Durand, and William T Freeman. The visual microphone: Passive recovery of sound from video. 2014.

[53] Guillermo Suarez-Tangil, Juan E Tapiador, Pedro Peris-Lopez, and Arturo Ribagorda. Evolution, detection and analysis of malware for smart devices. *IEEE communications surveys & tutorials*, 16(2):961–987, 2013.

[54] Pengfei Hu, Wenhao Li, Yifan Ma, Panneer Selvam Santhalingam, Parth Pathak, Hong Li, Huanle Zhang, Guoming Zhang, Xiuzhen Cheng, and Prasant Mohapatra. Towards unconstrained vocabulary eavesdropping with mmwave radar using gan. *IEEE Transactions on Mobile Computing*, 23(1):941–954, 2022.

[55] Yole Group. Metasurfaces break through: Turning speculation into reality, 2023. Accessed: 2025-03-20.

[56] Metalenz and STMicroelectronics. Metalenz and stmicroelectronics deliver world's first optical metasurface technology for consumer electronics devices, 2023. Accessed: 2025-03-20.

[57] Shiyu Cheng, Kaveh Pahlavan, Haowen Wei, Zhuoran Su, Seyed Reza Zekavat, and Ali Abedi. A study of interference analysis between mmwave radars and ieee 802.11 ad at 60 ghz bands. *International Journal of Wireless Information Networks*, 29(3):222–231, 2022.

[58] Adel Belouchrani, Karim Abed-Meraim, J-F Cardoso, and Eric Moulines. A blind source separation technique using second-order statistics. *IEEE Transactions on signal processing*, 45(2):434–444, 1997.

[59] Simon S Haykin. *Adaptive filter theory*. Pearson Education India, 2002.

[60] Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

Appendix A.
Meta-Review

The following meta-review was prepared by the program committee for the 2025 IEEE Symposium on Security and Privacy (S&P) as part of the review process as detailed in the call for papers.

A.1. Summary

This paper presents a countermeasure on audio-eavesdropping with mmWave radars. Without countermeasure, the radar signal from the attacker is modulated by small audio-induced vibrations on the victim. The countermeasure consists in reflecting the attacker's signal with a metasurface, modulating noise or even audio misinformation on top of it.

A.2. Scientific Contributions

- Establishes a New Research Direction
- Provides a Valuable Step Forward in an Established Field
- Creates a New Tool to Enable Future Science

A.3. Reasons for Acceptance

1) Establishes a New Research Direction. The paper takes a novel approach to defend against audio eavesdropping. The main contribution is that of turning the victim from a passive reflector (which leaks information through its vibrations) to an active reflector (which uses a metasurface to actively modulate the reflected signal with noise or even misinformation). Such approach might find applications beyond those discussed in the paper.
2) Provides a Valuable Step Forward in an Established Field / Creates a New Tool to Enable Future Science. Prior work has demonstrated the threat posed by mmWave radars capable of measuring minuscule victim movements, for example, by recovering audio from audio-induced vibrations. This paper contributes to this field by further replicating the attacks and investigating a countermeasure. Similarly, prior work has explored the use of metasurfaces for security, and this paper further proposes another application.

A.4. Noteworthy Concerns

1) Threat model and experimental evaluation: The experimental evaluation of the countermeasure (and corresponding attack) occurs in simplified conditions (e.g., very short distance, stationary victim). In such conditions, the practicality of the attack is very limited and questions weather the additional cost and complexity of the countermeasure are justified. On the one hand, it is reasonable to evaluate a countermeasure in positive conditions for the attacker, and attacks in more challenges conditions were discussed in prior work. In addition, the countermeasure injects signals at higher SNR and has thus an inherent advantage that would likely apply also at larger distances. On the other hand, only further experiments in more challenging scenarios would fully prove the relevance of the threat model and the performance of the countermeasure in realistic conditions.

Appendix B.
Response to the Meta-Review

1) We agree with the reviewers that testing at longer distances and under more complex conditions would strengthen the evaluation. Moreover, prior work has demonstrated the attack at distances of up to ten meters and under varying smartphone positions, motions, and user holding conditions [2], with key contributions focused on neural network denoising algorithms and radar sensor architectures to improve SNR and thereby extend the range. Among the various methods employed, we performed cubic interpolation to remove frame artifacts, polynomial regression to compensate for phase reset drifts, and filtering with cluster suppression to eliminate hardware-induced detrending and noise. However, we believe that fully replicating [2], including the design and implementation of neural networks and improved radar architectures, along with their experimental evaluations - on top of our primary contributions to audio misinformation development, on-phone metasurface design and implementation, and experimental evaluations - would be beyond the scope of a single conference paper but represents a promising direction for future research.