MetaGuardian: Enhancing Voice Assistant Security through Advanced Acoustic Metamaterials

Zhiyuan Ning[†‡#], Zheng Wang[◇], Zhanyong Tang[†‡§*]
[†]NorthWest University, China, [◇]University of Leeds, United Kingdom
[‡]Xi'an Key Laboratory of Advanced Computing and Software Security
[§]Shaanxi Key Laboratory of Passive Internet of Things and Neural Computing
[#]Shaanxi International Joint Research Centre for the Battery-Free Internet of Things
[†]ningzhiyuan@stumail.nwu.edu.cn, [◇]z.wang5@leeds.ac.uk, [†]zytang@nwu.edu.cn

Abstract

Voice assistants (VAs) have become integral to daily life, yet their always-on microphones make them attractive targets for attacks that threaten user privacy and safety. We present MetaGuardian, the first system to leverage acoustic metamaterials to defend against three major classes of attacks for VAs - inaudible, adversarial, and laser-based - within a single, portable design. Unlike prior defenses, MetaGuardian can be seamlessly integrated into the enclosures of commercial smart devices, providing strong protection without requiring software modification, hardware redesign, or costly machine learning models. MetaGuardian leverages mutual impedance effects between metamaterial units to extend the protection range to 16–40 kHz, effectively blocking wideband inaudible attacks. It also employs a carefully designed coiled space structure to disrupt adversarial signals while preserving normal VA operations. Its universal design allows flexible adaptation to different devices, striking a balance between portability and protection effectiveness. In controlled evaluations, MetaGuardian achieves a high defense success rate across all attack types, offering a practical and reliable foundation for securing VAs on smart devices.

CCS Concepts

• **Security and privacy → Malware and its mitigation**; **Artificial immune systems**.

1 Introduction

Voice assistants (VAs) such as Apple Siri, Google Assistant, and Amazon Alexa are now an essential part of modern mobile devices and smart home systems [25, 37, 39, 43, 47, 49, 68]. Their growing ubiquity, however, has also exposed them to a range of security threats [28, 44, 66], including inaudible, adversarial, and even laser-based attacks. Inaudible attacks embed malicious commands in ultrasonic or near-ultrasonic signals, which remain imperceptible to human hearing yet are reliably recognized by VAs [12, 14, 28, 44, 45, 54, 55, 63, 66]. Adversarial attacks use carefully crafted audio inputs that sound benign to humans but are interpreted as harmful commands by the system. Laser-based attacks go further, exploiting amplitude-modulated light to remotely inject commands. All three types of attack are highly covert, making them difficult to detect and even harder to defend against.

A variety of defenses have been explored. Software-based approaches monitor microphone input and attempt to detect abnormal signals, disabling the VA if a threat is suspected [32, 64, 65]. While attractive in principle, these solutions suffer from reliability issues across different microphone models, struggle to block malicious signals without degrading usability [21, 45, 59, 66], and are difficult to deploy on closed-source VA platforms. Hardware-based defenses [21, 29, 46] typically require device modification or additional active components, which increase cost, compromise portability, and face similar reliability issues in complex environments [21, 58, 59, 66].

Recent advances in acoustic metamaterials [16, 18, 27, 36, 38, 40, 41, 70, 72, 73] provide new opportunities to overcome the limitations of existing defenses. These materials exploit carefully engineered passive structures to manipulate sound, selectively blocking malicious signals while preserving normal sound signals. Unlike software-based defenses, metamaterials act directly on the acoustic channel before signals reach the microphone, and unlike hardware modifications, they require no invasive changes to the device. Their compact, passive, and low-cost nature makes them well-suited for unobtrusive integration into mobile devices and smart speakers. For instance, Figure 1 illustrates deployment scenarios where existing defenses are ineffective, but metamaterials can provide a robust and practical solution.

Despite this promise, current acoustic metamaterials face important challenges. Their narrow resonant frequency range - the small band of sound frequencies at which the structure naturally vibrates and can therefore block signals effectively [35] - means that more than a dozen units may be required to cover the spectrum of inaudible attacks, severely limiting portability. Attempts to defend against adversarial audio can also interfere with the recognition of legitimate

*Corresponding author

voice commands, reducing usability in practice. In addition, variations in device shape and microphone placement make it challenging to integrate acoustic metamaterials, further hindering deployment in real-world systems.

We present MetaGuardian, the first VA defense system that harnesses acoustic metamaterials to address the aforementioned limitations, delivering a comprehensive, practical, and portable solution. MetaGuardian introduces three key innovations. **First**, it exploits *the mutual impedance effect* - where nearby metamaterial units interact and extend each other's frequency response - to increase the filtering range to 16–40 kHz. This allows MetaGuardian to block inaudible attacks with only *three units*, reducing structural complexity and enabling seamless integration into mobile devices and smart speakers. **Second**, it incorporates *a labyrinth-style coiled metamaterial* - an acoustic structure that folds long sound paths into a compact volume, much like a maze forces travelers to take a longer route in a confined space. This design enables strong control over specific frequencies, allowing MetaGuardian to selectively distort critical bands (2–4 kHz) to disrupt adversarial inputs while minimally affecting legitimate voice commands. **Third**, it provides *a portable, universal design* adaptable to different microphone placements, ensuring usability by transmitting legitimate commands through reserved acoustic channels while suppressing malicious signals.

We showcase that MetaGuardian can be easily fabricated using low-cost resin 3D printing and requires no modification to the target device or its software stack, or the training of machine learning models, which can be costly and time-consuming. We evaluate the effectiveness of the MetaGuardian prototype on nine commercial devices against representative VA attacks: five adversarial attacks [14, 15, 54, 62, 63], three inaudible attacks [45, 55, 58], and one laser attack [50]. Extensive experiments show that MetaGuardian consistently defends against all tested inaudible and adversarial attacks within practical attack ranges reported in prior studies, while also providing protection against laser injection. Compared with existing defenses, MetaGuardian delivers the first *integrated, low-cost, and portable* solution for safeguarding VAs.

This paper makes the following contributions:

- It introduces the first metamaterial-based VA defense system that protects against inaudible, adversarial, and laser attacks without software or hardware modifications.
- It demonstrates, for the first time, how mutual impedance effects can be exploited to extend metamaterial filtering to 16–40 kHz, blocking wide-band inaudible attacks with compactness and efficiency.

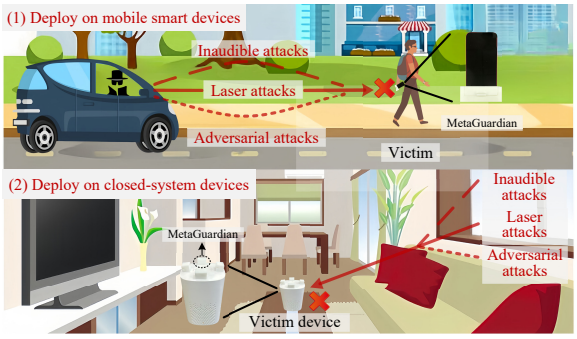


Figure 1: Deployment scenarios of MetaGuardian: (1) protecting mobile devices in public without sacrificing portability; (2) shielding smart speakers from inaudible, adversarial, and laser attacks.

- It presents a novel coiled metamaterial design that simultaneously disrupts adversarial inputs while preserving the usability of legitimate commands.
- It provides a universal and portable design that is manufacturable via consumer-grade 3D printing.

Online material. The 3D printing CAD files and demonstration videos of MetaGuardian are available at https://github.com/Meta-Guardian/MetaGuardian.

2 Background and Related Work

In this section, we introduce the relevant background and compare MetaGuardian with prior defense strategies and alternative solutions.

2.1 Covert Attacks on Voice Assistants

Voice assistants (VAs) are vulnerable to three covert attack types: adversarial, inaudible, and laser. Unlike traditional transcription-based attacks, these can be executed without the victim's awareness, making them a greater threat [13].

Adversarial attacks embed malicious audio into conversations or music to deceive voice assistants into executing unintended commands [19, 53, 56]. For example, CommanderSong [63] hides adversarial perturbations in songs, while VRIFLE [33] embeds them in user commands, enabling covert control of VAs.

Inaudible attacks exploit ultrasonic frequencies, typically between 16 and 40 kHz, to deliver hidden voice commands. These attacks exploit weaknesses in how commercial microphones process sound, particularly in the early stages of the analog signal chain. In a typical microphone, an acoustic sensor such as a microelectromechanical systems diaphragm converts sound waves into electrical signals. These signals are then passed to a preamplifier. Ideally, the amplifier should increase the signal strength without altering its structure.

However, due to limitations in device design, circuit implementation, and manufacturing processes, the amplifier often introduces nonlinear distortion when processing high-frequency signals. This distortion leads to the mixing of different frequencies. When an attacker sends an ultrasonic signal that carries a voice command, the nonlinear response of the amplifier causes frequency mixing. This process produces unintended low-frequency components that fall within the normal range of human speech. These components resemble the original voice command and are interpreted and executed by the voice assistant as if they had been spoken aloud by a person [28, 44, 45, 66]. Although placing filters before the amplifier can help reduce the impact of inaudible attacks through analog signal processing, both modifying commercial microphones and using external filters have practical challenges. Modifying built-in microphones is difficult because they are usually integrated into closed proprietary chips that do not offer accessible interfaces for hardware changes. Furthermore, the wide variation in circuit designs across devices leads to high costs and poor adaptability. Using external filters also introduces complications, as these solutions require additional acoustic sensing components and separate power supplies. This increases system complexity and deployment costs, and makes them unsuitable for everyday use.

Laser attacks use modulated laser beams to inject commands into microphones, operating stealthily at distances over 100 meters, posing severe risks to privacy and device security [50].

2.2 Software-based Defenses

Software-based approaches have been proposed to counter VA attacks. They employ varied tactics to counter voice threats. For example, EarArray [65] detects inaudible attacks via signal timing differences across microphones. NormDetect [32] improves this by detecting missing features of the attack signal without heavy data needs. MVP-EARS [64] reveals adversarial attacks through voice assistant transcription mismatches, and VSMask [52] blocks them with real-time perturbations.

Software solutions often have limited reliability and may block attack signals at the cost of disrupting the normal operation of VAs. Their deployment is further challenged by the lack of access to internal systems on commercial devices. A key limitation is that detection methods based on signal features do not generalize well across different platforms, due to variations in microphone sensitivity and frequency response (see also Section 6.3.1) [32, 65]. As a result, these methods often fail in real-world settings. Some defenses try to stop inaudible attacks by disabling the VA entirely, which undermines normal usability [32, 45, 58]. Moreover, as shown in Table 1, many commercial smart speakers restrict access to audio data for security reasons [32, 34, 65]. This restriction makes it difficult to test or deploy software defenses on real devices. Since simulation environments cannot fully reflect the diversity of hardware in actual products, evaluations based on them may lead to reduced effectiveness in practice.

Table 1: Smart speakers' audio access restrictions

Manuf.	Product Name	VA	Access Restr.
Amazon	Echo Series	Alexa	No
Apple	HomePod Series	Siri	No
Xiaomi	Xiaomi Speaker Series	Xiao AI	No
Huawei	Huawei AI Speaker Series	Xiaoyi	No

2.3 Hardware-based Defenses

Hardware-based solutions introduce changes to the hardware to defend against attacks on voice systems. For example, AIC [21] uses an additional speaker array to interfere with and block inaudible attacks. VocalPrint [29] uses millimeter wave probes to detect throat vibrations and confirm that the voice input is coming from a live human rather than a playback device. Similarly, the work presented in [46] uses a throat microphone to distinguish the user's voice from external speaker signals.

As hardware-based defenses require modifications to standard circuits or rely on non-portable active components, they have limited practical feasibility. Commercial devices usually adopt closed hardware architectures, making such invasive modifications challenging for end users. These modifications are often non-transferable across devices and can compromise functionality and stability, leading to compatibility issues [21, 29]. In addition, some hardware defenses depend on bulky, power-hungry components, such as speaker arrays or millimetre-wave radars [21, 29]. These solutions hinder portability and restrict deployment, particularly in outdoor or mobile settings. Furthermore, introducing additional hardware or circuit modifications increases system complexity and potential failure points. Attackers often exploit hardware-level traits, such as microphone nonlinearity [44, 45, 58]. While these defenses can reduce certain risks, they may also create new vulnerabilities, such as instability, that could serve as new entry points for attacks.

2.4 Acoustic Metamaterials

Acoustic metamaterials are engineered structures that control sound in ways ordinary materials cannot. By incorporating cavities, channels, or coils much smaller than the wavelength of sound, they can bend, block, absorb, or filter sound with high precision [16, 35, 72] . Much like a flute or bottle resonates at a specific tone, these structures "tune"

sound waves, but at a much finer scale. Arrays of such units can therefore act as compact, custom sound filters.

MetaGuardian leverages this principle to provide non-intrusive protection for smart devices without requiring changes to their software or hardware. Through carefully designed internal structures, MetaGuardian modifies the phase and amplitude of sound waves within targeted frequency ranges [30, 35, 61, 67], disrupting adversarial and inaudible attacks. When fabricated from opaque resin, these metamaterials can also block laser-based attacks. This enables MetaGuardian to deliver effective, low-cost, and external protection for VAs.

Unlike software or hardware defenses, acoustic metamaterials act directly on the acoustic channel through their passive physical structure. They require no power supply, are compact in size, and can be placed externally in front of a microphone, overcoming many of the limitations of conventional solutions. Nonetheless, challenges remain in broadening the filtering frequency range, preserving normal device functionality, and ensuring seamless integration across diverse device designs, which MetaGuardian is designed to address.

Metamaterials vs. analog filters. Acoustic metamaterials act like analog filters for blocking acoustic signals. However, as discussed in Section 2.1, analog filters are difficult to deploy at scale in commercial devices. In contrast, acoustic metamaterials intercept attack signals before they reach the microphone, preventing effective attack components from being generated inside the device. Therefore, MetaGuardian require no modification to the microphone hardware, offering lower deployment costs and greater adaptability.

3 Our Approach

MetaGuardian leverages acoustic metamaterials to build a VA defense system that is portable across devices and requires no modifications to the target device's hardware or software. Developing MetaGuardian requires addressing three key challenges: (1) *Expanding the filter range* of acoustic metamaterial units to provide comprehensive protection against inaudible attacks; (2) *Achieving robustness* against adversarial attacks while preserving accurate recognition of legitimate audio; (3) *Ensuring portability* across diverse devices while balancing functionality and protection. The following subsections (Sections 3.1–3.3) detail our solutions to these challenges.

3.1 Expanding Filtering Range

Traditional acoustic metamaterials often rely on Helmholtz-like resonators, which are small cavity-neck structures that trap and absorb sound energy at a specific frequency (much like how blowing across the top of a bottle produces a single tone). These resonators are effective for ultrasound filtering, but their narrow bandwidth makes it difficult to cover the full range of inaudible attack frequencies. To address this, we propose a solution based on the *mutual impedance effect*, which broadens the filtering range of metamaterial units and enables comprehensive defense against inaudible attacks.

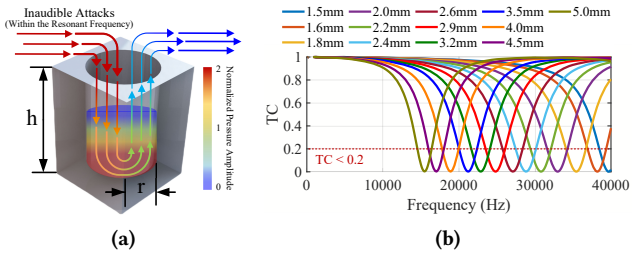


Figure 2: (a) Helmholtz-like acoustic metamaterial unit. (b) Filtering effect of 13 units across 16–40 kHz.

3.1.1 Narrowband metamaterials. Helmholtz-like acoustic metamaterials rely on their geometric structure to resonate at specific ultrasonic frequencies, effectively filtering targeted bands [35]. As shown in Figure 2a, these devices typically consist of a cylindrical cavity connected to a narrow neck. When external sound waves enter, the air inside resonates, absorbing energy near the target frequency and reducing the transmission of those waves.

The resonant frequency f_0 can be estimated as [35]:

$$f_0 = \frac{v}{4(h+r)} \tag{1}$$

where v is the speed of sound in air (343 m/s), h is the cavity depth, and r is the radius of the neck. By adjusting h and r, resonators can be tuned to specific frequencies.

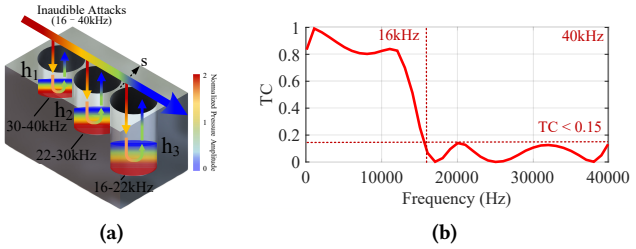
However, each unit only covers a narrow range, typically 1–2 kHz [35], while inaudible attacks span 16–40 kHz. As shown in Figure 2b, covering this range requires about 13 resonators, which increases size and complexity and makes integration with real devices impractical.

3.1.2 Broadband filtering with the mutual impedance effect. Recent studies [35, 48] show that the *mutual impedance effect* can shift and broaden resonant frequencies when metamaterial units are placed close together. We exploit this property to achieve broadband filtering with far fewer units.

Specifically, mutual impedance increases the total impedance of the system, which lowers and broadens the resonant frequency range. The effect depends strongly on how the units are placed:

Effect of spacing. The strength of mutual impedance is inversely related to the spacing S between units:

$$Z_{\text{mutual}} \propto \frac{1}{S}. \tag{2}$$



(a) (b)

Figure 3: (a) Inaudible Attack Defense Metamaterial(IADM) and (b) its filtering performance in 16-40 kHz range.

Smaller spacing increases acoustic coupling, making the resonance effect stronger. Based on acoustic coupling theory, this relationship is similar to how mutual inductance works in electromagnetics. To balance performance with manufacturability, we set the spacing to 0.1 mm, which maximizes coupling while remaining feasible for 3D printing.

Effect of arrangement. The layout of the units is also important. A linear arrangement produces stronger coupling than circular layouts, where destructive interference reduces the effect. Simulations confirm that linear spacing achieves the strongest broadband filtering.

Implementation. Using these insights, we linearly arranged three metamaterial units with 0.1mm spacing and heights $h_1 = 2$ mm, $h_2 = 3.2$ mm, and $h_3 = 4.8$ mm. To validate this configuration, we used the COMSOL Multiphysics framework [5] to simulate the resulting resonant structures and evaluate their acoustic response under our design parameters. COMSOL is a widely used multiphysics simulation platform known for its ability to accurately solve coupled acoustic-structural problems. It has been extensively validated in metamaterials and acoustics research, making its simulation results highly reliable [35, 65]. The simulation results show that our design expands the resonance range nearly fourfold, effectively covering the 16–40 kHz attack band. We refer to this configuration as the *Inaudible Attack Defense Metamaterial (IADM)*. As shown in Figure 3b, the IADM reduces ultrasonic transmission to below 15%, demonstrating strong protection against inaudible attacks.

3.2 Achieving Robustness

To address adversarial attacks, we propose a coiling-up space-structured metamaterial capable of amplifying signal amplitude within a specific frequency range, thereby disrupting or weakening the critical features of attack signals and neutralizing adversarial attacks [14]. However, if the interference frequency range is crucial for legitimate audio, it may affect the normal operation of the voice assistant. Therefore, precise analysis and the design of metamaterials tailored to that frequency range are necessary.

3.2.1 Selection of interference frequency bands. To ensure the intelligibility of legitimate audio while effectively interfering with adversarial attack signals, it is crucial to select an appropriate interference frequency band. The clarity of human speech (100-4000 Hz) primarily depends on the first (F1: 100-1000 Hz) and second formants (F2: 1000-2000 Hz) [11, 26], while the 2000-4000 Hz range mainly carries consonant details, contributing only about 10% of the total speech information entropy ($H_{\text{high}}/H_{\text{total}} \approx 10\%$) [22, 51]. Conversely, adversarial attacks typically embed perturbations in the 2000-4000 Hz frequency range to enhance their stealth, allowing them to interfere with the normal operation of speech recognition systems without being easily perceived by the human ear [31, 42, 57]. Therefore, interfering within this frequency range can maximize the suppression of adversarial attacks while preserving essential speech content.

Coiling-up space-structured metamaterials can effectively neutralize adversarial attack signals by amplifying perturbations and introducing nonlinear distortion. Adversarial attacks typically add a small perturbation $\delta(t)$ to the legitimate audio, with its power significantly lower than the original signal:

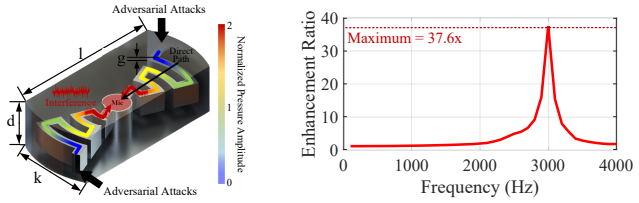
$$x_{\text{adv}}(t) = x_{\text{clean}}(t) + \delta(t), \quad P_\delta(f) \ll P_{x_{\text{clean}}}(f). \tag{3}$$

Metamaterials utilize frequency-selective resonance to significantly amplify signals within a specific band. Given a transmission gain $H(f)$, the processed signal is expressed as: $x_{\text{meta}}(t) = \mathcal{F}^{-1}\{H(f)X_{\text{adv}}(f)\}$., when $H(f) \gg 1$ (applied only to the 2000-4000 Hz range), the perturbation $\delta(t)$ is greatly amplified, introducing nonlinear distortion that disrupts attack features:

$$\tilde{\delta}(t) = \mathcal{F}^{-1}\{H(f)\Delta(f)\}. \tag{4}$$

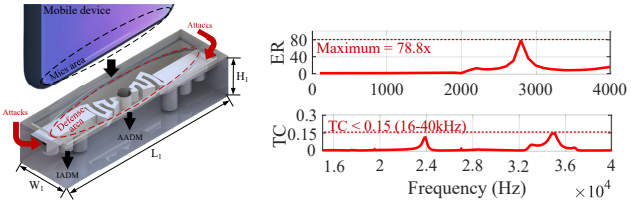
Therefore, this metamaterial design effectively weakens adversarial attacks.

Advantages over direct filtering. While modifying the IADM structure can also filter out the frequency band used in adversarial attacks, this band also carries important information for automatic speech recognition and speaker identification. As a result, direct filtering is likely to degrade these functions and significantly impair daily usage. In contrast, the space-wrapping metamaterial used by MetaGuardian selectively interferes with critical features of attack signals. Although it may introduce some impact on speech, it preserves legitimate audio to the greatest extent, making it a more practical and effective defense against adversarial attacks. Section 6.1 empirically shows the advantage of MetaGuardian over direct filtering.



(a) (b)

Figure 4: (a) Adversarial Attack Defense Metamaterial(AADM) and (b) its interference performance.



(a) (b)

Figure 5: (a) Mobile devices structure design and (b) its filtering and interference performance.

3.2.2 Metamaterial design for adversarial attack defense. We propose a novel coil space-structured acoustic metamaterial to enhance audio signals in the 2000-4000 Hz frequency range and achieve interference effects. As shown in Fig. 4a, the metamaterial adopts a slender design, effectively reducing its size and improving portability. It consists of two sets of helical spatial structures that extend the propagation path of sound waves to regulate the resonance frequency, generating strong resonance within the target frequency band. During resonance, the acoustic energy is concentrated and amplified, thereby enhancing signals in this frequency range to interfere with adversarial signals. The dimensions of the metamaterial are as follows: length $l = 15$ mm, width $k = 7.65$ mm, height $d = 4.75$ mm, and internal channel width $g = 0.8$ mm.

Initially, the resonant frequency f_r of the acoustic metamaterial determines its response and amplification capability for specific frequency signals, and is closely related to the internal path length L_{coiled}. The formula for calculating the resonant frequency f_r is:

$$f_r = \frac{c}{4L_{\text{coiled}}} \tag{5}$$

where c denotes the speed of sound in air, which is 343 m/s, and L_{coiled} represents the length of the coiling path within the metamaterial. As the frequency of the sound wave approximates the resonant frequency, the metamaterial demonstrates its most potent energy response, thereby amplifying signals within that particular frequency spectrum. By judiciously selecting an appropriate path length L_{coiled}, the resonant frequency of the metamaterial can be modulated to align with the designated frequency range.

In the proposed design, the specified target frequency range is 2000-4000 Hz, thereby setting the resonant center frequency as noted in $f_r = 3000$ Hz. Using Equation 5, the calculated coil path length is determined to be as indicated in $L_{\text{coiled}} = 28.5$ mm. This configuration ensures that the metamaterial produces a substantial enhancement effect within the designated target frequency range. Subsequently, after determining L_{coiled}, the sound pressure amplification factor G is calculated using the following equation:

$$G = \frac{n_r}{\lambda_0} \cdot \sqrt{\frac{2\rho c^2}{\lambda_0^2}} \tag{6}$$

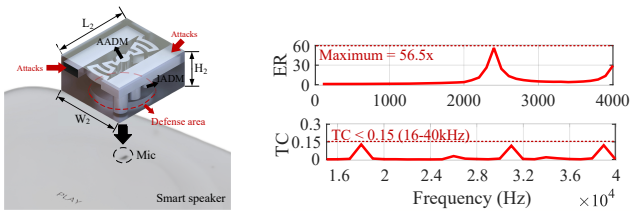
In this context, the refractive index $n_r = \frac{L_{\text{coiled}}}{L_{\text{blue}}}$ is defined as the quotient of the propagation speed of sound waves within the metamaterial and their speed in air. By calculating the path length ratio shown in Figure 4a, this refractive index can be estimated. When an adversarial attack passes through the metamaterial with a high refractive index n_r, the sound pressure is excessively amplified, leading to distortion. This metamaterial is designated as the *Adversarial Attack Defense Metamaterial* (AADM).

The COMSOL simulation results (Figure 4b) are consistent with theoretical predictions, showing enhanced sound energy within the 2000-4000 Hz frequency range, with a maximum gain of 37.6 times at 3000 Hz. Subsequently, we also verified in Section 6.1 and Section 6.2.1 that AADM effectively defends against adversarial attacks while maintaining the integrity of legitimate audio signals.

3.3 Ensuring Portability

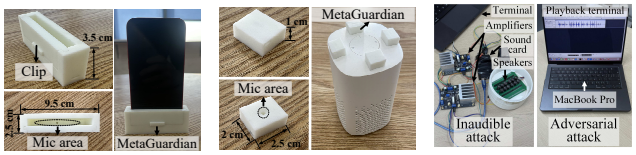
Although IADM and AADM each perform well in defense, MetaGuardian faces practical challenges due to significant differences in structure and microphone layouts among mainstream voice assistant devices. The key issue is how to integrate both into a universal defense structure that balances effective protection with device portability and functionality. To address this, we analyzed the structural features of mobile devices and smart speakers and designed dedicated universal defense solutions for each.

3.3.1 Structure design for mobile devices. When designing a universal structure for mobile devices, we first analyzed their common form factors - typically flat and elongated for easy portability. To preserve this portability, MetaGuardian adopts a similar shape. In addition, since most mobile devices use a bottom microphone for primary audio capture, the structure must be installed at that location for effective protection. Figure 5a illustrates our universal framework



(a) (b)

Figure 6: Smart speaker structure design (a) and its filtering and interference performance (b).



(a) (b) (c)

Figure 7: MetaGuardian prototypes for mobile devices (a) and smart speakers (b). Sub-figure (c) depicts the attack setup.

design. IADM and AADM units are arranged horizontally to fit the device shape. A recessed top secures the device and protects the microphone, while side channels (4 mm × 2 mm) allow legitimate voice signals to pass through. The 5 mm wall blocks 65 dB adversarial signals and resists laser attacks. Core dimensions are $L_1 = 40mm$, $W_1 = 25mm$, and $H_1 = 15mm$. To support different devices, only these three parameters need adjustment. For devices with multiple bottom microphones, additional AADM units can be positioned accordingly to enhance protection.

Impact on IADM and AADM performance. To evaluate the impact of the MetaGuardian structure design for mobile devices on defensive effectiveness against IADM and AADM, we used COMSOL to simulate its filtering performance in the ultrasonic range and its interference effects in the low-frequency range. As shown in Figure 5b, the structure effectively filters inaudible attacks within the 16-40 kHz range. The center frequency of low-frequency enhancement shifted to 2800 Hz, with the gain increasing to 78.8 times. We attribute this change to additional phase shifts along the channel path, which cause constructive interference at specific frequencies [10, 20]. This interference shifts the enhanced center frequency and increases the gain. Nevertheless, the variation remains within the acceptable interference frequency range discussed in Section 3.2, ensuring that adversarial attacks are effectively disrupted without impairing the recognition of legitimate commands.

3.3.2 Structure design for smart speakers. Smart speakers have microphones concentrated at the top in a circular layout. To fit this design, we developed a compact cubic structure that encloses a single microphone without obstructing buttons. Multiple such units can be combined to protect the entire microphone array. Figure 6a shows this structure. IADM and AADM are arranged in a zigzag pattern to reduce length and avoid blocking buttons. A circular recess at the bottom covers the microphone. The wall thickness matches that of the mobile device structure, allowing attack signals into the internal metamaterial. Dimensions are length $L_2 = 25mm$, width $W_2 = 20mm$, height $H_2 = 10mm$. The circular recess is adjustable to fit microphones of various shapes.

Impact on IADM and AADM performance. The COMSOL simulation results for the MetaGuardian structure design for smart speakers are shown in Figure 6b. The results confirm that the structure effectively filters inaudible attacks within the 16-40 kHz range. Compared to the MetaGuardian structure for mobile devices, the center frequency and gain of the low-frequency enhancement show slight variations, likely due to the shorter channel length producing a smaller additional phase shift. These variations are minor and do not affect the overall functionality of the structure.

4 Implementation

The MetaGuardian prototype is fabricated using resin 3D printing and includes two structural designs tailored for mobile devices and smart speakers (see Figure 7a and Figure 7b). The mobile version adopts a slender form to enhance portability, with a front clip to prevent slipping; the smart speaker version is more compact, with a bottom notch to preserve button functionality. Its modular design makes it easy to adapt to different microphone layouts. This structure balances portability and adaptability, and can be extended to various devices by adjusting design parameters.

Figure 7c shows the devices used in our experiments for inaudible and adversarial attacks: inaudible attacks are amplified through a power amplifier and transmitted via an ultrasonic transducer, while adversarial commands are played through the built-in speaker of a laptop (MacBook Pro).

5 Experimental Setup

5.1 Attack Systems

We tested MetaGuardian against three classes of known attacks for VAs: inaudible, adversarial, and laser attacks. For inaudible attacks, we reproduced three representative systems with different center frequencies: NUIT [55] (18 kHz), DolphinAttack [58] (25 kHz), and LipRead [45] (40 kHz), covering the typical attack range of 16–40 kHz. For adversarial audio, we implemented five open-source attacks: ALIF (2024) [15], KENKU (2023) [54], SMACK (2023) [62], CommanderSong (2018) [63], and Devil's Whisper (2020) [14]. We

Table 2: Nine representative attack systems evaluated.

System name	Attack type	Compatible devices
KENKU [54]	Adversarial	iPhone 16 Pro, Pixel 8 Pro, Echo Dot 5th, HomePod mini
SMACK [62]	Adversarial	iPhone 14 Pro, Echo Dot 5th
ALIF [15]	Adversarial	iPhone 14 Pro, Pixel 8 Pro, Echo Dot 5th
CommanderSong [63]	Adversarial	iPhone 14 Pro
Devil's Whisper [14]	Adversarial	iPhone 16 Pro, Pixel 8 Pro, Echo Dot 5th, HomePod mini
DolphinAttack [58], NUIT [55], LipRead [45]	Inaudible	All devices
Light Commands [50]	Laser	All devices

Table 3: Nine test devices from five major vendors.

Brand	Model	Type	VA (OS)
	iPhone 16 Pro	Mobile device	Siri (iOS 18)
	iPhone 14 Pro	Mobile device	iFlytek (7.0.4062)
Apple	HomePod mini	Smart speaker	Siri (18.2)
Google	Pixel 8 Pro	Mobile device	Google Assistant (Android 14)
	Xiaomi 14	Mobile device	XiaoAI (HyperOS 2)
Xiaomi	XiaoAI Play 2	Smart speaker	XiaoAI (1.62.26)
Huawei	Mate 60 Pro	Mobile device	Xiaoyi (HarmonyOS 4)
Huawei	AISpeaker 2e	Smart speaker	Xiaoyi (HarmonyOS 2)
Amazon	Echo Dot 5th	Smart speaker	Alexa (9698496900h)

also evaluated a laser-based attack, Light Commands [50], verifying its penetration capability with a laser pointer and a photosensor. Table 2 summarizes the tested systems used in our evalaution

5.2 Test Devices

We tested MetaGuardian on nine smart devices: five smartphones and four smart speakers, covering leading brands including Apple [4], Google [6], Xiaomi [7], Huawei [2], and Amazon [3]. These include flagship models and widely used consumer products covering a variety of use cases and price ranges. Table 3 lists the specifications.

5.3 Evaluation Metrics

We consider three complementary evaluation metrics: *Protection Success Rate (PSR)*, the percentage of failed attacks out of 30 attempts per system; *Word Interference Rate (WIR)*, the ratio of destroyed to total attack command keywords; and *Command Recognition Rate (CRR)*, the percentage of legitimate commands correctly recognized.

5.4 Evaluation Environments

We made efforts to test MetaGuardian under realistic attack conditions. For adversarial attacks, we used 65 dB voice commands, including common tasks such as *Open the door, Play music, Make a call, Send a message, Turn on the light, Transfer money, Navigate to my office,* and *Make a credit card payment*. For inaudible attacks, the same commands were

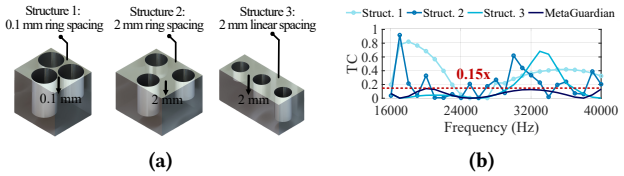


Figure 8: Three distinct structures (a) and comparison of transmission coefficients (TC) (b).

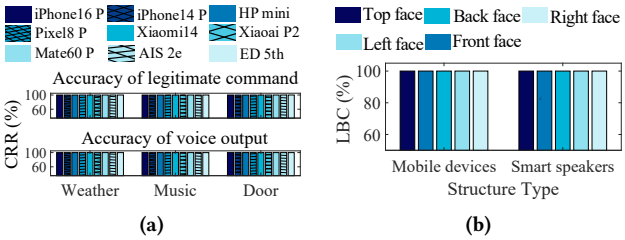


Figure 9: (a) Impact on commands input & playback, (b) Laser light-blocking coefficient (LBC).

transmitted using a 3-watt ultrasonic speaker. Laser attack tests used a laser pointer and photosensor setup.

All experiments were conducted in an open laboratory environment with a background noise level of approximately 43 dB. Table 4 summarizes the test objectives, while Figure 7c shows the attack devices.

6 Experimental Results

Highlights of our evaluation are:

- MetaGuardian extends the filtering range to 16–40kHz for ultrasonic signals while preserving normal voice command functionality (Section 6.1);
- MetaGuardian reliably defends against inaudible, adversarial, and laser-based attacks across diverse conditions (Section 6.2);
- Compared to existing software and hardware defenses, MetaGuardian offers improved reliability, compactness, and portability (Section 6.3).

We note that these results were obtained in a controlled environment, where factors such as user movement and ambient noise were minimized. The performance of MetaGuardian in more unconstrained, real-world settings may be affected by these additional variables.

6.1 Filtering Performance

We evaluated MetaGuardian's ability to filter ultrasonic signals using the Avisoft-Bioacoustics CM16/CMPA, an external ultrasonic measurement microphone, to record the

Table 4: Summary of experimental objectives and design.

Objective	Label	Focus	Description
Filtering performance	A1	Band coverage	Tested filtering from 16–40 kHz across different unit arrangements.
Impact on normal usage	A2	Usability	Measured command recognition during input and playback.
Adversarial defense	B1	Attack robustness	Five devices tested against five adversarial attacks.
Inaudible defense	B2	Ultrasound robustness	Nine devices tested against three inaudible attacks.
Laser defense	B3	Laser robustness	Tested with laser pointer at different angles.
Multi-angle defense	B4	Robustness across angles	Measured defense under different attack directions.
Precision interference	B5	Keyword disruption	Assessed ability to distort adversarial command words.
Anti-interference	B6	Environmental robustness	Tested under background noise interference.
Prior work reliability	C1	Cross-device robustness	Evaluated consistency of existing defenses across devices.
Comparison	C2	Benchmarking	Compared MetaGuardian with prior defense strategies.

strength of signals passing through it under different unit arrangements and spacings (see Figure 8a). As shown in Figure 8b , MetaGuardian achieves strong filtering performance consistent with COMSOL simulations (Section 3.1.2). The ring structure with 0.1 mm and 2 mm spacing (Structure 1 in Figure 8) is effective only in the 25–30 kHz range, whereas the linear structure with 0.1 mm spacing (Structure 3 in Figure 8) covers 16–28 kHz. These results show that a linear arrangement with reduced spacing significantly enhances the mutual impedance effect.

We also verified that MetaGuardian does not interfere with legitimate usage. Standard commands such as *What is the weather*, *Play music*, and *Open the door* were tested using both Google Cloud TTS [1] synthesized voices and recordings from 10 male and 10 female volunteers. As shown in Figure 9a, devices equipped with MetaGuardian successfully responded to all commands, and playback was accurately recognized by other devices, yielding a 100% command recognition rate. These results confirm that MetaGuardian preserves normal voice assistant and audio playback functions while providing effective protection.

6.2 Defense Performance

We evaluated the performance of MetaGuardian against various attacks under controlled conditions.

6.2.1 Adversarial attacks. We evaluated the system against five representative adversarial attacks (Table 2) on five VA-enabled devices. Figure 10 presents the attack success rates (PSR) in this setting. At distances where these attacks typically achieve high success, including KENKU [54] (70% at 0.3 m), CommanderSong [63] (82% at 1.5 m), SMACK [62] (64.7% at 0.5 m), Devil's Whisper [14] (90% at 2 m), and ALIF [15] (85.7% at 0.3 m), MetaGuardian maintained a 100% defense success rate. Even under more challenging conditions, with attacks launched from 0.1 m at 65 dB playback volume, defense success remained above 97% for all five attacks across nine devices. This robustness is due to the AADM structure's high-gain amplification in the 2000–4000 Hz range, which

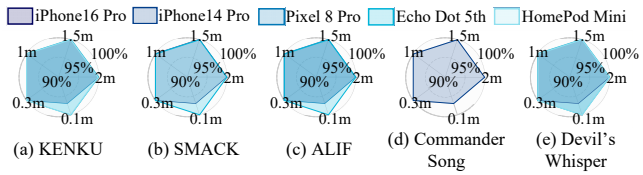


Figure 10: Adversarial attack defense at various ranges.

effectively disrupts adversarial signals while preserving the recognition of legitimate commands.

6.2.2 Inaudible attacks. To evaluate MetaGuardian's effectiveness against inaudible attacks, we tested nine devices at various distances and recorded the PSR in a controlled environment. The results are shown in Figure 11. Within the maximum effective ranges of three common attacks, DolphinAttack achieved 100% success at 19.8 meters, LipRead 50% at 7.62 meters, and NUIT over 80% at 3.8 meters, while MetaGuardian consistently maintained a 100% PSR. Even when the attack distance was reduced to 0.5 meters, the PSR for all three attacks remained above 93%. A slight decline in defense performance at closer distances is attributed to reduced signal attenuation, which allows part of the attack energy to exceed MetaGuardian's suppression threshold. However, inaudible attacks typically require conspicuous equipment such as speaker arrays, power amplifiers, and external power supplies, which are difficult to deploy discreetly at short range. As a result, the practical threat in such scenarios remains limited.

6.2.3 Laser attack. The main weakness of laser attacks is their inability to penetrate opaque barriers. We used a 60 mW laser pointer (the same maximum power as in Light Commands [50]) to illuminate two MetaGuardian structures from five angles, and measured their light-blocking coefficients with a TA636A light sensor to evaluate the protective effect. The results are shown in Figure 9b. At all tested angles,

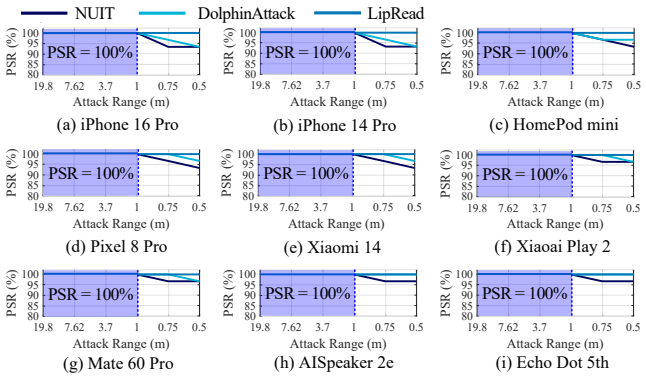


Figure 11: Inaudible attack defense at various ranges.

the laser pointer achieved 100% light blocking when shining on MetaGuardian, effectively preventing laser transmission. Analysis shows that MetaGuardian significantly attenuates the laser energy through optical absorption and refraction, blocking the attack commands carried by the laser and causing the attack to fail.

6.2.4 Multi-angle defenses in adversarial and inaudible attacks . In real-world scenarios, attacks may come from multiple directions. To evaluate MetaGuardian 's defense performance at different angles, we conducted adversarial and inaudible attacks from 15°, 30°, and 60° angles at distances of 0.1 m and 0.5 m, respectively, and recorded the attack success rate (PSR). The results are shown in Figures 12 and 13. For adversarial attacks, MetaGuardian consistently achieved a PSR exceeding 96% across all tested angles. For inaudible attacks, the PSR remained above 93% at all angles, with defense effectiveness improving as the angle increased, reaching 100% at 60°. This improvement is attributed to the optimized wall thickness design in MetaGuardian (see Section 3.3), which effectively blocks some attack signals, forcing the remaining signals to pass through the metamaterial's internal structure where they encounter interference.

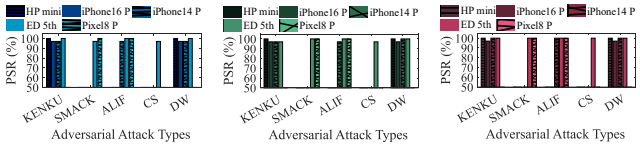
6.2.5 Precision interference. When defending against attacks containing multiple keywords, the system's ability to precisely interfere with each keyword is crucial. To evaluate MetaGuardian's interference effectiveness, we launched adversarial and inaudible attacks at distances of 0.1 m and 0.5 m on multiple mobile devices with speech-to-text capabilities (including iPhone 16 Pro, iPhone 14 Pro, Pixel 8 Pro, Xiaomi 14, and Mate 60 Pro) and calculated the Word Error Rate (WIR). The results are shown in Figures 14a and 14b. The results show that MetaGuardian achieves a WIR exceeding 95% in adversarial attacks, with deviations within 5%. Its optimized structure effectively disperses and absorbs keyword signal energy, hindering accurate recognition. In inaudible attacks, MetaGuardian maintains a WIR above 92.5%, with deviations controlled within 7%, demonstrating stable and effective defense capabilities and validating its effectiveness against complex attacks.

6.2.6 Anti interference. We evaluated MetaGuardian's anti-interference capability in outdoor settings. To this end, we tested devices in ~75 dB ambient noise and during user movement at 2m/s while launching attacks (Figures 15a and 15b). We measured the *Word Interference Rate (WIR)*, a higher-is-better metric (Section 5.3). MetaGuardian achieved a WIR of 98% against adversarial attacks and over 95% against inaudible attacks in noisy environments; during motion, WIR for both attack types remained above 97%, demonstrating robust reliability. Leveraging its passive structure, MetaGuardian alters sound-wave phases through material properties to disrupt targeted frequencies, providing stable protection without active signal analysis and remaining resilient to noise, temperature, and other environmental variations.

6.3 Compared to Prior Work

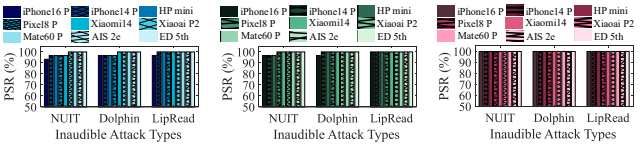
6.3.1 Reliability across microphones. Variations in the frequency response of microphones across different devices cause significant differences in the received audio signals, affecting the accuracy of defense methods based on signal feature detection [29, 32, 45, 58, 65]. We selected the classic LipRead method [45] for testing (other defense methods use similar signal feature extraction approaches). Under the same environment, the "turn on hotspot" command was recorded 30 times using different devices, and the average values of three features—power, autocorrelation coefficient, and amplitude skew—were calculated and combined into a comprehensive score. The results show that the differences in these three features across devices reached 17%, 22%, and 80.97%, respectively, causing some devices (such as iPhone 14 Pro, Xiaomi 14, and Pixel 8 Pro) to misclassify the attack command as legitimate (see Figure 16b). In contrast, MetaGuardian defends the microphone directly with a physical structure, portable across devices.

6.3.2 Advantages of MetaGuardian. We compare MetaGuardian with recent defense approaches for VAs to highlight its advantages. As shown in Table 5, five mainstream software-based defenses require disabling the voice assistant upon detecting an attack, which disrupts normal usage and is difficult to deploy in closed systems. Although these methods achieve over 90% defense success rates, they are, as discussed in Section 6.3.1, susceptible to variations in microphone characteristics across devices. In contrast, MetaGuardian employs a passive physical structure that directly disrupts attack signals outside the microphone, without modifying system logic or relying on software support, offering greater stability and broader compatibility.



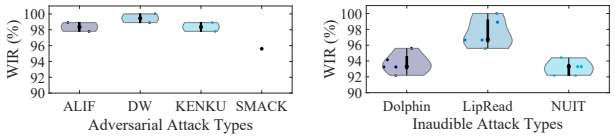
(a) Attack angle: 15° (b) Attack angle: 30° (c) Attack angle: 60°

Figure 12: Adversarial attack defense at various angles.



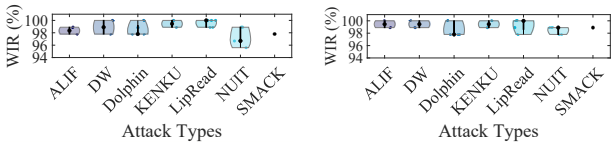
(a) Attack angle: 15° (b) Attack angle: 30° (c) Attack angle: 60°

Figure 13: Inaudible attack defense at various angles.



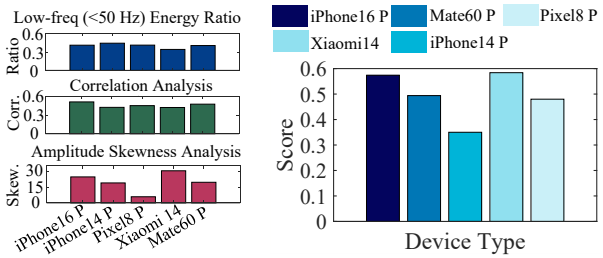
(a) (b)

Figure 14: WIR against adversarial (a) and inaudible (b) attacks.



(a) (b)

Figure 15: (a) WIR in a noisy environment, (b) WIR in a mobile environment.



(a) (b)

Figure 16: Feature responses of different devices to the same command (a) and comprehensive scores (b).

Existing hardware-based defense methods, such as AIC [21], VocalPrint [29], and the approach proposed by Sahidullah et al. [46], achieve defense success rates above 90%. However, they rely on active components such as speaker arrays, millimeter-wave radar, or continuously worn headsets, which reduce system reliability and portability. In contrast, MetaGuardian adopts a passive design that requires no device modifications or user intervention, offering strong compatibility and adaptability. Moreover, MetaGuardian can be seamlessly integrated with existing software and hardware defenses, demonstrating excellent synergy across different defense strategies.

7 Discussions

Naturally, there is room for improvement and further work. We highlight several promising directions below.

Countermeasures for dynamic attacks. The current implementation of MetaGuardian successfully mitigates fixed-band attacks through filtering and amplification. An exciting avenue for future work is extending its resilience against dynamic attacks like frequency-hopping and other adaptive attacks [17] . This can be achieved by using tunable acoustic metamaterials based on piezoelectric materials [8, 69] to dynamically adjust their operating frequency bands to enable real-time adaptation.

Improving portability. We show that MetaGuardian can be fabricated in a compact form factor suitable for everyday use. While the resin-based construction provides structural stability, portability can be further improved. Incorporating flexible metamaterials, such as elastomers [9, 23] , would allow for lighter and more adaptable designs.

Electromagnetic interference (EMI) defense. Our MetaGuardian already offers strong protection against voice and laser-based threats. A natural next step is to extend its capabilities to defend against electromagnetic interference (EMI) attacks, where malicious signals can be injected without using the acoustic channel. Integrating shape memory alloys [24, 71] could provide effective electromagnetic shielding, paving the way for a multilayer, multi-modal defense system that combines acoustic and electromagnetic resilience.

Impact on ultrasonic sensing. Our current design filters signals in the 16–40 kHz band, which may affect ultrasonic sensing applications such as proximity detection, gesture recognition, and acoustic analysis. This can be improved by using tunable acoustic metamaterials to selectively and dynamically adjust frequency bands to balance protection with functionality.

Table 5: Performance compared to prior research

System name	Function intact	Closed system def.	No modify	Portable	Multi-attack def.
DolphinAttack [66]	No	No	Yes	Yes	No
LipRead [45]	No	No	Yes	Yes	No
NormDetect [32]	No	No	Yes	Yes	No
EarArray [65]	No	No	Yes	Yes	No
VoShield [60]	No	No	Yes	Yes	Yes
AIC [21]	Yes	No	Yes	No	No
VocalPrint [29]	Yes	Yes	Yes	No	Yes
Sahidullah et al. [46]	Yes	Yes	Yes	No	Yes
MetaGuardian	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**

8 Conclusion

We have presented MetaGuardian, an acoustic metamaterial-based protection system designed to defend voice assistants against inaudible, adversarial, and laser attacks. Unlike prior approaches, MetaGuardian requires no modifications to software or hardware. Its design leverages mutual impedance to expand the filtering range and reduce device size, enabling frequency-targeted defense while preserving legitimate audio transmission. The system is adaptable to a wide range of devices and deployment scenarios. Our extensive experiments show that MetaGuardian provides effective and consistent protection across multiple attack types and hardware platforms, making it a reliable, practical solution.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (Grant No. 62372373) and the Shaanxi Science and Technology Innovation Team Program (Grant No. 2024RSCXTD05).

References

[1] https://cloud.google.com/speech-to-text. Google Text-to-Speech AI. Last accessed: 2025-3-1.

[2] https://consumer.huawei.com/cn/phones/. Huawei. Last accessed: 2025-1-20.

[3] https://www.amazon.com/smart-home-devices/b?ie=UTF8&node=9818047011. Amazon. Last accessed: 2025-1-20.

[4] https://www.apple.com.cn/iphone/. Apple. Last accessed: 2025-1-20.

[5] https://www.comsol.com/. COMSOL. Last accessed: 2025-1-20.

[6] https://www.google-mobile.cn/. Google. Last accessed: 2025-1-20.

[7] https://www.mi.com/. Xiaomi. Last accessed: 2025-1-20.

[8] Andrea Bacigalupo, Maria Laura De Bellis, and Diego Misseroni. 2020. Design of tunable acoustic metamaterials with periodic piezoelectric microstructure. *Extreme Mechanics Letters* 40 (2020), 100977.

[9] Katia Bertoldi, Vincenzo Vitelli, Johan Christensen, and Martin Van Hecke. 2017. Flexible mechanical metamaterials. *Nature Reviews Materials* 2, 11 (2017), 1–11.

[10] Liyun Cao, Zhichun Yang, Yanlong Xu, Zhaolin Chen, Yifan Zhu, Shi-Wang Fan, Krupali Donda, Brice Vincent, and Badreddine Assouar. 2021. Pillared elastic metasurface with constructive interference for flexural wave manipulation. *Mechanical Systems and Signal Processing* 146 (2021), 107035.

[11] Laurel H Carney, David A Cameron, Kameron B Kinast, C Evelyn Feld, Douglas M Schwarz, U-Cheng Leong, and Joyce M McDonough. 2023. Effects of sensorineural hearing loss on formant-frequency discrimination: Measurements and models. *Hearing Research* 435 (2023), 108788.

[12] Guangke Chen, Yedi Zhang, Zhe Zhao, and Fu Song. 2023. {QFA2SR}:{Query-Free} Adversarial Transfer Attacks to Speaker Recognition Systems. In *32nd USENIX Security Symposium (USENIX Security 23)*. 2437–2454.

[13] Yiwei Chen, Wenhao Li, XiuZhen Cheng, and Pengfei Hu. 2024. A survey of acoustic eavesdropping attacks: Principle, methods, and progress. *High-Confidence Computing* (2024), 100241.

[14] Yuxuan Chen, Xuejing Yuan, Jiangshan Zhang, Yue Zhao, Shengzhi Zhang, Kai Chen, and XiaoFeng Wang. 2020. {Devil's} whisper: A general approach for physical adversarial attacks against commercial black-box speech recognition devices. In *29th USENIX Security Symposium (USENIX Security 20)*. 2667–2684.

[15] Peng Cheng, Yuwei Wang, Peng Huang, Zhongjie Ba, Xiaodong Lin, Feng Lin, Li Lu, and Kui Ren. 2024. ALIF: Low-cost adversarial audio attacks on black-box speech platforms using linguistic features. In *2024 IEEE Symposium on Security and Privacy (SP)*. IEEE, 1628–1645.

[16] Ruizhi Dong, Man Sun, Fangshuo Mo, Dongxing Mao, Xu Wang, and Yong Li. 2021. Recent advances in acoustic ventilation barriers. *Journal of Physics D: Applied Physics* 54, 40 (2021), 403002.

[17] Tianyu Du, Shouling Ji, Jinfeng Li, Qinchen Gu, Ting Wang, and Raheem Beyah. 2020. Sirenattack: Generating adversarial audio for end-to-end acoustic systems. In *Proceedings of the 15th ACM Asia conference on computer and communications security*. 357–369.

[18] Yong Ge, Hong-xiang Sun, Shou-qi Yuan, and Yun Lai. 2019. Switchable omnidirectional acoustic insulation through open window structures with ultrathin metasurfaces. *Physical Review Materials* 3, 6 (2019), 065203.

[19] Taesik Gong, Alberto Gil CP Ramos, Sourav Bhattacharya, Akhil Mathur, and Fahim Kawsar. 2019. Audidos: Real-time denial-of-service adversarial attacks on deep audio models. In *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*. IEEE, 978–985.

[20] Paola Gori, Claudia Guattari, Francesco Asdrubali, Roberto de Lieto Vollaro, Alessio Monti, Davide Ramaccia, Filiberto Bilotti, and Alessandro Toscano. 2016. Sustainable acoustic metasurfaces for sound control. *Sustainability* 8, 2 (2016), 107.

[21] Yitao He, Junyu Bian, Xinyu Tong, Zihui Qian, Wei Zhu, Xiaohua Tian, and Xinbing Wang. 2019. Canceling inaudible voice commands against voice control systems. In *The 25th Annual International Conference on Mobile Computing and Networking*. 1–15.

[22] Kamran Heydari, Ali Akbar Tahaei, Akram Pourbakht, Hamid Haghani, and Ahmadreza Nazeri. 2021. The relationship between psychoacoustic and electrophysiological assessments of temporal resolution. *Journal of the American Academy of Audiology* 32, 03 (2021), 171–179.

[23] Shan Jiang, Xuejun Liu, Jianpeng Liu, Dong Ye, Yongqing Duan, Kan Li, Zhouping Yin, and YongAn Huang. 2022. Flexible metamaterial electronics. *Advanced Materials* 34, 52 (2022), 2200070.

[24] Xiaojian Jiang, Hao Yu, Haohao Lu, Yinsong Si, Yubing Dong, Yaofeng Zhu, Chen Qian, and Yaqin Fu. 2023. Anisotropic shape memory composite for dynamically adjustable electromagnetic interference shielding. *ACS Applied Polymer Materials* 5, 6 (2023), 4400–4410.

[25] Tae-Kook Kim. 2020. Short research on voice control system based on artificial intelligence assistant. In *2020 international conference on electronics, information, and communication (ICEIC)*. IEEE, 1–2.

[26] Ettien Koffi. 2024. A COMPREHENSIVE REVIEW OF FORMANTS: LINGUISTIC AND SOME PARALINGUISTIC APPLICATIONS. *Linguistic Portfolios* 13, 1 (2024), 2.

[27] Hoyeong Kwon, Dimitrios Sounas, Andrea Cordaro, Albert Polman, and Andrea Alù. 2018. Nonlocal metasurfaces for optical signal processing. *Physical review letters* 121, 17 (2018), 173004.

[28] Gen Li, Zhichao Cao, and Tianxing Li. 2023. EchoAttack: Practical Inaudible Attacks To Smart Earbuds. In *Proceedings of the 21st Annual International Conference on Mobile Systems, Applications and Services*. 383–396.

[29] Huining Li, Chenhan Xu, Aditya Singh Rathore, Zhengxiong Li, Hanbin Zhang, Chen Song, Kun Wang, Lu Su, Feng Lin, Kui Ren, et al. 2020. Vocalprint: exploring a resilient and secure voice authentication via mmwave biometric interrogation. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*. 312–325.

[30] Hong-Ze Li, Xu-Chang Liu, Qi Liu, Shuang Li, Jin-Shui Yang, Li-Li Tong, Sheng-Bo Shi, Rüdiger Schmidt, and Kai-Uwe Schröder. 2023. Sound insulation performance of double membrane-type acoustic metamaterials combined with a Helmholtz resonator. *Applied Acoustics* 205 (2023), 109297.

[31] Jiguo Li, Xinfeng Zhang, Jizheng Xu, Siwei Ma, and Wen Gao. 2021. Learning to fool the speaker recognition. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 17, 3s (2021), 1–21.

[32] Xinfeng Li, Xiaoyu Ji, Chen Yan, Chaohao Li, Yichen Li, Zhenning Zhang, and Wenyuan Xu. 2023. Learning normality is enough: a software-based mitigation against inaudible voice attacks. In *32nd USENIX Security Symposium (USENIX Security 23)*. 2455–2472.

[33] Xinfeng Li, Chen Yan, Xuancun Lu, Zihan Zeng, Xiaoyu Ji, and Wenyuan Xu. 2023. Inaudible adversarial perturbation: Manipulating the recognition of user speech in real time. *arXiv preprint arXiv:2308.01040* (2023).

[34] Zhuohang Li, Cong Shi, Tianfang Zhang, Yi Xie, Jian Liu, Bo Yuan, and Yingying Chen. 2021. Robust detection of machine-induced audio attacks in intelligent audio systems with microphone array. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*. 1884–1899.

[35] Joshua S Lloyd, Cole G Ludwikowski, Cyrus Malik, and Chen Shen. 2023. Mitigating inaudible ultrasound attacks on voice assistants with acoustic metamaterials. *IEEE Access* 11 (2023), 36464–36470.

[36] Mark B Lundeberg, Yuanda Gao, Reza Asgari, Cheng Tan, Ben Van Duppen, Marta Autore, Pablo Alonso-González, Achim Woessner, Kenji Watanabe, Takashi Taniguchi, et al. 2017. Tuning quantum nonlocal effects in graphene plasmonics. *Science* 357, 6347 (2017), 187–191.

[37] Michal Luria, Guy Hoffman, and Oren Zuckerman. 2017. Comparing social robot, screen and voice interfaces for smart-home control. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 580–628.

[38] Y-N Lv, A-W Liu, Y Tan, C-L Hu, T-P Hua, X-B Zou, YR Sun, C-L Zou, G-C Guo, S-M Hu, et al. 2022. Fano-like resonance due to interference with distant transitions. *Physical review Letters* 129, 16 (2022), 163201.

[39] Yash Mittal, Paridhi Toshniwal, Sonal Sharma, Deepika Singhal, Ruchi Gupta, and Vinay Kumar Mittal. 2015. A voice-controlled multi-functional smart home automation system. In *2015 Annual IEEE India Conference (INDICON)*. IEEE, 1–6.

[40] Adam Overvig and Andrea Alù. 2022. Diffractive nonlocal metasurfaces. *Laser & Photonics Reviews* 16, 8 (2022), 2100633.

[41] Adam C Overvig, Stephanie C Malek, and Nanfang Yu. 2020. Multifunctional nonlocal metasurfaces. *Physical Review Letters* 125, 1 (2020), 017402.

[42] Namgyu Park and Jong Kim. 2024. Toward robust ASR system against audio adversarial examples using agitated logit. *ACM Transactions on Privacy and Security* 27, 2 (2024), 1–26.

[43] Adam Rogowski. 2012. Industrially oriented voice control system. *Robotics and Computer-Integrated Manufacturing* 28, 3 (2012), 303–315.

[44] Nirupam Roy, Haitham Hassanieh, and Romit Roy Choudhury. 2017. Backdoor: Making microphones hear inaudible sounds. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*. 2–14.

[45] Nirupam Roy, Sheng Shen, Haitham Hassanieh, and Romit Roy Choudhury. 2018. Inaudible voice commands: The {Long-Range} attack and defense. In *15th USENIX Symposium on Networked Systems Design and Implementation (NSDI 18)*. 547–560.

[46] Md Sahidullah, Dennis Alexander Lehmann Thomsen, Rosa Gonzalez Hautamäki, Tomi Kinnunen, Zheng-Hua Tan, Robert Parts, and Martti Pitkänen. 2017. Robust voice liveness detection and speaker verification using throat microphones. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 26, 1 (2017), 44–56.

[47] Sonali Sen, Shamik Chakrabarty, Raghav Toshniwal, and Ankita Bhaumik. 2015. Design of an intelligent voice controlled home automation system. *International Journal of Computer Applications* 121, 15 (2015).

[48] Chao Shen, Yu Liu, and Lixi Huang. 2021. On acoustic absorption mechanisms of multiple coupled quarter-wavelength resonators: Mutual impedance effects. *Journal of Sound and Vibration* 508 (2021), 116202.

[49] Amit Kumar Sikder, Leonardo Babun, Z Berkay Celik, Hidayet Aksu, Patrick McDaniel, Engin Kirda, and A Selcuk Uluagac. 2022. Who's controlling my device? Multi-user multi-device-aware access control system for shared smart home environment. *ACM Transactions on Internet of Things* 3, 4 (2022), 1–39.

[50] Takeshi Sugawara, Benjamin Cyr, Sara Rampazzi, Daniel Genkin, and Kevin Fu. 2020. Light commands:{Laser-Based} audio injection attacks on {Voice-Controllable} systems. In *29th USENIX Security Symposium (USENIX Security 20)*. 2631–2648.

[51] Daniélli Rampelotto Tessele, Hélinton Goulart Moreira, Fernanda Soares Aurélio Patatt, Glória Cristina de Souza Streit, Larine da Silva Soares, and Michele Vargas Garcia. 2022. Descending audiometric configuration: tonal means, speech perception and audiological hearing disadvantage. *Audiology-Communication Research* 27 (2022), e2661.

[52] Yuanda Wang, Hanqing Guo, Guangjing Wang, Bocheng Chen, and Qiben Yan. 2023. Vsmask: Defending against voice synthesis attack via real-time predictive perturbation. In *Proceedings of the 16th ACM Conference on Security and Privacy in Wireless and Mobile Networks*. 239–250.

[53] Zhibo Wang, Hongshan Yang, Yunhe Feng, Peng Sun, Hengchang Guo, Zhifei Zhang, and Kui Ren. 2023. Towards transferable targeted adversarial examples. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 20534–20543.

[54] Xinghui Wu, Shiqing Ma, Chao Shen, Chenhao Lin, Qian Wang, Qi Li, and Yuan Rao. 2023. {KENKU}: Towards Efficient and Stealthy Black-box Adversarial Attacks against {ASR} Systems. In *32nd USENIX Security Symposium (USENIX Security 23)*. 247–264.

[55] Qi Xia, Qian Chen, and Shouhuai Xu. 2023. {Near-Ultrasound} Inaudible Trojan (Nuit): Exploiting Your Speaker to Attack Your Microphone. In *32nd USENIX Security Symposium (USENIX Security 23)*. 4589–4606.

[56] Meng Xue, Kuang Peng, Xueluan Gong, Qian Zhang, Yanjiao Chen, and Routing Li. 2023. Echo: Reverberation-based Fast Black-Box Adversarial Attacks on Intelligent Audio Systems. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 3 (2023), 1–24.

[57] Hiromu Yakura and Jun Sakuma. 2018. Robust audio adversarial example for a physical attack. *arXiv preprint arXiv:1810.11793* (2018).

[58] Chen Yan, Guoming Zhang, Xiaoyu Ji, Tianchen Zhang, Taimin Zhang, and Wenyuan Xu. 2019. The feasibility of injecting inaudible voice commands to voice assistants. *IEEE Transactions on Dependable and Secure Computing* 18, 3 (2019), 1108–1124.

[59] Qiben Yan, Kehai Liu, Qin Zhou, Hanqing Guo, and Ning Zhang. 2020. Surfingattack: Interactive hidden attack on voice assistants using ultrasonic guided waves. In *Network and Distributed Systems Security (NDSS) Symposium*.

[60] Qiang Yang, Kaiyan Cui, and Yuanqing Zheng. 2023. VoShield: Voice liveness detection with sound field dynamics. In *IEEE INFOCOM 2023-IEEE Conference on Computer Communications*. IEEE, 1–10.

[61] Xiaocui Yang, Fei Yang, Xinmin Shen, Enshuai Wang, Xiaonan Zhang, Cheng Shen, and Wenqiang Peng. 2022. Development of adjustable parallel helmholtz acoustic metamaterial for broad low-frequency sound absorption band. *Materials* 15, 17 (2022), 5938.

[62] Zhiyuan Yu, Yuanhaur Chang, Ning Zhang, and Chaowei Xiao. 2023. {SMACK}: Semantically Meaningful Adversarial Audio Attack. In *32nd USENIX Security Symposium (USENIX Security 23)*. 3799–3816.

[63] Xuejing Yuan, Yuxuan Chen, Yue Zhao, Yunhui Long, Xiaokang Liu, Kai Chen, Shengzhi Zhang, Heqing Huang, Xiaofeng Wang, and Carl A Gunter. 2018. {CommanderSong}: A systematic approach for practical adversarial voice recognition. In *27th USENIX security symposium (USENIX security 18)*. 49–64.

[64] Qiang Zeng, Jianhai Su, Chenglong Fu, Golam Kayas, Lannan Luo, Xiaojiang Du, Chiu C Tan, and Jie Wu. 2019. A multiversion programming inspired approach to detecting audio adversarial examples. In *2019 49th annual IEEE/IFIP international conference on dependable systems and networks (DSN)*. IEEE, 39–51.

[65] Guoming Zhang, Xiaoyu Ji, Xinfeng Li, Gang Qu, and Wenyuan Xu. 2021. EarArray: Defending against DolphinAttack via Acoustic Attenuation.. In *NDSS*.

[66] Guoming Zhang, Chen Yan, Xiaoyu Ji, Tianchen Zhang, Taimin Zhang, and Wenyuan Xu. 2017. Dolphinattack: Inaudible voice commands. In *Proceedings of the 2017 ACM SIGSAC conference on computer and communications security*. 103–117.

[67] Jin Zhang, Wei Rui, Chengrong Ma, Ying Cheng, Xiaojun Liu, and Johan Christensen. 2021. Remote whispering metamaterial for non-radiative transceiving of ultra-weak sound. *Nature Communications* 12, 1 (2021), 3670.

[68] Wenkai Zhang, Zihao An, Zhendong Luo, Wenyu Li, Zhao Zhang, Yimei Rao, Che Fai Yeong, and Feng Duan. 2016. Development of a voice-control smart home environment. In *2016 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. IEEE, 1697–1702.

[69] Xiaodong Zhang, Jing Nie, Jinhong He, Fengbin Lin, and Yang Liu. 2025. Digitally Controlled Piezoelectric Metamaterial for Low-Frequency and High-Efficiency Sound Absorption. *Materials* 18, 9 (2025), 2102.

[70] Yingxin Zhang, Yao Wei Chin, Xiang Yu, Milan Shrestha, Gih-Keong Lau, Boo Cheong Koo, Kun Liu, and Zhenbo Lu. 2023. Ventilated acoustic metasurface with low-frequency sound insulation. *JASA Express Letters* 3, 7 (2023).

[71] Shu Zhu, Qingya Zhou, Mengya Wang, Jackson Dale, Zhe Qiang, Yuchi Fan, Meifang Zhu, and Changhuai Ye. 2021. Modulating electromagnetic interference shielding performance of ultra-lightweight composite foams through shape memory function. *Composites Part B: Engineering* 204 (2021), 108497.

[72] Yi-Fan Zhu, Aurélien Merkel, Krupali Donda, Shiwang Fan, Liyun Cao, and Badreddine Assouar. 2021. Nonlocal acoustic metasurface for ultrabroadband sound absorption. *Physical Review B* 103, 6 (2021), 064102.

[73] Yi-Fan Zhu, Xin-Ye Zou, Bin Liang, and Jian-Chun Cheng. 2015. Acoustic one-way open tunnel by using metasurface. *Applied Physics Letters* 107, 11 (2015).