MetaGuardian: Enhancing Voice Assistant Security through Advanced Acoustic Metamaterials

Zhiyuan Ning
ningzhiyuan@stumail.nwu.edu.cn
NorthWest University
Xi'an, Shaanxi, China

Zheng Wang
z.wang5@leeds.ac.uk
University of Leeds
Leeds, United Kingdom

Zhanyong Tang*
zytang@nwu.edu.cn
NorthWest University
Xi'an, Shaanxi, China

Abstract

We present MetaGuardian, a voice assistant (VA) protection system based on acoustic metamaterials. MetaGuardian can be directly integrated into the enclosures of various smart devices, effectively defending against inaudible, adversarial and laser attacks without relying on additional software support or altering the underlying hardware, ensuring usability. To achieve this, MetaGuardian leverages the mutual impedance effects between metamaterial units to extend the signal filtering range to 16-40 kHz to effectively block wide-band inaudible attacks. Additionally, it adopts a carefully designed coiled space structure to precisely interfere with adversarial attacks while ensuring the normal functioning of VAs. Furthermore, MetaGuardian offers a universal structural design, allowing itself to be flexibly adapted to various smart devices, striking a balance between portability and protection effectiveness. In controled evaluation environments, MetaGuardian achieves a high defense success rate against various attack types, including adversarial, inaudible and laser attacks.

1 Introduction

Voice assistants (VAs), such as Apple Siri, Google Assistant, and Amazon Alexa, have become widely integrated into mobile devices and smart home systems [21, 34, 36, 40, 45, 47, 68]. However, their widespread adoption has also exposed them to various security threats [24, 41, 66], including inaudible, adversarial, and laser-based attacks. Inaudible attacks[24, 41, 42, 53, 66] embed malicious voice commands within ultrasonic or near-ultrasonic signals, making them imperceptible to human hearing but still recognizable by the VA[9, 11, 52, 63]. Adversarial attacks involve carefully crafted audio inputs that sound benign to users but are intentionally designed to be misinterpreted as harmful commands by the VA. Laser-based attacks exploit amplitude-modulated light to remotely inject commands into the system. These attack methods are highly covert, making detection and mitigation particularly challenging.

Efforts have been made to create hardware- and software-based solutions to mitigate VA attacks. Software-based solutions focus on detecting attack signals entering the microphone and alerting the user to disable the voice assistant when a threat is identified. However, they face reliability issues caused by differences in microphone models and often struggle to effectively block attack signals while maintaining the normal functionality of the voice assistant [18, 42, 58, 66]. Moreover, as mainstreamed Vas are usually closed systems, it is difficult to integrate and deploy a software-based solution. On the other hand, hardware-based defense solutions typically require modifications to commercial devices or the integration of additional active components. The former demands significant time and cost investments, while the latter may compromise the portability and practicality of mobile devices and likewise face reliability challenges in complex environments [18, 56, 59, 66].

Recent advancements in acoustic metamaterials [13, 15, 23, 33, 35, 37, 38, 69–71] present a promising alternative to conventional defense strategies of VAs. Acoustic metamaterials manipulate sound waves through meticulously designed passive physical structures, enabling them to selectively block attack signals within specific frequency ranges while ensuring the normal operation of VAs, offering exceptionally high reliability. Unlike software-based solutions, acoustic metamaterials can effectively interfere with attack signals before they reach the microphone and do not rely on the device's operating system. This allows for broad deployment, even on closed-system devices. Their passive and compact nature enables seamless integration into smart device exteriors without requiring invasive modifications to the hardware, thereby preserving both circuit integrity and device portability. Figure 1 illustrates two typical defense scenarios that existing methods struggle to address effectively.

Although acoustic metamaterials offer the potential for protecting VAs from various attacks, developing a comprehensive defense system remains challenging. One major drawback of traditional acoustic metamaterials is their **narrow resonant frequency range**, which requires the combination of more than 13 units to effectively filter a broad spectrum of inaudible attack signals. This significantly compromises device portability. Additionally, when defending

*Corresponding author

against *adversarial attacks*, protective measures can interfere with the recognition accuracy of legitimate voice commands within overlapping frequency ranges, reducing usability in real-world applications. Furthermore, variations in device shape and microphone placement of the target device make it challenging to integrate acoustic metamaterials.

We present MetaGuardian, a new VA defense system based on acoustic metamaterials. MetaGuardian is designed to overcome the aforementioned limitations of traditional acoustic metamaterials. First, it implements a filtering mechanism leveraging the mutual impedance effect between metamaterial units. This enables effective filtering of inaudible attacks in the 16–40 kHz range with just **three units**, significantly reducing structural complexity while maintaining a compact volume of only $0.795cm^3$. Second, MetaGuardian enhances defense against *adversarial attacks* by integrating labyrinth-style coiled acoustic metamaterial. This design selectively amplifies signals in the 2000–4000 Hz range, distorting critical frequencies to disrupt adversarial attacks while ensuring minimal impact on legitimate voice commands. Third, MetaGuardian provides a universal and portable design suitable for the two primary types of VA-equipped devices: mobile devices and smart speakers. The system adapts to structural and microphone placement characteristics, allowing for direct external mounting. Additionally, by strategically arranging metamaterials within reserved signal channels, the design ensures seamless transmission of legitimate commands while effectively disrupting attack signals.

We demonstrate that MetaGuardian can be manufactured using low-cost resin 3D printing technology and does not require users to train machine learning models. We evaluated MetaGuardian performance on nine smart devices against five adversarial attacks [11, 12, 52, 62, 63], three inaudible attacks [43, 53, 57], and one laser attack [48]. In a controlled evaluation environment , we show that MetaGuardian is compatible with various devices and can effectively defend against laser attacks. Moreover, within the effective attack distances identified by related studies, MetaGuardian successfully defends against all tested cases of five adversarial attacks and three inaudible attacks. Compared to existing defense solutions, MetaGuardian offers an innovative and efficient security protection mechanism for voice assistants.

This paper makes the following contributions:

- It presents the first acoustic metamaterial-based system that can effectively defend against inaudible, adversarial, and laser attacks without requiring software support or hardware modifications.
- It is the first to leverage mutual impedance effects to extend acoustic meta-matrial's filtering range to 16-40 kHz, blocking wide-band inaudible attacks while maintaining device functionality and portability.
- It demonstrates how a portable VA protection system can be built through low-cost 3D printing.

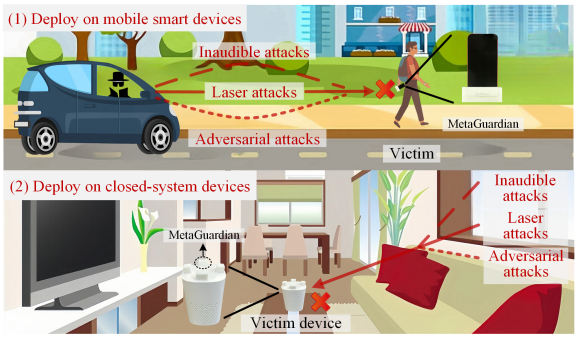


Figure 1: Example deployment scenarios of MetaGuardian: (1) protecting mobile devices in public spaces without compromising portability and passivity; (2) interfering and blocking attacks on closed-system smart speakers.

Online material. The 3D printing CAD files for MetaGuardian and demonstration videos of system deployment can be downloaded from 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 [10].

Adversarial attacks embed malicious audio into conversations or music to deceive voice assistants into executing unintended commands [16, 51, 54]. For example, CommanderSong [63] hides adversarial perturbations in songs, while VRIFLE [29] embeds them in user commands, enabling covert control.

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 [24, 41, 43, 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 [48].

2.2 Software-based Defenses

Software-based approaches have been proposed to counter VA attacks. They employ varied tactics to counter voice threats. For examples, EarArray [65] detects inaudible attacks via signal timing differences across microphones. NormDetect [28] 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 [50] blocks them with real-time perturbations.

Software-based 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 many 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 5.4.1) [28, 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 [28, 43, 57]. Moreover, as shown in Table 1, many commercial smart speakers restrict access to audio data for security reasons [28, 30, 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 [19] uses an additional speaker array to interfere with and block inaudible attacks. VocalPrint [25] 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 [44] uses a throat microphone to distinguish the user's voice from external speaker signals.

As hardware-based defences 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 [19, 25]. In addition, some hardware defenses depend on bulky, power-hungry components, such as speaker arrays or millimetre-wave radars [19, 25]. 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 [41, 43, 57]. 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

MetaGuardian is based on acoustic metamaterials and does not require changes to the software and hardware systems of the end-user devices. Using the macroscopic design of their internal structures, acoustic materials can modify the phase

and amplitude of sound waves within particular frequency ranges [26, 31, 61, 67], thus providing the potential to disrupt adversarial and inaudible attacks. Moreover, acoustic metamaterials can be constructed from opaque resin materials, which endows them with the ability to prevent the penetration of laser attacks.

Unlike software and hardware defenses, acoustic metamaterials interact with sound waves purely through their passive physical structure, require no power supply, are compact in size, and can be placed externally to the device's microphone, thereby circumventing the limitations of traditional solutions. However, they still face challenges such as expanding the filtering frequency range, maintaining device functionality, and achieving seamless integration across different devices, making it difficult to develop a comprehensive defense system.

Metamaterials vs. analog filters . Acoustic metamaterials act like analog filters that can effectively block acoustic attacks. 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 System Design for MetaGuardian

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 entails 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 portability, 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 use Helmholtz-like resonators to filter ultrasound, but their narrow filtering bandwidth limits their ability to cover a wide range of inaudible attack frequencies. To address this limitation, we propose a solution based on the mutual impedance effect, which expands the bandwidth of the metamaterial units, enabling comprehensive defense against inaudible attacks.

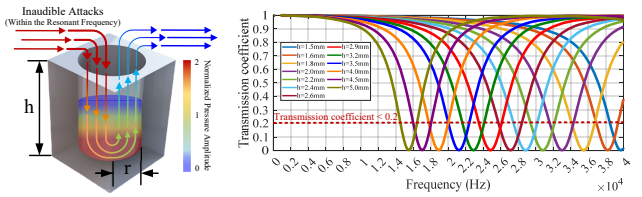


Figure 2: (a) Helmholtz-like acoustic metamaterial unit, (b) filtering effect of 13 units in 16-40 kHz.

3.1.1 Principles and Narrowband Limitations of Metamaterials. Acoustic metamaterials similar to Helmholtz resonators take advantage of their unique geometric structure to resonate with specific ultrasonic frequencies, allowing efficient filtering of ultrasound waves [31]. As shown in Figure 2a, these acoustic metamaterials consist of a cylindrical cavity and a circular neck. When external sound waves enter the resonator, their energy interacts with the air inside the cavity, leading to significant absorption of sound wave energy near the resonant frequency and greatly attenuating the energy of external sound waves passing through the cavity.

Therefore, to match the resonance frequency of ultrasound, it is necessary to design an appropriate cavity structure of the acoustic metamaterial. The resonance frequency f_0 of the acoustic metamaterial can be calculated using the following formula [32]:

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

where v is the speed of sound in air (typically 343 m/s), h is the depth of the cylindrical cavity, and r is the radius of the narrow neck. By adjusting these parameters, highly efficient filters targeting specific ultrasonic frequency bands can be precisely designed.

However, the resonance frequency of acoustic metamaterials is highly dependent on the precise matching of their geometric structure, which limits a single fixed-structure metamaterial unit to filtering a relatively narrow frequency range. Although LLOYD et al. [31] pointed out that when $r = 1.5$ mm, acoustic metamaterials can achieve a wider filtering bandwidth, the resonance of a single metamaterial unit is still confined to a frequency range of 1-2 kHz. In contrast, the frequency range of inaudible attacks spans a much broader range of 16-40 kHz. As shown in Figure 2b, to reduce the transmission coefficient to approximately 0.2 and effectively defend against such a wide range of inaudible attacks [31], around 13 metamaterial units are required. This significantly increases system complexity and dramatically reduces portability, making it difficult to integrate with devices.

3.1.2 Broadband Filtering via Mutual Impedance Effect. Recent studies [31, 46] have shown that mutual impedance can tune the resonant frequency and broaden the frequency range. Inspired by this, we leverage this effect to achieve broadband filtering with fewer units, breaking the limitations of traditional methods.

Specifically, mutual impedance enhances the total impedance Z_{total} of the system, which influences the resonant frequency through the following equation: $f_r = \frac{1}{2\pi}\sqrt{\frac{1}{m_{\text{eff}} Z_{\text{total}}}}$, where m_{eff} represents the effective mass, reflecting the inertia of the structure under a specific vibration mode. It is determined jointly by the material and structure of the metamaterial unit. As Z_{total} increases, the resonant frequency extends toward the lower frequency range, thereby expanding the overall frequency range. The total impedance Z_{total} of the system can be expressed as:$Z_{\text{total}} = \sum_{i=1}^{N} Z_i + Z_{\text{mutual}}$. This indicates that enhancing the mutual impedance effect is essential for expanding the resonant frequency range.

To achieve this, we further investigated the correlation between the strength of the mutual impedance effect and the spatial configuration of metamaterial units, leading to a critical finding: *The separation and arrangement of the units significantly impact the mutual impedance effect.*

The effect of unit spacing on mutual impedance. We have ascertained that the magnitude of the mutual impedance effect is intricately associated with the distance S that separates the units. Specifically, the mutual impedance exhibits an inverse proportionality to this distance. This correlation is articulated by the following equation:

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

wherein S signifies the distance between adjacent units. We posit that the fundamental cause of this correlation is attributable to the influence of S on the coupling effect between units. Reduced separation between units intensifies the coupling effect, thereby augmenting the impedance interaction and the resultant mutual impedance.

To further substantiate this relationship, we constructed a spatial correlation model based on acoustic coupling theory. Analogous to the mutual inductance theory in electromagnetics, the acoustic mutual impedance can be represented as the spatial integral of the sound pressure fields of adjacent units:

$$Z_{\text{mutual}} = \frac{1}{j\omega} \int_V \left(p_1 \cdot v_2^*\right) dV \tag{3}$$

Here, p_1 represents the sound pressure radiated by the first unit, v_2^* is the complex conjugate of the vibration velocity of the adjacent unit, and ω is the angular frequency. When the unit spacing is much smaller than the wavelength of the sound wave ($S \ll \lambda$), the sound pressure field approximately follows a spherical wave decay ($p \propto 1/S$), meaning that the closer the distance, the stronger the sound pressure. Additionally, the vibration velocity is in phase with the sound pressure, indicating that an increase in sound pressure simultaneously enhances the vibration velocity. Therefore, the integration result satisfies Equation 2, confirming the inverse relationship between mutual impedance and distance.

In addition, we developed an equivalent RLC circuit model, treating each metamaterial unit as a resonant RLC circuit and using mutual inductance M to represent the mutual impedance:

$$Z_{\text{mutual}} = j\omega M = j\omega \left(k\sqrt{L_1 L_2}\right) \tag{4}$$

Here, the coupling coefficient $k \propto 1/S$, which is also inversely related to the distance. This model further provides theoretical support for the regulation of mutual impedance.

To optimize the mutual impedance effect, we set S to **0.1 mm**, a distance that maximizes the mutual impedance while meeting the precision requirements of 3D printing, ensuring that the units do not overlap and avoiding structural interference.

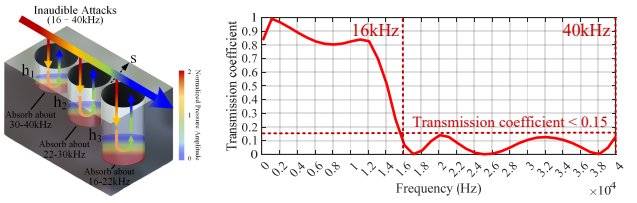
The effect of unit arrangement on mutual impedance. The arrangement of the units also plays a key role in the strength of the mutual impedance effect. Through an analysis of different arrangements, we found that a linear arrangement significantly enhances the mutual impedance effect. This is because a linear arrangement only involves direct coupling between adjacent units, thereby avoiding the weakening of the mutual impedance effect caused by complex interactions among multiple units in more intricate layouts, such as circular arrangements. The mutual impedance under different configurations can be expressed as follows:

$$Z_{\text{mutual}} = \begin{cases} \frac{1}{S} \cdot \alpha, & \text{Linear} \\ \frac{1}{S}\left(\alpha \cdot f_{\text{loss}}(N)\right), & \text{Circular} \end{cases} \tag{5}$$

Where α is the coupling coefficient, which depends on the unit arrangement structure, material properties, and the surrounding medium of the units. In a linear arrangement, α is primarily determined by direct coupling between adjacent units. For a circular arrangement, there exists multi-path interference between units, and the mutual impedance weakening factor $f_{\text{loss}}(N)$ can be expressed as:

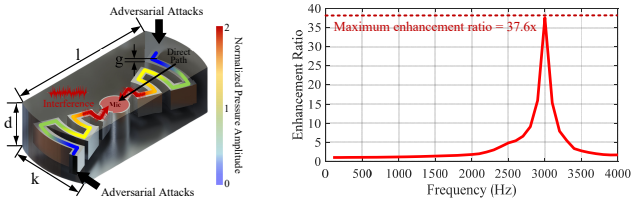
$$f_{\text{loss}}(N) = \frac{1}{N} \sum_{m=1}^{N} \sin^2\left(\frac{\pi m}{N}\right) \tag{6}$$

As the number of metamaterial units N increases, the function $f_{\text{loss}}(N)$ decreases, indicating that phase mismatches between non-adjacent units cause energy loss, weakening the mutual impedance effect in circular arrangements. Moreover, $f_{\text{loss}}(N)$ satisfies $0 < f_{\text{loss}}(N) < 1$, showing that additional coupling in circular arrangements reduces overall



(a) (b)

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



(a) (b)

Figure 4: (a) Adversarial Attack Defense Metamaterial(AADM), (b) its interference effects.

mutual impedance. To verify this, we tested filtering effects for different arrangements and spacings in experiments (see Section 5.2.1).

By linearly arranging metamaterial units with a spacing of 0.1 mm, COMSOL simulations show that the resonance frequency range of a single unit expands nearly fourfold. Three units of different heights ($h_1 = 2mm$, $h_2 = 3.2mm$, $h_3 = 4.8mm$) were selected to cover the inaudible attack frequency band of 16–40 kHz, forming the Inaudible Attack Defense Metamaterial (IADM). As shown in Figure 3b, simulation results indicate that the IADM reduces the ultrasonic wave transmission coefficient in this band to below 15%, demonstrating significant defense effectiveness.

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 [11, 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) [8, 22], while the 2000-4000 Hz range mainly carries consonant details, contributing only about 10% of the total speech information entropy ($H_{\text{high}}/H_{\text{total}} \approx 10\%$) [20, 49]. 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 [27, 39, 55]. 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:

$$x_{\text{adv}}(t) = x_{\text{clean}}(t) + \delta(t), \quad P_{\delta}(f) \ll P_{x_{\text{clean}}}(f). \tag{7}$$

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:

$$\tilde{\delta}(t) = \mathcal{F}^{-1}\{H(f)\Delta(f)\}. \tag{8}$$

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. In Section 5.2.2, we provide empirical evidence showing the advantage of MetaGuardian over direct filtering.

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:

$$f_r = \frac{c}{4L_{\text{coiled}}} \tag{9}$$

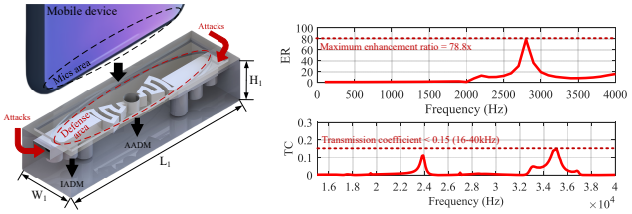
where c denotes the speed of sound in air, which is 343 m/s, and L_{coiled} represents the length of the coiling path within the metamaterial. As the frequency of the sound wave approximates the resonant frequency, the metamaterial demonstrates its most potent energy response, thereby amplifying signals 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 9, 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:

$$G = \frac{n_r}{\lambda_0} \cdot \sqrt{\frac{2\rho c^2}{\lambda_0^2}} \tag{10}$$

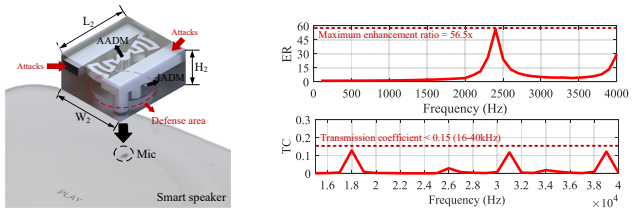
In this context, the refractive index $n_r = \frac{L_{\text{coiled}}}{L_{\text{blue}}}$ is defined as the quotient of the propagation speed of sound waves within the metamaterial and their speed in air. By calculating the path length ratio shown in Figure 4a, this refractive index can be estimated. When 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



(a) (b)

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



(a) (b)

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

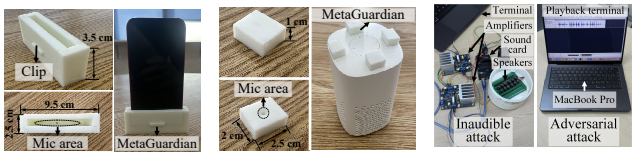
verified in Section 5.2.2 and Section 5.3.1 that AADM effectively defends against adversarial attacks while maintaining the integrity of legitimate audio signals.

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 Universal Structure Design for Mobile Devices. When designing a universal MetaGuardian 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 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



(a) (b) (c)

Figure 7: MetaGuardian prototypes for (a) mobile devices, (b) smart speakers, and (c) attack devices.

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 [7, 17]. 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 Universal 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 universal 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.

Additionally, 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 Evaluation

5.1 Experimental Setup and Methodology

Test targets. To comprehensively and accurately evaluate the performance of MetaGuardian in defending against attack signals, we reproduced three types of inaudible attacks with different center frequencies: NUIT [53] (18 kHz), DolphinAttack [57] (25 kHz), and LipRead [43] (40 kHz), effectively covering the typical attack range of 16-40 kHz. Additionally, we reproduced five authoritative open-source adversarial attacks: ALIF (2024)[12], KENKU (2023)[52], SMACK (2023)[62], CommanderSong (2018) [63] and Devil's Whisper (2020) [11] as well as a laser attack, Light Commands [48] (we verified its ability to penetrate MetaGuardian using a laser pointer and a photosensor). Detailed information on these systems is presented in Table 2.

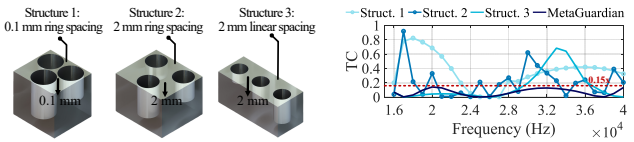
Test devices. To verify the broad applicability of MetaGuardian, we selected five smartphones and four smart speakers for protection effectiveness testing, covering well-known brands such as Apple [4], Google [5], Xiaomi [6], Huawei [2], and Amazon [3]. The selected devices include flagship models and highly practical products from these brands in recent years, spanning different types, usage scenarios, and price ranges, and are widely used in personal and home environments. This selection aims to ensure MetaGuardian's compatibility and effectiveness across various

Table 2: Tested on nine authoritative attack systems.

System name	Attack type	Compatible devices
KENKU [52]	Adversarial	iPhone 16 Pro, Pixel 8 Pro, Echo Dot 5th, HomePod mini
SMACK [62]	Adversarial	iPhone 14 Pro, Echo Dot 5th
ALIF [12]	Adversarial	iPhone 14 Pro, Pixel 8 Pro, Echo Dot 5th
CommanderSong [63]	Adversarial	iPhone 14 Pro
Devil's Whisper [11]	Adversarial	iPhone 16 pro, Pixel 8 Pro Echo Dot 5th, HomePod mini
DolphinAttack [57], NUIT [53], LipRead [43]	Inaudible	All devices
Light Commands [48]	Laser	All devices

Table 3: Tested on nine models from six VAs.

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



(a) (b)

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

hardware and environments. Detailed specifications of all test devices are listed in Table 3.

Evaluation metrics. We use three distinct evaluation metrics to comprehensively assess the performance of MetaGuardian across multiple attack scenarios and usage conditions.

Protection Success Rate (PSR) measures defense performance against MetaGuardian by counting failed attacks from 30 attempts with each system.

Word Interference Rate (WIR) gauges MetaGuardian's interference on attack command keywords by the ratio of destroyed to total words.

Commands Recognition Rate (CRR) through evaluates MetaGuardian's impact on legitimate commands by the ratio of recognized to total commands.

Experiment design. We conducted experiments to evaluate the defense capabilities of MetaGuardian. Adversarial attacks used 65 dB voice commands, including *Open the door, Play music, Make a call, What's the time, Send a message, Turn on the light, Transfer money, Airplane mode on, Navigate to my office,* and *Make a credit card payment.* Inaudible attacks transmitted the same commands using a 3-watt ultrasonic speaker. The attack devices are shown in Figure 7c. Experiments were conducted in an open laboratory with a background noise level of approximately 43 dB to minimize environmental interference and signal loss.

Table 4 details the experiments. Experiments A1 and A2 validated MetaGuardian's practical applicability. Experiments B1-B3 evaluated its defense against adversarial, inaudible, and laser attacks. Experiments B4-1, B4-2, and B5-1, B5-2 assessed its handling of complex attacks. Experiment C compared MetaGuardian's advantages to existing defense strategies.

5.2 System Filtering Performance and Legitimate Signal Impact

5.2.1 A1 - Filtering Effect of MetaGuardian . To verify the optimized MetaGuardian's ability to efficiently filter ultrasonic signals, we used the Avisoft-Bioacoustics CM16/CMPA to measure the ultrasonic signal strength passing through it and compared the effects of different unit arrangements and spacings on filtering (see Figure 8a).

As shown in Figure 8b, MetaGuardian performs excellently in filtering, consistent with COMSOL simulations. The ring structure with 0.1 mm and 2 mm spacing (Structure 1) is effective only in the 25-30 kHz range, while the linear structure with 0.1 mm spacing (Structure 3) is effective only in the 16-28 kHz range. The results indicate that using a linear arrangement with reduced spacing can significantly enhance the mutual impedance effect.

5.2.2 A2 - Impact on Normal Usage . Before evaluating MetaGuardian 's defense performance, it is essential to ensure it does not interfere with the input and output of legitimate commands. Therefore, we tested its impact on standard commands such as *What is the weather*, *Play music* and *Open the door*, using voice samples synthesized by Google Cloud TTS [1] and real speech from 10 male and 10 female volunteers.

As shown in Figure 9, devices equipped with MetaGuardian successfully responded to all voice commands, and the commands played back were accurately recognized by other devices, achieving a 100% command recognition rate. These results indicate that devices integrated with MetaGuardian can operate voice assistant and audio playback functions normally, ensuring a good user experience.

5.3 Overall Defense Performance

We evaluated the performance of MetaGuardian against various attacks under controlled conditions. It is important to note that the results of MetaGuardian were obtained in

Table 4: Experimental includes two improvement verifications, seven defense evaluations, and one comparison.

Test objectives	Label	Test focus	Description
Filtering performance	A1 (Sec.5.2.1)	Filtering effect of MetaGuardian	We tested MetaGuardian's 16-40 kHz filtering across different unit arrangements.
Impact on normal usage	A2 (Sec.5.2.2)	Impact on normal functionality	We measured recognition accuracy during input and playback to assess impact.
Defense performance	B1 (Sec.5.3.1)	Adversarial attack defense capability	We tested five MetaGuardian-integrated devices against five adversarial attacks.
	B2 (Sec.5.3.2)	Inaudible attack defense capability	Nine MetaGuardian-equipped devices were tested against three inaudible attacks.
	B3 (Sec.5.3.3)	Laser attack defense capability	MetaGuardian's laser defense was tested using a laser pointer at different angles.
	B4 (Sec. 5.3.4)	Multi-angle defense capability	We evaluated MetaGuardian's multi-angle defense against attacks.
	B5 (Sec. 5.3.5)	Precision interference in attacks	We evaluated MetaGuardian's effectiveness in disrupting command keywords.
	B6 (Sec. 5.3.6)	Anti-interference capability	We tested MetaGuardian 's defense under environmental interference.
Compared to prior work	C1 (Sec. 5.4.1)	Prior work's reliability affected	To validate our viewpoint, we tested the cross-device reliability of prior work.
	C2 (Sec.5.4.2)	Advantages of MetaGuardian	We compared MetaGuardian with existing defense strategies in various aspects.

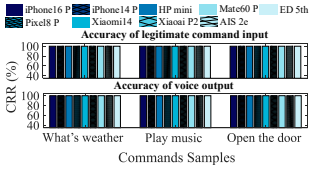


Figure 9: Impact on commands input & playback.

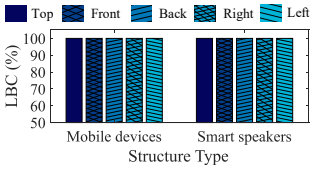


Figure 10: Laser light-blocking coefficient (LBC)

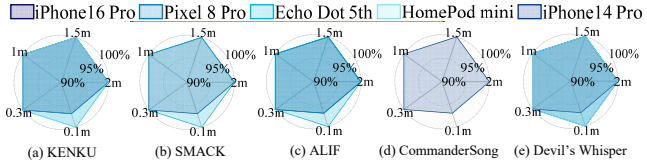


Figure 11: Adversarial attack defense at various ranges.

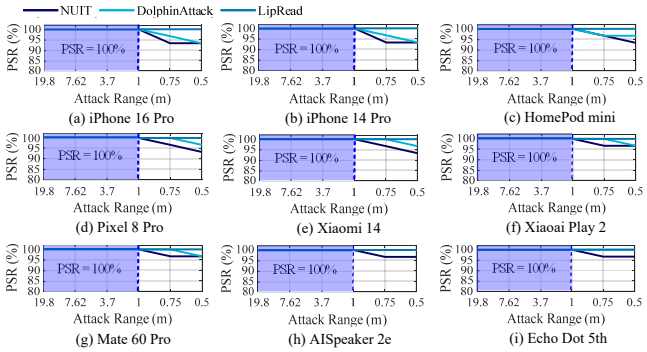


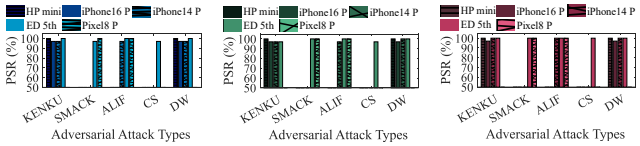
Figure 12: Inaudible attack defense at various ranges.

a controlled environment, which minimizes some variables like user movement and background noise. Performance of MetaGuardian in unconstrained settings may be influenced by these additional factors.

5.3.1 B1 - Adversarial Attack Defense Capability . We evaluated MetaGuardian against five representative adversarial attacks (Table 2) on five VA-enabled devices. Figure 11 presents the attack success rates (PSR) in this setting. At distances where these attacks typically achieve high success, including KENKU [52] (70% at 0.3 m), CommanderSong [63] (82% at 1.5 m), SMACK [62] (64.7% at 0.5 m), Devil's Whisper [11] (90% at 2 m), and ALIF [12] (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 effectively disrupts adversarial signals while preserving the recognition of legitimate commands.

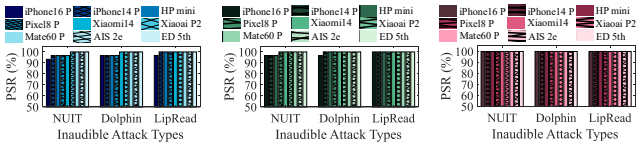
5.3.2 B2 - Inaudible Attack Defense Capability . To evaluate MetaGuardian 's effectiveness against inaudible attacks, we tested nine devices at various distances and recorded the PSR in a controlled environment. The results are shown in Figure 12. Within the maximum effective ranges of three common attacks, 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.

5.3.3 B3 - Laser Attack Defense Capability . 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 [48]) to illuminate two MetaGuardian structures from five angles, and measured



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

Figure 13: Adversarial attack defense at various angles.



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

Figure 14: Inaudible attack defense at various angles.

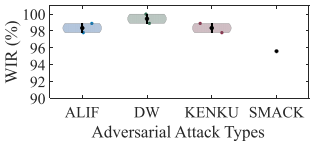


Figure 15: WIR against adversarial attacks.

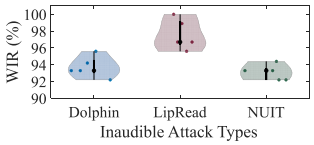


Figure 16: WIR against inaudible attacks.

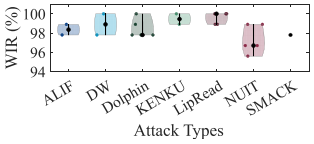


Figure 17: WIR in a noisy environment.

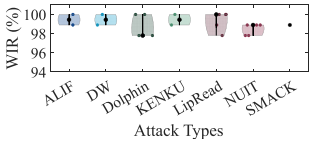


Figure 18: WIR in a mobile environment.

their light-blocking coefficients with a TA636A light sensor to evaluate the protective effect. The results are shown in Figure 10. At all tested angles, 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.

5.3.4 B4 - Multi-angle Defense 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 13 and 14. 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.

5.3.5 B5 - Precision Interference in Adversarial and Inaudible Attacks . When defending against multi-keyword attacks, precise interference with each keyword is essential. To evaluate MetaGuardian 's effectiveness, we launched adversarial and inaudible attacks at 0.1 m and 0.5 m on various speech-to-text mobile devices (iPhone 16 Pro, iPhone 14 Pro, Pixel 8 Pro, Xiaomi 14, Mate 60 Pro) and measured the WER, as shown in Figures 15 and 16. MetaGuardian achieves a WER over 95% in adversarial attacks (deviation $\leq$ 5%) by dispersing and absorbing keyword signal energy to hinder recognition. In inaudible attacks, it maintains WER above 92.5% (deviation $\leq$ 7%), demonstrating stable, effective defense against complex attacks.

5.3.6 B6 - Anti-interference Capability . To evaluate the system's anti-interference capability in outdoor conditions, we conducted tests in an environment with approximately 75 dB ambient noise. Volunteers carrying devices equipped with MetaGuardian moved at a speed of 2 m/s while attacks were launched, and the word identification rate (WIR), a higher-is-better metric, was measured, as shown in Figures 17 and 18. MetaGuardian achieved a WIR of 98% against adversarial attacks and over 95% against inaudible attacks in noisy conditions. While in motion, the WIR for both attack types exceeded 97%, demonstrating high reliability. MetaGuardian employs a passive structure that requires no signal analysis. By altering the phase of sound waves through its material properties, it nonselectively interferes with specific frequencies. This approach is inherently resistant to variations in noise, temperature, and other environmental factors, enabling stable and continuous protection.

5.4 Compared to Prior Work

5.4.1 C1 - Reliability Impacted by Microphone Differences of Prior Work . 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 [25, 28, 43, 57, 65]. We selected the classic LipRead method [43] for testing (other defense methods use similar signal feature extraction approaches). Under the same environment, the "turn

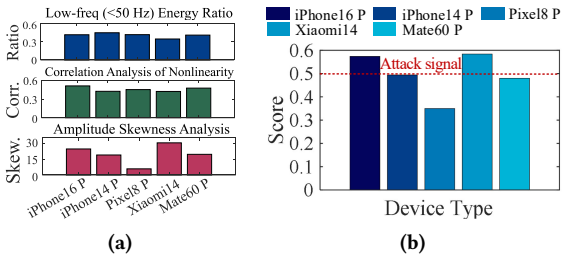


Figure 19: (a) Feature responses of different devices to the same command, (b)comprehensive scores.

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 19b). In contrast, MetaGuardian defends the microphone directly with a physical structure, avoiding the impact caused by hardware differences.

5.4.2 C2 - Advantages of MetaGuardian . We conducted a comparative analysis of MetaGuardian and recent defense approaches to evaluate its advantages in usability. 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 5.4.1, 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.

Existing hardware-based defense methods, such as AIC [19], VocalPrint [25], and the approach proposed by Sahidullah et al. [44], 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.

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 [43]	No	No	Yes	Yes	No
NormDetect [28]	No	No	Yes	Yes	No
EarArray [65]	No	No	Yes	Yes	No
VoShield [60]	No	No	Yes	Yes	Yes
AIC [19]	Yes	No	Yes	No	No
VocalPrint [25]	Yes	Yes	Yes	No	Yes
Sahidullah et al. [44]	Yes	Yes	Yes	No	Yes
MetaGuardian	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**

6 Discussion and Future Work

Mismatch with dynamic attacks . Current system have fixed filtering or amplification bands, making them ineffective against adaptive attacks like frequency hopping, and exposing critical vulnerabilities against complex threats.

Portability limitations . While these designs offer some protection, they are often bulky, heavy, and unattractive, reducing portability and user experience. They are unsuitable for scenarios requiring compactness and discretion.

Electromagnetic interference (EMI) defense . The current design remains vulnerable to EMI attacks, which can inject malicious signals without using the acoustic channel, weakening the effectiveness of MetaGuardian and similar acoustic-based defenses.

Impact on ultrasonic sensing . Filtering out the 16–40 kHz band may interfere with ultrasonic sensing functions in modern voice assistants, such as proximity detection, gesture recognition, and acoustic analysis, thereby affecting user experience with these features.

Future directions. To address current limitations, future work can explore tunable acoustic metamaterials using piezoelectric materials or shape memory alloys, enabling real-time, electrically controlled impedance adjustment to balance defense and sensing, and adapt to dynamic or frequency-hopping attacks. To resist EMI attacks, shielding or active suppression can be incorporated to build a multilayer defense for enhanced practicality and security.

7 Conclusion

We have presented MetaGuardian, an acoustic metamaterial-based VA protection system that blocks inaudible, adversarial, and laser attacks without software support. By leveraging mutual impedance, it expands the filtering range and reduces size, enabling frequency-targeted defense while maintaining audio transmission. Its adaptable design supports diverse devices. Experiments confirm MetaGuardian is effectively in protecting VA systems across attack types and hardware platforms, making it a reliable, practical solution.

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.google-mobile.cn/. Google. Last accessed: 2025-1-20.

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

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

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

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] 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).

[30] 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.

[31] 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.

[32] Joshua S Lloyd, Cole G Ludwikowski, Cyrus Malik, and Chen Shen. 2023. Mitigating Inaudible Ultrasound Attacks on Voice Assistants With Acoustic Metamaterials. *IEEE Access* (2023).

[33] 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.

[34] 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.

[35] 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.

[36] 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.

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

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

[39] 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.

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

[41] 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.

[42] 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.

[43] 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.

[44] 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.

[45] 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).

[46] 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.

[47] 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.

[48] 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.

[49] 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.

[50] 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.

[51] 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.

[52] 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.

[53] 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.

[54] 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.

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

[56] 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.

[57] 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.

[58] 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*.

[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] 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).

[70] 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.

[71] 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).