From One Form of Energy to Another: Laser-Induced Injection Attacks on Acoustic Sensing

LUPENG ZHANG, Nanyang Technological University, Singapore
MINHAO CUI, Seoul National University, South Korea
WENWEI LI, Peking University, China
XUEFU DONG, The University of Tokyo, Japan
QING WANG, Delft University of Technology, Netherlands
LEI WANG, Dalian University of Technology, China
DAQING ZHANG, Institut Polytechnique de Paris, France
LILI QIU, The University of Texas at Austin, USA
JIE XIONG*, Nanyang Technological University, Singapore

Wireless sensing has gained significant research interest in recent years. However, the focus has primarily been on improving sensing performance, such as enhancing accuracy, and little attention has been paid to the security aspects of wireless sensing. In this paper, we demonstrate that acoustic signal–based sensing, widely regarded as the safest wireless sensing modality due to its physical characteristics, can be stealthily attacked and compromised. The core of this stealthy injection attack lies in exploiting the "photoacoustic effect" to convert the energy of laser light into vibrations, inducing acoustic signals, which are then used to compromise acoustic sensing systems. It is important to note that the laser can be invisible to the human eye, and the acoustic signals it generates are inaudible to human ears. To make the attack even stealthier, we propose the concept of "Smartphone-Defined Laser" to use commodity smartphones to control low-cost laser hardware ($0.60), eliminating the need for bulky and expensive signal generators. Through hardware and software co-design, we successfully demonstrate the attack with smartphones and cheap laser hardware. Comprehensive experiments show that the proposed attack can compromise the state-of-the-art acoustic sensing techniques (both chirp-based and continuous wave-based), achieving a high average success rate of 96.1% across different tasks even when the target is 50 m away. We hope our findings raise awareness of the security risks associated with acoustic sensing and encourage further research into enhancing its security.

CCS Concepts: • **Human-centered computing → Ubiquitous and mobile computing**.

Additional Key Words and Phrases: Photoacoustic effect, Acoustic sensing, Injection attacks

*Corresponding author

Authors' Contact Information: Lupeng Zhang, Nanyang Technological University, Singapore, Singapore, LUPENG001@e.ntu.edu.sg; Minhao Cui, Seoul National University, Seoul, South Korea, minhaocui@snu.ac.kr; Wenwei Li, Peking University, Beijing, China, liwenwei@pku.edu.cn; Xuefu Dong, The University of Tokyo, Tokyo, Japan, dongxuefu@mcl.iis.u-tokyo.ac.jp; Qing Wang, Delft University of Technology, Delft, Netherlands, qing.wang@tudelft.nl; Lei Wang, Dalian University of Technology, Dalian, China, lei.wang@dlut.edu.cn; Daqing Zhang, SAMOVAR Lab, Telecom SudParis, Institut Polytechnique de Paris, Paris, France, dqzhang@sei.pku.edu.cn; Lili Qiu, The University of Texas at Austin, Austin, TX, USA, lili@cs.utexas.edu; Jie Xiong (corresponding author), Nanyang Technological University, Singapore, Singapore, jie.xiong@ntu.edu.sg.

ACM 2474-9567/2026/6-ART72
https://doi.org/10.1145/3810215

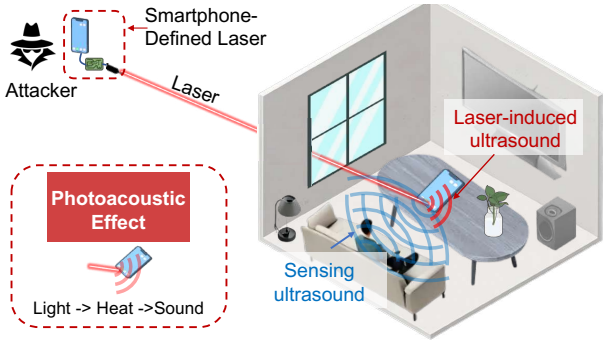


Fig. 1. Attack scenario of our proposed LightSaber: The attacker remotely and covertly compromises acoustic sensing via the photoacoustic effect by injecting laser signals with a smartphone-defined laser system.

ACM Reference Format:
Lupeng Zhang, Minhao Cui, Wenwei Li, Xuefu Dong, Qing Wang, Lei Wang, Daqing Zhang, Lili Qiu, and Jie Xiong. 2026. From One Form of Energy to Another: Laser-Induced Injection Attacks on Acoustic Sensing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 10, 2, Article 72 (June 2026), 30 pages. https://doi.org/10.1145/3810215

1 INTRODUCTION

Over the past decade, wireless sensing has garnered increasing attention from both academia and industry. This new sensing technique enables a wide range of meaningful applications in the realm of human-computer interaction [8, 30, 52], smart health [7, 15, 65], and environmental monitoring [11, 12, 42]. Various wireless signals have been exploited for sensing, including Wi-Fi [50, 56, 64], mmWave [49, 58, 67], LoRa [21, 53, 54], and acoustic signals [7, 51, 65]. Among them, acoustic-based wireless sensing exhibits the greatest commercialization potential due to its fine sensing granularity and the widespread availability of devices equipped with speakers and microphones. Some existing commercial devices, such as Google Nest [14], Amazon Echo [3], and Xiaomi Sound Pro [1], have already integrated acoustic sensing capabilities.

While most research has focused on improving sensing performance, significantly less attention has been given to the security aspects of wireless sensing—an oversight that can lead to serious consequences. For instance, in RF-based health monitoring systems, an attacker could obstruct the detection of a patient's breathing pauses, potentially resulting in life-threatening consequences. Although a few studies revealed security flaws in RF-based wireless sensing [20, 55, 69] by injecting an attack signal, similar threats to acoustic sensing have rarely been discussed. This is because acoustic sensing is generally considered more resilient to remote injection attacks, since the ultrasound signals used for injection attenuate quickly over distance [5]. This rapid attenuation makes it challenging to covertly inject malicious acoustic signals from a distance. Acoustic signals can also be easily obstructed, even by a transparent glass wall.

In this paper, we ask this question: *"Is acoustic sensing really as secure as commonly believed?"* By exploiting the photoacoustic effect, we demonstrate practical and stealthy injection attacks on acoustic sensing systems. The basic idea is shown in Figure 1. While ultrasound signals attenuate quickly, laser beams can propagate over long distances in the air with very little energy loss. Thus, our system leverages this property to remotely deliver energy to the target, where the laser beam causes a local temperature increase. When the energy intensity of the laser beam varies, the temperature variation causes mechanical vibrations of the target surface, which in turn generate sound [17]. What is exciting is that the characteristics (e.g., frequency and amplitude) of the generated sound can be precisely controlled by adjusting the parameters of the injected laser beam. In this way, we can use

the intangible[1] laser to stealthily deliver energy to the target, where it is converted into the attack acoustic signal to compromise acoustic sensing.

Prior studies [40, 66] have demonstrated laser-induced audio injection attacks on voice-command (VC) systems. In these settings, VC systems operate in a passive listening mode, which allows an attacker to inject a voice command that can be recognized at arbitrary times. In contrast, acoustic sensing systems actively transmit CW/FMCW sensing signals to infer target physical states (e.g., gestures and respiration) from subtle, time-varying perturbations in the reflected signals. Consequently, to attack an acoustic sensing system, an attacker must craft more delicate attack signals that both (i) closely resemble legitimate CW/FMCW sensing waveforms and (ii) are timing-constrained by the target devices' sensing timeline. Moreover, existing laser attacks typically rely on bulky and expensive laboratory equipment (e.g., function generators and bench-top laser drivers), which are also less stealthy for real-world attacks. As a result, prior works cannot be directly applied to acoustic sensing systems. Building a laser-based injection attack for acoustic sensing is non-trivial, and several challenges must be tackled before it becomes a functioning system. To compromise acoustic sensing, we consider two attack strategies: (i) *failure attacks* that inject random signals or noise to disrupt sensing, and (ii) *fake result attacks* that inject carefully designed signals to manipulate sensing outputs. We term these two attack goals "Failure Attack" and "Fake Result Attack", respectively.

Challenge 1: Performing fake result attacks without having knowledge of the legitimate signal is challenging. Compared to disrupting acoustic sensing with noise, injecting signals to make the acoustic sensing system generate fake results (e.g., causing the vital sign sensing system to obtain an incorrect respiration rate) is more challenging. To achieve a fake result attack, existing RF-based methods, such as those using Wi-Fi [20, 69] and radar [19], typically assume that the attacker is close to the target and can capture the legitimate signal. Then, based on the content and timing of the legitimate signal, attack signals are designed and injected. However, in our case, since we aim to attack the sensing system stealthily from a distance, the assumption that the attacker can obtain the legitimate signal is no longer valid.

In acoustic sensing, two different types of signals are commonly used: single-tone continuous wave (CW) and frequency-modulated continuous wave (FMCW). While the CW signal is primarily used to capture Doppler information, the FMCW (chirp) signal can capture fine-grained displacement (distance change) information.

Realistic CW Spoofing: For CW signals, the Doppler frequency shift can provide speed information for sensing. To generate a signal containing meaningful speed information (e.g., gesture motion speed), we reverse-engineer the sensing process. Specifically, based on the relationship between human motion and signal variation [35], we determine how to adjust signal parameters to generate signal samples that encode targeted motion information (e.g., motion speed, direction, and duration). However, the generated signal remains too "ideal", posing a risk of being detected by the sensing system. In real life, even during a single gesture, a human's speed is not constant but varies throughout the motion. Furthermore, when the same person performs the same gesture multiple times, variations in speed and motion duration naturally occur. To make the generated signal more closely resemble a true human-reflected signal, we propose the dynamic frequency shift warping (DFSW) method, inspired by the dynamic time warping (DTW) method. This method incorporates realistic variations into the generated attack signal to mimic real-life nonlinearity and diversity, making the attack more convincing and harder to detect.

FMCW Spoofing without Timing Knowledge: Unlike CW-based sensing, which relies solely on analyzing the received signals, FMCW-based systems utilize both the transmitted (TX) and received (RX) signals to extract time difference information [26]. This time difference information is then converted into distance for sensing. In our case, since the attacker is remote, it is hard to know the legitimate TX signal's timing to directly manipulate the sensing result. Therefore, we exploit the unique property of FMCW-based sensing, where sensing typically does not rely on absolute distance information, but rather on distance variation over time. This variation is

[1]The laser power is as low as 5 mW, which is not enough for the human body to feel any heat.

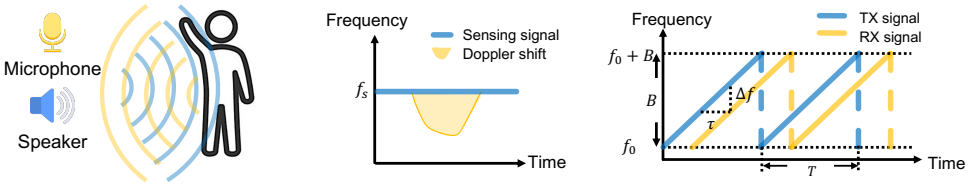
obtained by periodically sending chirps (each chirp provides one distance estimate) and calculating the distance multiple times each second. With this observation, we can eliminate the need to know the timing of the legitimate signal. Instead, we can manipulate the distance information obtained at the sensing system by intelligently adjusting the time intervals between the injected malicious chirps. In this way, by controlling the time interval, we can inject the results we want (i.e., fake results) without knowing the timing of the legitimate signal.

Challenge 2: It is challenging to find a practical and stealthy device capable of effectively driving a modulated laser to attack acoustic sensing systems remotely. To effectively attack an acoustic sensing system, we need to precisely control the generated acoustic signals by adjusting the laser signals. We initially considered using microcontroller boards such as Arduino [4] or Raspberry Pi [47]. However, they are unable to generate the required signals with sufficiently accurate timing and Doppler shifts due to their low clock accuracy and the limited number of bits available to represent the signal. On the other hand, high-performance arbitrary waveform generators (AWGs) can produce high-quality signals for our attack. However, they are expensive and bulky, making them impractical and not stealthy for real-world attacks.

To address this challenge, we propose the novel concept of Smartphone-Defined Laser that leverages the most ubiquitous mobile devices ever created—smartphones—to modulate laser signals for the proposed attack. The key insight enabling smartphone control of the laser driver is that the acoustic signals generated by the laser through the photoacoustic effect fall within the operating frequency range of the smartphone's audio module. Based on the photoacoustic effect, the frequency of the generated acoustic signal depends on the rate of change of the laser intensity. Therefore, we only need a device capable of generating alternating voltage signals at the frequency of acoustic sensing signals (e.g., $18 - 22\ kHz$). This objective aligns well with the capability of a smartphone's audio module. Thus, on the hardware side, we simply need to connect the standard 3.5 mm audio jack of the smartphone to a laser driver. On the software side, we modify the LibAS [46] acoustic sensing framework into an attack tool, further lowering the barrier to executing the proposed attack. This design effectively turns a smartphone into a versatile and powerful controller for carrying out the attack.

We implement our system LightSaber using multiple smartphones including the low-end Redmi 14C ($68.00), a cheap commodity laser driver ($11.80) and a 5 mW infrared laser ($0.60). We conduct attacks on state-of-the-art CW-based gesture recognition systems (EchoWrite 1.0 [52], 2.0 [71]) and an FMCW-based breathing monitoring system (LaSense [27]). Results show that we can achieve an effective stealthy attack on both CW-based and FMCW-based acoustic sensing. We evaluate the system on various low-cost lasers (ranging from $0.60 to $5.00) with different wavelengths and power levels, as well as on multiple smartphones (Samsung S21 Ultra, Pixel 7Pro, Redmi K30S, and Redmi 14C), demonstrating the versatility and adaptability of our system. The key contributions are summarized below:

- We propose LightSaber, which leverages the photoacoustic effect to uncover critical security vulnerabilities in acoustic sensing. Our findings show that acoustic sensing is not as secure as commonly believed.
- We propose separate signal processing pipelines for CW and FMCW sensing, respectively, enabling practical and stealthy remote attacks without the need to access legitimate signals.
- We propose the novel concept of a Smartphone-Defined Laser, which leverages ubiquitous smartphones to drive lasers and execute the proposed attack. We release the software at https://github.com/lupengz/LightSaber to benefit the community. We believe this concept can inspire the research community to explore smartphones as a viable attack platform.
- We evaluate the effectiveness of the proposed attack under various scenarios, using different smartphones and lasers. Comprehensive real-life experiments show that a stealthy attack can be achieved even at a distance of 50 m. Additionally, we discuss potential countermeasures to enhance the security of acoustic sensing systems.



(a) Acoustic sensing setup (b) Doppler shift spectrogram of the right-swipe gesture (c) ToF between TX and RX chirp signals

Fig. 2. Two main types of acoustic sensing.

2 BACKGROUND

In this section, we first introduce the principles of acoustic sensing and then present an overview of the photoacoustic effect.

2.1 Acoustic Sensing

A typical acoustic sensing setup is shown in Figure 2(a). It consists of a device with a speaker to emit acoustic signals and a microphone to capture the reflected signals from the target. We present the theoretical foundations of the two main types of acoustic sensing (CW-based and FMCW-based) to explain the rationale behind our design.

CW-based acoustic sensing. Acoustic sensing systems commonly use single-tone continuous wave (CW) signals to capture human motion by analyzing the Doppler shift caused by target movement. CW signals are widely used in gesture recognition [52, 71], driving behavior recognition [59, 62], and human presence detection (Amazon Echo [3]). Doppler shift refers to the frequency shift in the signals reflected from the target. Doppler shift can be expressed as:

$$\Delta f = f_s \times \left(\frac{v_s + v_r}{v_s - v_r}\right) - f_s = \frac{2 f_s v_r}{v_s - v_r} \approx \frac{2 f_s v_r}{v_s}, \tag{1}$$

where f_s, v_s, and v_r represent the frequency of the sensing signal, speed of sound, and radial speed of the target to the microphone, respectively. Since v_s ($\sim 340\ m/s$) is much larger than v_r, we can make the approximation in Equation 1. Figure 2(b) illustrates the Doppler shift pattern when a user performs a right-swipe gesture using a hand. By analyzing Doppler shift patterns, sensing systems can recognize the corresponding gestures.

FMCW-based acoustic sensing. FMCW signals, also known as chirp signals, are waves whose frequencies changing over time. FMCW signals can estimate distance by measuring the time-of-flight (ToF), making them widely used in object proximity detection (e.g., Google Nest [14]) and vital sign monitoring [27, 32, 33, 65]. Figure 2(c) illustrates multiple chirp signals, and they can be represented as:

$$x(t) = \cos\left(2\pi\left(f_0 t + \frac{B}{2T} t^2\right)\right), \tag{2}$$

where f_0, B, and T are the initial frequency, bandwidth, and duration of a chirp, respectively. A target at distance R reflects the chirp back, introducing a round-trip delay $\tau = 2R/v_s$, where v_s is the speed of sound. At the receiver, the received (delayed) chirp is mixed with the transmitted chirp and passed through a low-pass filter. This produces a low-frequency mixed signal whose frequency Δf is proportional to the delay τ: $\Delta f = \frac{B}{T}\tau$. Thus, by estimating Δf (e.g., via FFT on the mixed signal), the system can recover τ and compute the target distance as:

$$R = \frac{v_s}{2}\tau = \frac{v_s T}{2B}\Delta f. \tag{3}$$

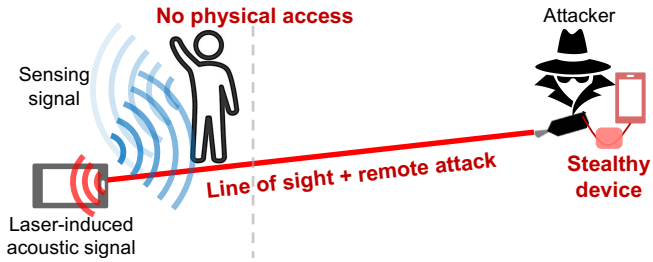


Fig. 3. The threat model of LightSaber.

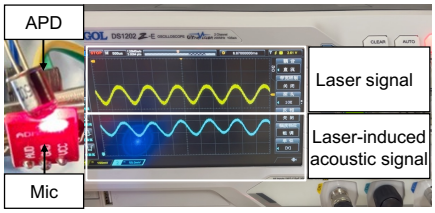


Fig. 4. Laser-induced acoustic waveforms.

2.2 Photoacoustic Effect

The photoacoustic effect is a physical phenomenon in which sound waves are generated when light, with varying intensity, strikes an object. When light signals with a modulation frequency[2] of f arrive at an object, the object's surface material periodically absorbs and releases optical energy, leading to thermal expansion and contraction, which generates sound waves as below:

$$v(t) \propto \alpha \cdot \frac{1}{\rho C} \cdot \frac{P}{2\pi f} \cdot \cos(2\pi f t + \varphi), \tag{4}$$

where $v(t)$ is the light-induced acoustic signal, α is the light absorption coefficient of the object, ρ is the density of the object, C is the specific heat of the object, P is the power of the light signal, f is the modulation frequency of the light signal, and φ is the phase of the modulation.

3 THREAT MODEL

In this section, we define the attack scenario, outlining the target victims, attack goals, and the attacker's capabilities and knowledge assumptions to clarify the attack's applicability and real-world feasibility.

3.1 Attack Scenario

As shown in Figure 3, an attacker discreetly uses a smartphone to control a laser, which remotely injects laser signals into the victim's sensing device.

Target victim. We assume that the victim is cautious and alert to unusual stimuli, such as unexpected sounds, visible light, or heat, which could raise suspicion. In line with the mainstream acoustic sensing systems [26, 27, 52, 71], we assume that the sensing device (*e.g.*, smartphone or laptop) is stationary during the sensing process. This is because acoustic sensing infers target context from signal variations, and signal variations induced by device motion become entangled with those caused by target movements, leading to sensing failures or significant accuracy degradation [29]. We additionally consider the case in which the victim's sensing device is held in hand, and therefore undergoes involuntary hand motions.

Attack goals. An attacker aims to achieve the following goals by injecting attack signals:

(1) Failure attack: The attacker injects signals that obscure legitimate sensing signals. These injected signals prevent the system from extracting meaningful information from legitimate sensing signals.

(2) Fake result attack: A more advanced and harmful attack is the fake result attack, where the injected signals cause the sensing system to produce incorrect results. For instance, in smart homes, an attacker can inject the laser through a glass window, causing an acoustic vital-sign sensing system to capture an incorrect respiration rate.

[2]The modulation frequency refers to the rate at which the intensity of the light signal is varied, rather than the frequency of the light signal.

3.2 Attacker Capability and Knowledge

We summarize the attacker's knowledge and capabilities as follows:

- No Physical Access: The attacker has no physical access to the victim's device.
- Device Parameters: Similar to prior works [24, 38, 40], the attacker can obtain a similar device or access publicly available product information to learn basic parameters such as signal type, frequency range, and general processing flow.
- Line of Sight: Similar to prior works [40, 66], we assume that the attacker has a line of sight to the target device. This is reasonable, as attackers can remain hidden and wait for an opportunity to strike. Additionally, glass windows and open spaces are common in many environments, making this assumption feasible.
- Attack Feedback: Like other remote attacks [40, 66], the attacker can leverage visual feedback. For example, in fake result attacks, if the injected signal successfully triggers a gesture command (e.g., unlocking a door), the physical response of door opening is observable.

4 FEASIBILITY OF PHOTOACOUSTIC ATTACKS

In this section, we conduct a feasibility study on remotely injecting malicious acoustic signals into the sensing system by leveraging the photoacoustic effect. We use an infrared laser to trigger the photoacoustic effect. In the preliminary experiments, we use a commodity 5 mW, 650 nm laser diode[3] as the transmitter and a Micro-Electro-Mechanical Systems (MEMS) microphone commonly used in acoustic sensing (SparkFun ADMP401) as the receiver. The preliminary experiment results are the following:

Signal strength. Based on the experiment results, the power of the laser-induced 18 kHz acoustic signal is around 32 dB when the laser source is 50 m away, comparable to an acoustic sensing signal at a distance of 0.5 m. This is because:

- *Negligible propagation loss:* The laser retains sufficient energy even over long distances.[4]
- *Strong directionality:* The laser beam is directed at the microphone, triggering the photoacoustic effect.
- *Microphone diaphragm:* The diaphragm's low density and high elasticity contribute to larger laser-induced acoustic signals, as described by Equation 4.

Signal frequency. We vary the laser intensity change rate in the range of $18 - 22$ kHz and confirm that the laser-induced acoustic signal has exactly the same frequency. Therefore, we can flexibly control the frequency of the acoustic signals generated by the photoacoustic effect by modulating the laser signal accordingly. We present the 18 kHz result in Figure 4.

5 LightSaber DESIGN

To conduct remote laser attacks on an acoustic sensing system by the photoacoustic effect, we need to address the following software and hardware challenges:

- **Challenge 1:** How to synthesize CW and FMCW attack signals that reliably cause the sensing system to produce attacker-desired outputs, without access to the legitimate signal?
- **Challenge 2:** How to find a practical and stealthy device capable of effectively modulating lasers to generate acoustic attack signals?

To address these challenges, we design the LightSaber attack, which integrates both software and hardware components, as illustrated in Figure 5. On the software side, we design two tailored signal generation strategies

[3]Note that to ensure the safety of the experimenter's eyes, we select a laser whose propagation path is invisible, while the light spot is visible in the benchmark experiments. Lasers with both propagation path and light spot invisible are tested in Section 7.
[4]Laser attenuation in air is approximately 2×10^{-5} dB/m [36], whereas 20 kHz ultrasonic waves attenuate at about 1 dB/m [31], making sound attenuation 5×10^4 times larger than laser attenuation.

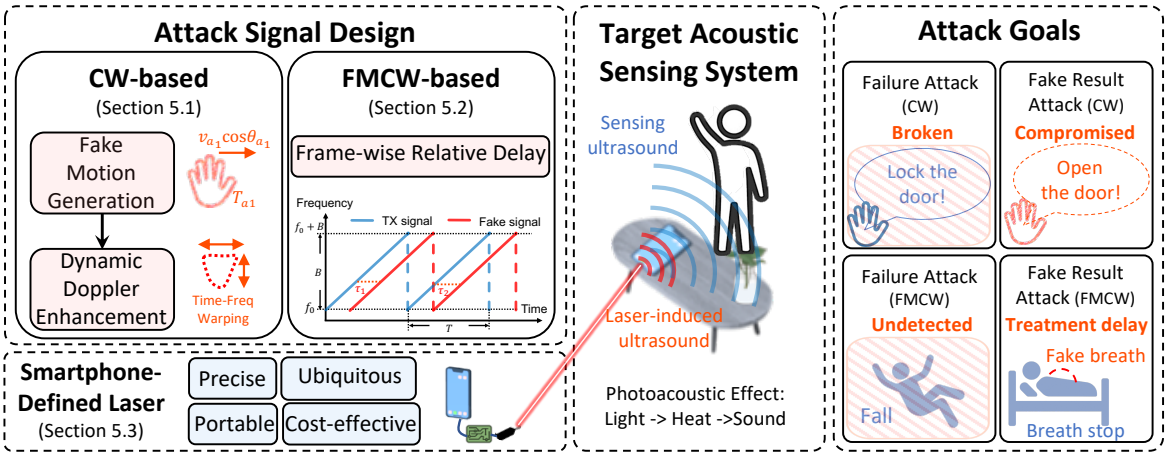
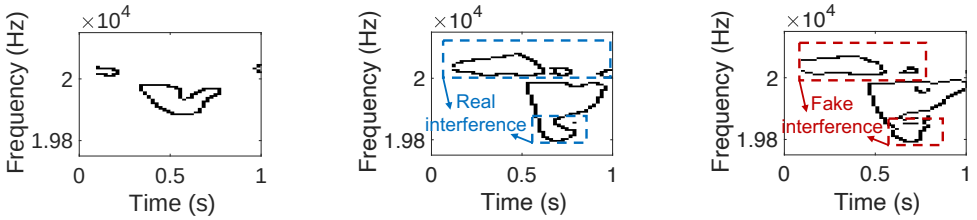


Fig. 5. Overview of the LightSaber Design. We design CW- and FMCW-based attack signals to spoof or disrupt sensing results, enabled by a smartphone-defined laser that injects ultrasound via the photoacoustic effect. These laser-induced signals are received by the target system, achieving failure (e.g., missed gestures or fall alerts) and fake result attacks (e.g., spoofed gestures or breathing), without physical contact or access to legitimate signals.



(a) W/o interference (b) W/ real interference (c) W/ fake interference

Fig. 6. Sensing failure examples of CW-type.

for remotely attacking acoustic sensing systems: a CW-based approach leveraging motion models and dynamic Doppler enhancement (Section 5.1), and an FMCW-based strategy based on relative delay manipulation (Section 5.2). On the hardware side, we propose the concept of a Smartphone-Defined Laser, enabling a precise, portable, ubiquitous, and cost-effective attack paradigm (Section 5.3).

5.1 CW-based Sensing Attack

After validating the feasibility of using laser-induced acoustic signals for attack, we proceed to design attack signals for CW-based acoustic sensing systems.

5.1.1 Failure Attack. A straightforward way to cause sensing failure is by injecting strong Gaussian white noise to overwhelm the sensing signal. However, this method produces a distinct spectral pattern and an abrupt SNR drop, making the attack readily apparent.

Our goal is to subtly mislead the system by injecting interference to cause misclassification or sensing failure without exposing the presence of an attack. To achieve this, we replace obvious white noise with fake

"motion interference" signals measured in the presence of real-world interfering motions (e.g., background body movements, undefined gestures, and other environmental disturbances). For example, in gesture recognition, user gestures typically produce distinct time-frequency patterns. Figure 6(a) shows a spectrogram of a clean swipe gesture. In realistic environments, background motion may introduce additional frequency shifts, as shown in Figure 6(b), leading the system to misclassify the gesture or conclude that no valid gesture is present. Inspired by this, we collect signals from random body movements and reuse them as injected attack signals. By overlaying these signals onto the legitimate received waveform during sensing, we create frequency patterns like the one in Figure 6(c), which successfully mislead the system and cause sensing failure.

5.1.2 Fake Result Attack. In CW-based acoustic sensing, the target's speed is inferred from the Doppler shifts in the received signal. Therefore, generating Doppler shifts that convey meaningful motion information is crucial for our fake result attack. While generative models have been employed to create fake signals in RF sensing attacks [37], they require large amounts of high-quality training data and retraining for different environments. Other approaches, which manually design attack signals [20, 69], rely on metasurfaces or relays near the sensing device to capture, delay, or modify legitimate signals. *Note that these methods do not apply to our laser-induced remote attack, where the attacker does not have access to the legitimate signals.*

Factors affecting Doppler shift. To generate synthetic Doppler shift patterns that accurately mimic real-world motions, we study the factors that affect Doppler shift patterns: *(1) Sensing setup:* The relative position between the target and the device can affect the Doppler shift pattern. *(2) User diversity:* For the same gesture, the user's motion speed and duration can vary. As shown in Figure 7, different users and setups produce significantly different Doppler shift patterns for the same sliding gesture. The Doppler frequency shift can be modeled using the real-time hand radial velocity as:

$$\Delta f = \frac{2 f_s v_r(t)}{v_s} = \frac{2 f_s v_h(t) \cos \theta_h(t)}{v_s}, \tag{5}$$

where $v_h(t)$ is the speed of the hand and $\theta_h(t)$ is the angle between the hand's movement direction and the line from the hand to the microphone. As shown in Figure 7(a) and 7(c), different setups (e.g., variations in phone location and orientation) lead to distinct signal variations, even when the same gesture is performed. Additionally, different users' habitual gesture speeds $v_h(t)$ and duration further impact the amplitude and duration of the frequency shift.

Signal generation. Based on the relationship between the Doppler shift and signal variations, we can generate attack signals that mimic the Doppler patterns of specific movements (e.g., a particular gesture). We use Figure 8 to illustrate our mimic model. The attacker sets the motion duration $[T_{a_1}, T_{a_2}]$, motion speed $v_a(t)$, and motion direction $\theta_a(t)$ based on reasonable physical constraints (e.g., gesture speed typically ranges from 1 to 3 m/s). With these parameters, the attacker can derive the real-time frequency of the attack signal as:

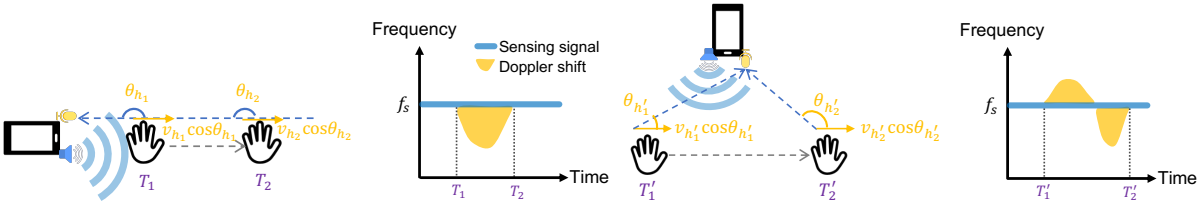
$$f_a(t) = f_s + \frac{2 f_s v_a(t) \cos \theta_a(t)}{v_s}. \tag{6}$$

The attack signal can then be generated as:

$$a(t) = A_0 \cos\left(2\pi \int_{T_{a_1}}^{t} f_a(\tau)\, d\tau\right), \quad t \in [T_{a_1}, T_{a_2}] \tag{7}$$

where A_0 is the signal amplitude.

Design of dynamic frequency shift warping. To enhance stealthiness, a straightforward approach is to adopt the Dynamic Time Warping (DTW) scheme to warp frequency shift patterns. As shown in Figure 9(a), we use DTW to warp an "ideal" frequency shift pattern generated by our model. However, DTW only adjusts shifts along the time axis and does not account for realistic frequency variations. We propose the Dynamic Frequency Shift Warping (DFSW) method, which applies nonlinear warping to both the time and frequency dimensions, as



(a) User A's setup and motion (b) User A's Doppler shift (c) User B's setup and motion (d) User B's Doppler shift

Fig. 7. Impact of user and setup variability on Doppler shift patterns.

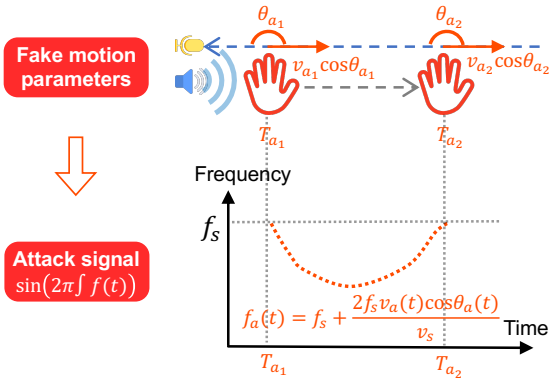
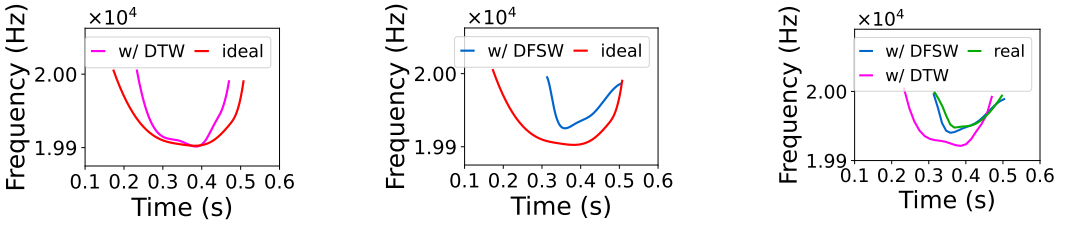


Fig. 8. Illustration of proposed mimic model.



(a) DTW: time-only (b) DFSW: dual-axis (c) Comparison of different warping

Fig. 9. Comparison of frequency shift warping methods.

shown in Figure 9(b). This approach makes the generated signal more closely resemble real sensing signals. The detailed steps of DFSW are presented in Algorithm 1.

Compared to DTW, which finds an optimal alignment path $\phi(t)$ under constraints (monotonicity, continuity), DFSW introduces stochastic, nonlinear, and parameterized distortions in both time and frequency domains with physical interpretations.

- **Continuity and smoothness:** The warping functions $\Delta t(t, f)$ and $\Delta f(t, f)$ are designed to be continuous and differentiable, ensuring smooth variations in the time-frequency space rather than the abrupt, piecewise transitions often seen in DTW alignments.

Algorithm 1: Dynamic Frequency Shift Warping (DFSW)

Input : Attack signal $a(t)$, time distortion ratio μ, total duration T_a, frequency distortion ratio γ, center frequency f_s, decay coefficient λ

Output : Warped time-frequency representation $F'(t, f)$

1. Compute original time-frequency representation:
   $F(t, f) \leftarrow \text{STFT}(a(t))$; *// STFT: Short-Time Fourier Transform*
2. **Time-domain warping:**
   Sample $\beta_t \sim \mathcal{U}(-\mu, \mu)$; *// $\mathcal{U}(-\mu, \mu)$: Uniform distribution controlling time warping intensity*
   Compute $\Delta t(t, f) \leftarrow \beta_t \cdot T_a \cdot g_t(t, T_a)$; *// $g_t(\cdot)$: General nonlinear time-warping function*
3. **Frequency-domain warping:**
   Sample $\beta_f \sim \mathcal{U}(-\gamma, \gamma)$; *// $\mathcal{U}(-\gamma, \gamma)$: Uniform distribution controlling frequency warping range*
   Compute $\Delta f(t, f) \leftarrow \beta_f \cdot f_s \cdot g_f(f, \lambda)$; *// $g_f(\cdot)$: General nonlinear frequency-warping function*
4. **Apply DFSW:**
   $F'(t, f) \leftarrow F\left(t + \Delta t(t, f),\ f + \Delta f(t, f)\right)$;

return $F'(t, f)$;

- **Time-frequency two-dimensional-warping:** The goal is to apply realistic warping in both time and frequency dimensions to introduce nonlinearity and diversity. The warping functions $g_t(\cdot)$ and $g_f(\cdot)$ are general models that can be tailored to specific motions (e.g., walking, hand waving) based on physical principles.
- **Stealthiness and diversity enhancement:** Directly generated Doppler signals are often overly idealized and lack real-world variability. DFSW mitigates this by stochastically sampling the parameters of $g_t(\cdot)$ and $g_f(\cdot)$, introducing structure-preserving randomness that enhances signal diversity and stealthiness without sacrificing physical plausibility.

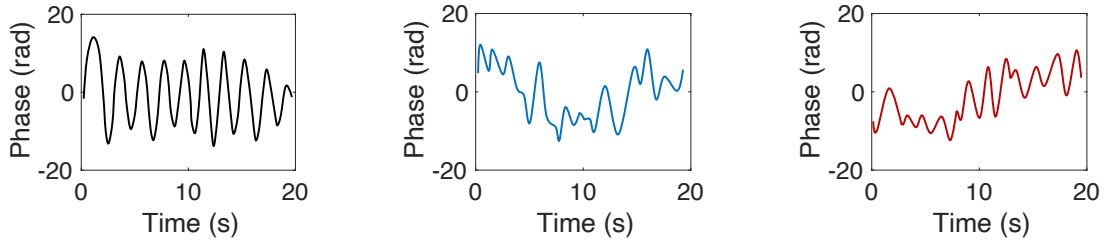
5.2 FMCW-based Sensing Attack

After designing attack signals for the CW-based acoustic sensing system, we proceed to design signals for another widely used FMCW-based system.

5.2.1 Failure Attack. Similar to the sensing failure strategy used in CW-based systems, we target FMCW-based sensing by exploiting non-target or random motion patterns to induce sensing failure. However, due to the specific properties of FMCW signals—where both legitimate and attack signals are chirps—there is no need to replicate exact signal patterns. Instead, we inject multiple interference chirps within each chirp frame to mimic the presence of multiple moving objects in the environment. To avoid triggering anomaly detection, the injected signals are designed to match the legitimate chirp's bandwidth B and duration T, which are typically public specifications or can be estimated using off-the-shelf devices. This ensures that the injected signal is both realistic and stealthy.

We take respiration monitoring as an example. Under normal conditions, the device transmits periodic chirps, which reflect off the human body and produce stable phase variations at the receiver, as shown in Figure 10(a). These variations can be used to estimate respiration rates. However, real-world disturbances—such as nearby human movement or background motion—create dynamic multipath effects that distort the phase trajectory, making it chaotic (Figure 10(b)) and preventing reliable signal interpretation. Following this insight, we generate our attack by injecting crafted chirp-based signals that simulate similar disturbances. The resulting received signal becomes unstable, as shown in Figure 10(c), misleading the system and causing it to fail in extracting accurate respiration rates.

5.2.2 Fake Result Attack. Similar to the failure attack, we design attack signals with bandwidth B and duration T for fake result attacks to ensure that the chirp signal remains consistent with legitimate sensing signals.



(a) W/o interference (b) W/ real interference (c) W/ fake interference

Fig. 10. Sensing failure examples for FMCW-based sensing.

Unlike CW-based sensing, which relies solely on the received signals, FMCW-based sensing depends on both the transmitted (TX) and received (RX) signals to calculate the time-of-flight (ToF) between them for sensing [26]. Therefore, if an attacker injects fake RX signals without knowledge of the timing of the TX signal, they cannot manipulate the ToF information to spoof the sensing results.

The principle of FMCW-based sensing. *Note that FMCW-based acoustic sensing systems typically do not rely on absolute distance information, but rather on distance variations over time.* Distance change, i.e., displacement, contains the complete information required for most applications. As shown in Figure 11, using breathing monitoring as an example, the sensing system does not rely on the absolute position of the chest but instead tracks its relative motion over time to infer the breathing rate.

Distance variation-based attack. Based on the insight above, we propose a signal injection strategy for FMCW-based sensing. As shown in Figure 11, the attacker stealthily injects fake signals with the same duration T and bandwidth B as the legitimate signal to avoid detection. Without access to the timing of the legitimate TX signal, the injected fake signal has a random delay δ relative to the TX signal. The insight here is that while this random delay prevents us from obtaining the correct absolute distance, it does not hinder our ability to manipulate the targeted relative distance over time. Take respiration sensing as an example. To create a fake breath motion, we need to inject a series of chirps that simulate the periodic displacement variation between the chest and the sensing device. We inject the first chirp *chirp 1* with a random delay of δ. We do not care about the absolute distance extracted at the receiver using *chirp 1*. We then inject the second chirp *chirp 2*. Note that we can control the interval (τ_1) between *chirp 1* and *chirp 2*. By adjusting the intervals between adjacent injected chirps (i.e., τ_2, τ_3,..., τ_n) as shown in Figure 11, we can control the variation of the measured distances, effectively simulating the motion of the human chest. This allows us to generate fake respiration patterns without knowing the timing of the legitimate signal.

5.3 Smartphone-defined Laser

To mount an effective acoustic attack, we need to precisely shape and control the laser's amplitude to produce the desired acoustic attack waveform. Low-cost embedded platforms (e.g., Arduino [4], Raspberry Pi [47]) are attractive for their size and availability, but their limited clock stability and DAC resolution (low bit depth) prevent them from synthesizing the high-fidelity, low-distortion waveforms our attack requires. On the other hand, arbitrary waveform generators (AWGs) can generate highly accurate signals, yet they are bulky, expensive, and operationally conspicuous; moreover, their MHz–GHz capabilities far exceed the tens-of-kHz bandwidth typical of acoustic sensing, making them overkill for this task. Therefore, we need a pragmatic middle ground—a solution that balances sufficient precision and waveform flexibility with compactness, low cost, and stealth.

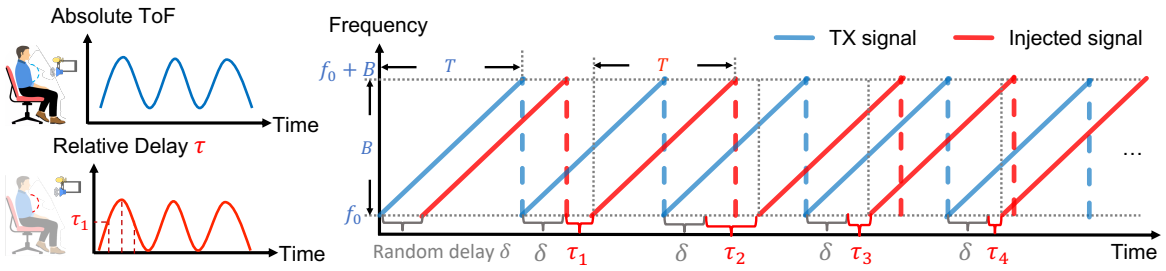
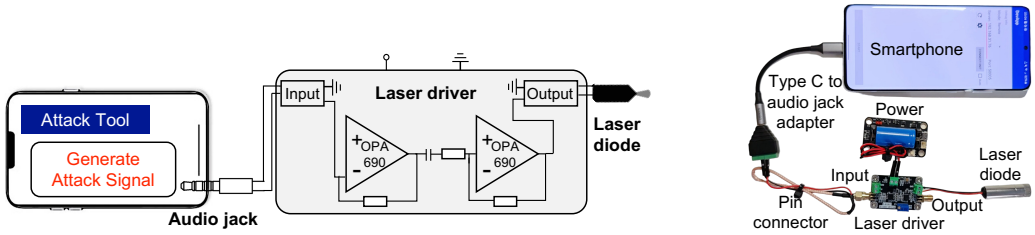
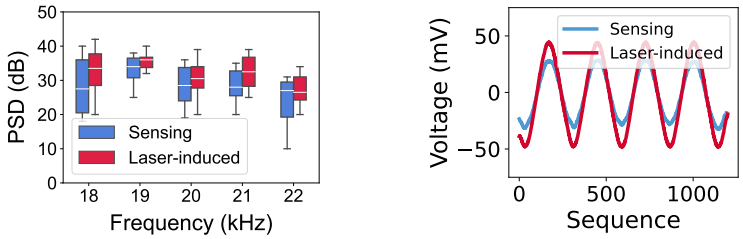


Fig. 11. Illustration of the relative distance variation-based FMCW attack strategy.



(a) Schematic (b) Prototype

Fig. 12. Smartphone-defined laser system.



(a) Signal energy (b) Signal quality

Fig. 13. Quality of Smartphone-defined laser-induced signal.

5.3.1 Concept of Smartphone-defined Laser. We exploit smartphones, the most ubiquitous mobile devices ever created, to modulate laser signals for the proposed attack. Because the induced acoustic signal's frequency follows the laser intensity modulation rate, we only need a device capable of generating an alternating (AC) voltage in the acoustic sensing band (typically $18 - 22$ *kHz*). This band lies squarely within the frequency response of smartphone audio hardware. Furthermore, modern smartphones can output relatively high-quality analog signals, making them well-suited to drive the laser modulation for the attack. Thus, on the hardware side, the smartphone's standard 3.5 *mm* audio output is interfaced to a laser driver. On the software side, we adapt the LibAS acoustic-sensing framework [46] by replacing its sensing-signal generator with an attack-signal generator (Figure 12(a)). This hardware–software combination yields a compact, low-cost, and easy-to-deploy attack tool that substantially lowers implementation complexity.

5.3.2 Effectiveness of Smartphone-defined Laser. To validate whether the Smartphone-Defined Laser can achieve the proposed attack goals, we systematically compared laser-induced acoustic signals, real acoustic sensing signals reflected from the target, and acoustic attack signals directly generated by a speaker. In this experiment, we built the Smartphone-Defined Laser system, as shown in Figure 12(b), which consists of a smartphone, a commodity laser driver ($11.80), and a 5 mW, 650 nm laser diode ($0.60), to generate and transmit attack signals. We tested the proposed attack using various smartphones and laser diodes in Section 7.5.

Comparison with legitimate sensing signals. To compare legitimate sensing and laser-induced attack signals, we conduct an experiment where the victim is 0.5 m from the sensing device, while the attacker injects FMCW-type attack signals from a 10 m distance using the Smartphone-Defined Laser. Both signals are collected using the same receiver. As shown in Figures 13(a) and 13(b), we compare the received signals in terms of amplitude and signal similarity across different frequencies. The power spectral density (PSD) analysis confirms that the energy distribution of the laser-induced signals closely aligns with that of legitimate sensing signals. Notably, we can easily adjust the laser diode's supply voltage to match the amplitude of the acoustic sensing signal. Figure 13(b) presents the measured 18 kHz attack signals, showing that the waveform and voltage of our attack signals are similar to the legitimate signals.

6 IMPLEMENTATION

Prototype of LightSaber. We implement our LightSaber prototype, as shown in Figures 14 and 15. On the hardware side, the attacker smartphone generates the analog attack signals and outputs them as an audio voltage signal through a USB-C to a 3.5 mm audio adapter (e.g., a UGREEN Type-C to 3.5 mm adapter with built-in DAC [2], $7.73). The audio signal is then fed into a low-cost commodity laser driver board employing the TI OPA690 operational amplifier [44] ($11.80), which converts the input waveform into a corresponding modulation of the laser-diode drive current. For long-distance experiments, an optional TI LM386 audio amplifier [43] ($0.22) can be inserted between the smartphone's audio output and the laser driver input to increase the input signal amplitude. The laser driver is powered by commodity batteries, and we vary the supply according to the attack distance; for example, at 50 m we use a battery pack made of three 18650 cells (11 V), while the driver board supports a 5–12 V supply range. The modulated signal is finally emitted by commodity laser diode modules operating at 650 nm or 980 nm, with output power ranging from 5 mW to 180 mW (typically $0.60–5.00 per module). To evaluate generalizability across victim devices, we test four commodity smartphones (Samsung S21 Ultra, Pixel 7 Pro, Redmi K30S, and Redmi 14C), whose retail prices range from $68.00 to $749.00. We further validate the attack using six laser diodes that span both wavelengths and all power levels (Figure 17), to evaluate how laser characteristics affect attack performance. Note that the 650 nm laser is path-invisible but has a visible light spot, whereas the 980 nm laser is both path-invisible and light spot-invisible. For the completely invisible 980 nm laser, a co-aligned 650 nm visible laser spot can be used to find the direction, which then guides the alignment of the invisible laser beam. In open areas, i.e., the corridor in Figure 14 and the campus in Figure 15, we employ a 650 nm laser diode that produces only a small visible spot (to avoid prolonged exposure of experimenters to unknown laser beams) at 5 mW for safety.[5] For software, we modify the widely used LibAS [46] acoustic sensing framework into an attack tool for signal generation.

Victim sensing system reproduction. For CW-based acoustic sensing systems, we replicate the state-of-the-art gesture recognition systems, EchoWrite 1.0 (based on Dynamic Time Warping) [52] and EchoWrite 2.0 (based on Convolutional Neural Network) [71]. We construct a gesture dataset based on the setup of EchoWrite 1.0 [52]. Ten participants (5 males and 5 females) are recruited to perform six stroke gestures as shown in Figure 16. These gestures cover the fundamental motion primitives commonly used in gesture-based systems, while more complex gestures can be represented as combinations of these basic strokes. This results in a total of 3,600 testing

[5]In the U.S., 5 mW is the maximum allowed power for consumer-grade 650 nm laser pointers.

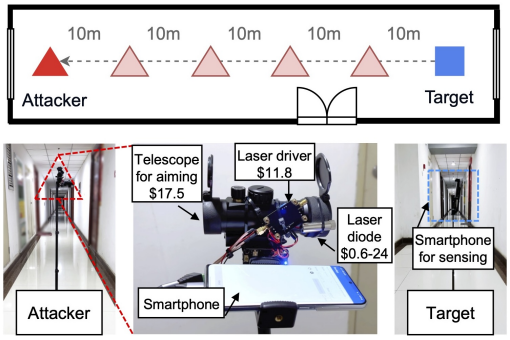


Fig. 14. Experiment setup for different attack ranges and LightSaber's prototype.

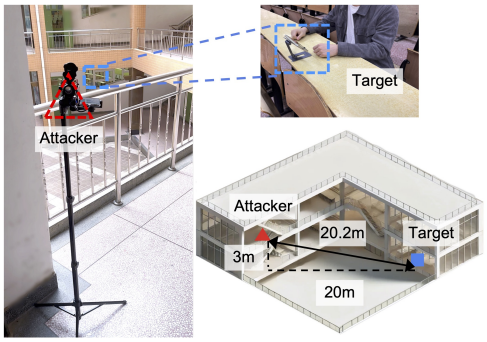


Fig. 15. Experiment setup for the low-power cross-building attack.

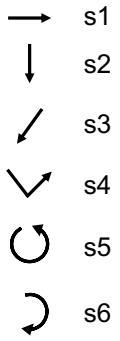


Fig. 16. Basic strokes.

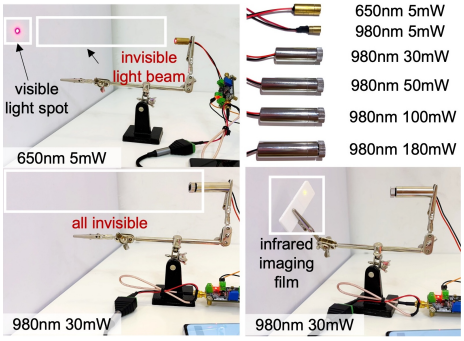
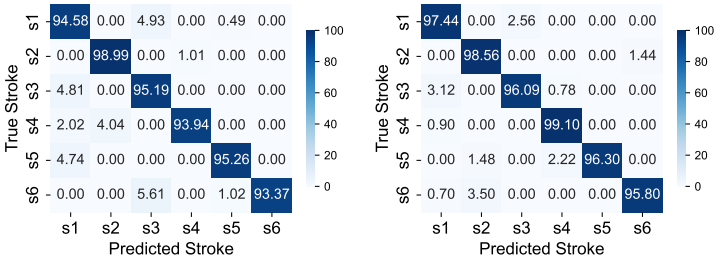


Fig. 17. Different laser diodes for attacking.



(a) Result of EchoWrite 1.0 (b) Result of EchoWrite 2.0

Fig. 18. Reproduction of CW-based sensing systems.

(a) Initial localization (b) Vertical alignment (c) Horizontal alignment

Fig. 19. Aiming workflow (visible laser spot shown for illustration only; actual attacks use an invisible 980 nm infrared laser).

instances. We then apply the same preprocessing pipeline as EchoWrite 1.0 and 2.0 to the collected dataset. For FMCW-based sensing systems, we replicate the state-of-the-art respiration monitoring system, LaSense [27]. Following LaSense, we use smartphones and the LibAS tool to implement the system with the same sensing parameters (frequency band 18 *kHz*-22 *kHz*, chirp duration 150 *ms*, and sampling rate 48 *kHz*).

Overall attack implementation workflow. (1) *Device setup and groundtruth measurement.* By default, the target sensing device (Samsung S21 Ultra) is positioned naturally for respiration monitoring and gesture recognition. For respiration experiments, ground-truth breathing rates are recorded using an Apple Watch Series 9, while for gesture experiments, ground-truth labels are obtained from manually annotated video recordings. The

attacker's equipment (Redmi 14C smartphone by default, laser driver, laser diode, telescope, and optical scope) is set up at the specified distance (e.g., 10 m). (2) *Signal generation.* The attacker uses our modified LibAS-based [46] attack tool together with a smartphone-defined laser platform to generate attack signals tailored to the attack objectives and signal design. (3) *Aiming and alignment.* We then align the laser to the target microphone using the following steps: i. Initial localization: Use a telescope to visually locate the victim device; ii. Coarse aiming: Widen the beam to a 10–20 cm spot to reliably illuminate the device body (as shown in Figure 19(a)). Note that the visible 650 nm laser is used solely for illustration. In real-world attacks, an attacker would use a 980 nm infrared laser, which is invisible to the naked eye and can only be observed using an infrared camera, enabling a stealthy aiming; iii. Vertical alignment: Gradually narrow the beam to a 5–10 cm spot (comparable to the width of the device's bottom side) and vertically steer it to ensure that the bottom side of the device, where the microphone is located, remains covered by the beam spot (as shown in Figure 19(b)); iv. Horizontal alignment: Further narrow the beam to a 0.5–1 cm spot (slightly larger than the typical 0.1–0.5 cm microphone openings, providing lateral tolerance) and horizontally steer it to ensure that the microphone opening is covered by the beam spot, as illustrated in Figure 19(c). The DIANA aiming scope in our system is an off-the-shelf optical scope that is co-mounted with the laser source to assist in localizing the invisible laser spot. Specifically, the relative offset between the laser emitter and the scope's crosshair is fixed, allowing us to infer the laser spot location directly from the observed crosshair position. (4) *Signal injection and metric computation.* With the legitimate sensing application running, we trigger the laser for attack signal injection.

7 EVALUATION

We first evaluate the performance of the attack on the CW-based system in Section 7.1 and on the FMCW-based system in Section 7.2. We then compare our method with speaker-based attacks in Section 7.3. After that, we evaluate the system's attack range in Section 7.4. Finally, we evaluate the system's generalizability using various smartphones and laser diodes in Section 7.5, as well as its robustness under different environmental conditions in Section 7.6.

Metrics. We use the Attack Success Rate (ASR), which measures the proportion of successful attacks among all attack attempts. We define success criteria separately for the CW-based gesture recognition system and the FMCW-based respiration monitoring system as follows.

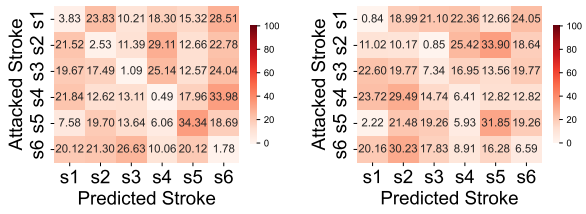
(1) CW-based gesture recognition system.

- *Failure attack:* The sensing system fails to recognize the victim's true gesture (i.e., the predicted label is different from the ground-truth label) [52, 71].
- *Fake result attack:* The sensing system outputs the attacker-desired gesture label [52, 71].

(2) FMCW-based respiration monitoring system.

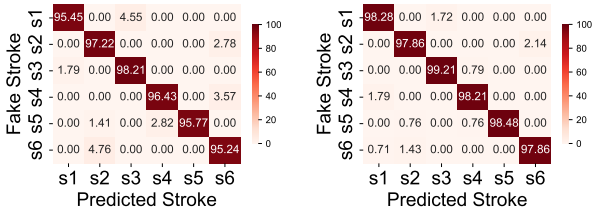
- *Failure attack:* We adapt two commonly used metrics to quantify sensing failures, corresponding to two stages of breathing detection: detection and monitoring. In the detection stage, the sensing system leverages the autocorrelation [64] of segmented signals to determine whether breathing motion is present. An autocorrelation value below an empirical threshold of 0.8 indicates that respiratory periodicity has been disrupted. In the monitoring stage, the system estimates the breathing rate. If the estimated rate deviates from the ground truth by more than 1 bpm [27], the respiration frequency estimation is considered to have failed.
- *Fake result attack:* The sensing system outputs the attacker-desired respiration rate; that is, the difference between the estimated breathing rate and the attacker-injected desired breathing rate is less than 1 bpm [27, 29].

Default evaluation setup. Unless otherwise specified, the attacker is located 10 m away and injects attack signals using a 5 mW, 980 nm laser, driven by a Redmi 14C smartphone, and the target device under attack is a Samsung S21 Ultra.



(a) EchoWrite 1.0 (b) EchoWrite 2.0

Fig. 20. Failure attack results in CW-based systems.



(a) EchoWrite 1.0 (b) EchoWrite 2.0

Fig. 21. Fake results in CW-based systems.

7.1 Attacks on CW-based Systems

Targeted sensing systems. To validate our implementation of the CW-based sensing systems, we replicate the evaluation of EchoWrite 1.0 [52] and EchoWrite 2.0 [71] using the same gesture set and experimental setup. Each experiment is repeated 50 times to ensure consistency. Figure 18 shows the confusion matrices of both systems. EchoWrite 1.0 achieves 95.23% accuracy using DTW, while EchoWrite 2.0 reaches 97.15% with a CNN-based model, confirming the reliability of our reproduction. These results establish a baseline for evaluating the effectiveness of our subsequent attack methods.

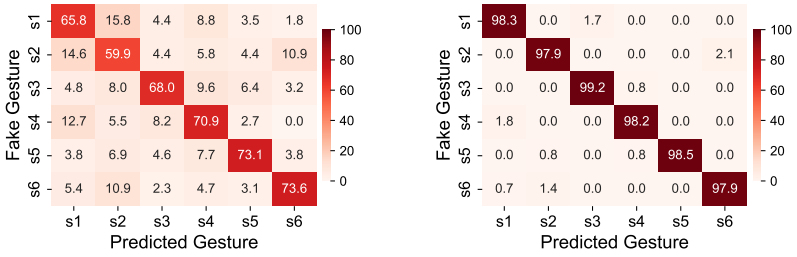
Failure attack. The victim is positioned 0.2 m from the sensing transceiver, while the attacker, located 10 m away, injects attack signals using a 5 mW, 650 nm laser. As shown in Figure 20, after injecting interference signals into the EchoWrite 1.0 and EchoWrite 2.0 systems, the average sensing accuracy drops to 7.57% and 8.97%, respectively. This demonstrates that our attack signals can effectively disrupt CW-based sensing, impacting both the DTW (signal processing-based) and CNN (deep learning-based) classifiers.

Fake result attack. Here, the attacker injects false signals from 10 m away to validate the fake result attack. As shown in Figure 21, our attack signals can effectively deceive both DTW-based and CNN-based sensing systems, achieving an average ASR of 96.47% and 98.30%. The attack remains consistently effective across different gestures.

Ablation study of DFSW. In both settings, we keep the same attack pipeline and laser power; the only difference lies in the injected signals. Figure 22 shows the confusion matrices for fake-result attacks in these two cases. Without DFSW (as shown in Figure 22(a)), the average classification accuracy across the six gestures is around 70%, and different target gestures are frequently confused with each other. With DFSW enabled (as shown in Figure 22(b)), the average classification accuracy increases to about 98%. This performance gap is mainly because, without DFSW, we inject a naive CW-based fake waveform that only roughly mimics the Doppler trend (e.g., an overall up-and-down shift), but does not follow the physically plausible time–frequency evolution of real hand motions (e.g., how Doppler changes with hand size, movement speed, and gesture trajectory). As a result, its Doppler pattern may deviate from the natural distribution of real gestures. It may be treated as noise or an out-of-distribution signal by the system's preprocessing step and classifier. Moreover, it lacks diversity across both the time and frequency domains, which further weakens the stealthiness. In contrast, with DFSW, we inject a carefully crafted attack waveform generated by our mimic-model-guided design, which follows hand-motion physics in both the time and frequency domains and introduces natural random variations. This design better matches the Doppler patterns of real gestures, leading to higher attack success rates while remaining harder to distinguish from legitimate sensing signals.

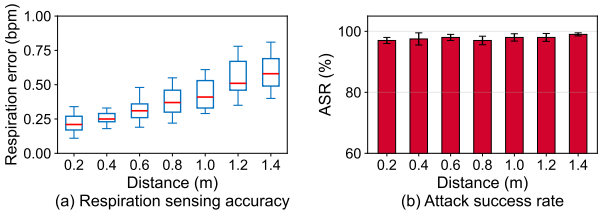
7.2 Attacks on FMCW-based System

Targeted existing sensing system. We replicate an FMCW-based sensing system, LaSense [27], with the victim positioned 0.2 m from the transceiver, breathing naturally for 60 seconds, and repeating the process 20



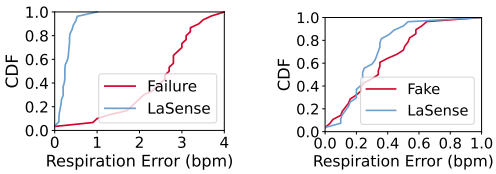
(a) Fake attack without DFSW (b) Fake attack with DFSW

Fig. 22. Ablation study of DFSW, comparing confusion matrices of fake-gesture attacks without DFSW and with DFSW.



(a) Respiration sensing accuracy (b) Attack success rate

Fig. 23. Reproduction of FMCW-based sensing systems.



(a) Failure attack (b) Fake result attack

Fig. 24. Attack results on FMCW-based systems.

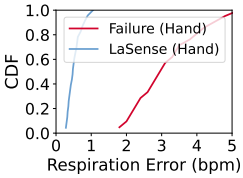


Fig. 25. Results on handheld FMCW-based system.

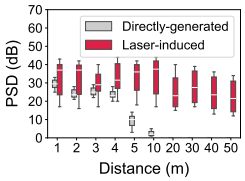
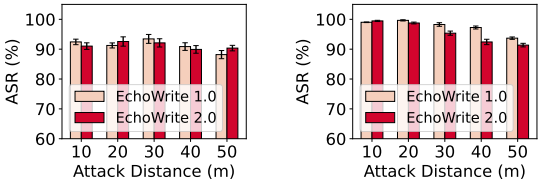


Fig. 26. Laser-induced vs. directly-generated.



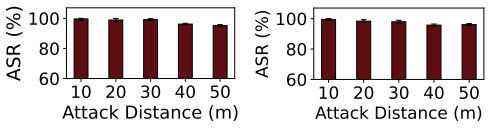
(a) Failure attack (b) Fake result attack

Fig. 27. Attack distance impact on CW-based systems.

times. Figure 23(a) presents the results of the reproduced system. As shown, the median respiration error is approximately 0.25 *bpm*, indicating high consistency with the original performance and confirming the accuracy of our replication.

FMCW respiration setup and distance choice. To validate that our attack remains effective across realistic deployment distances, we evaluate the FMCW-based respiration sensing system at multiple device–victim distances ranging from 0.2 m to 1.4 m under handheld usage scenarios. For each distance, we first measure the baseline respiration sensing error without any attack, and then evaluate the failure-attack performance. Figure 23 shows that, without attack, the FMCW system maintains sub-1 bpm error across all distances, although the error gradually increases as the distance grows. Figure 23(b) shows that the attack remains highly effective, with an ASR of approximately 95–100%, across all device–victim distances from 0.2 m to 1.4 m.

Failure attack. Here, we adopt the same setup as in attacking CW-based systems. Figure 24(a) presents the CDF of the respiration error under attack. The error significantly increases, with the median error reaching 2.8 *bpm*, demonstrating that the attack effectively distorts the sensing signal.



(a) Failure attack (b) Fake result attack

Fig. 28. Attack Distance vs. FMCW Performance.

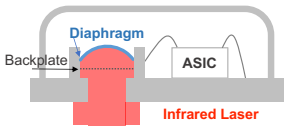


Fig. 29. Laser strikes MEMS mic.

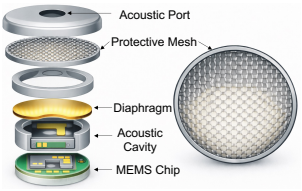


Fig. 30. Typical microphone stack.

Fake result attack. Here, the attacker injects carefully crafted signals that mimic real human respiration patterns, misleading the system into detecting the breathing of a non-existent user. Figure 24(b) shows a median respiration error of ~ 0.3 *bpm*, all below 1 *bpm*, indicating successful injection of fake breathing. Moreover, the error distribution closely resembles that of real measurements, further confirming the effectiveness of our attack.

Robustness of acoustic sensing under handheld device motion. We further validate the attack feasibility when the sensing device is handheld. The attacker is positioned approximately 5 meters away and performs a failure attack by aiming the laser at the device's microphone to induce sensing failure (the aiming process of *LightSaber* is detailed in Section 6 (Implementation)). To mitigate the effects of hand-induced motion, we employ a laser diode with an optical lens to slightly enlarge the beam diameter, increasing tolerance to device movements while maintaining sufficient optical intensity for the attack. We report both the baseline sensing performance of LaSense [27] under a handheld scenario and the performance under attack. To preserve baseline accuracy under handheld motion, we incorporate a motion-mitigation technique inspired by Liu *et al.* [29] on top of LaSense. Figure 25 shows that the failure attack remains feasible in handheld scenarios with mild hand motions.

Justification of the autocorrelation threshold. For the victim's respiration pattern without attacks, the autocorrelation is very high and concentrated (mean ≈ 0.97, variance $\approx 8.6 \times 10^{-5}$, min = 0.946, max = 0.986), confirming a strong and stable periodic pattern. Under our failure attacks, the autocorrelation drops sharply (mean ≈ 0.048, variance $\approx 3.3 \times 10^{-4}$, min = 0.013, max = 0.108), indicating that the injected interference effectively destroys the original respiration pattern. Due to this large separation between the two distributions, we observe that using thresholds in the range 0.5-0.8 produces nearly the same attack results. In our implementation, we empirically set the threshold to 0.8 to detect corrupted patterns, which provides a clear separation between the distributions of legitimate and attacked respiration patterns.

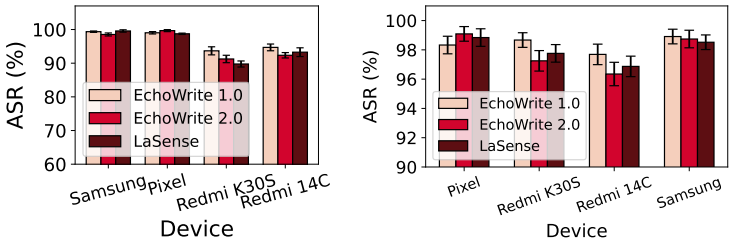
7.3 Comparison with Speaker-based Attacks

We use the speaker of a smartphone (Samsung S21 Ultra) to directly generate an acoustic attack signal and compare its power spectral density (PSD) with that of our laser-induced acoustic attack signal across distances ranging from 10 *m* to 50 *m*. As Figure 26 shows, at shorter ranges, both the laser-induced and directly-generated acoustic attack signals exhibit similar characteristics. However, as the distance increases, directly-generated acoustic signals suffer significant propagation loss, while laser-induced attack signals maintain a stable PSD. The underlying reason is the low attenuation of the laser beams in air, its concentrated energy, and the fact that it directly generates acoustic signals at the microphone. This shows that laser-based signals can enable remote injection, extending the feasibility of attacks beyond traditional acoustic methods.

7.4 Impact of Attack Distance

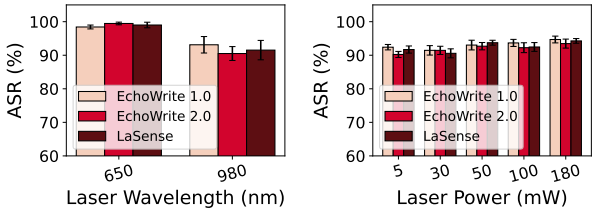
We investigate the key factor in long-range attacks—the impact of distance on the system's performance. We conduct end-to-end attacks at varying distances (from 10 *m* to 50 *m*, in 10 *m* increments). For CW-based sensing, Figure 27(a) shows that the ASR for the failure attack remains above 90% within 30 *m* and above 88% at 50 *m*,

Fig. 31. Microphone port close-ups.



(a) Different driver phones (b) Different sensing phones

Fig. 32. Robustness to phones.



(a) Impact of laser wavelength (b) Impact of laser power

Fig. 33. Robustness to laser parameters.

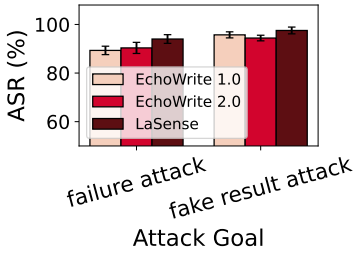


Fig. 34. Robustness to cross-building attack.

demonstrating the attack's long-range effectiveness. Similarly, Figure 27(b) shows that the fake result attack achieves over 85% success at all tested distances. For FMCW-based sensing, Figure 28 shows that even at 50 m, the ASR remains above 97.8%, demonstrating its capability of a long-range attack. Furthermore, Figure 28 shows that the ASR exceeds 96% across all tested distances. To mitigate signal attenuation due to beam expansion over long distances, we attach a lens to the laser diode to refocus the beam and minimize energy loss. Based on the laser diode's datasheet, the attack could theoretically reach distances of up to 100 m.

7.5 The Effect of Device Diversity

We further evaluate the generalizability of our system using different smartphones and laser diodes as attack devices, as well as different smartphones as victim devices. Without loss of generality, we conduct fake result attacks on the CW-based sensing system using the same experimental setup as in previous experiments.

Robustness to different microphone port and enclosure designs. In commodity smartphones, the microphone diaphragm is typically accessed via an open acoustic port to enable effective sound transmission. As illustrated in Figure 29 and Figure 30, this port is typically covered by an acoustically transparent (and often optically transmissive) protective mesh, rather than being fully blocked by opaque housing components. Even when commonly used protective phone cases are installed, the case openings are typically aligned with the microphone ports and therefore do not obstruct the optical path. As a result, although the diaphragm itself is not directly exposed, commodity smartphones still provide a usable optical entry point through the acoustic port. To validate performance across different microphone ports and enclosure designs, we conduct attacks on the CW-based sensing system using the same experimental setup as in our prior experiments, targeting four representative commercial smartphones spanning different market tiers (Samsung S21 Ultra, Pixel 7Pro, Redmi K30S, and Redmi 14C). Among them, Pixel 7Pro uses a larger slit-shaped port, while the others use circular

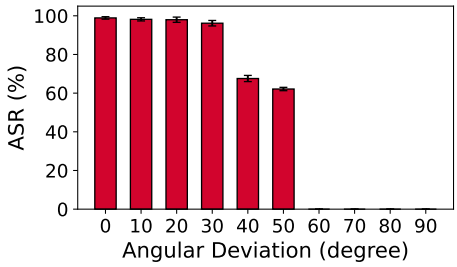
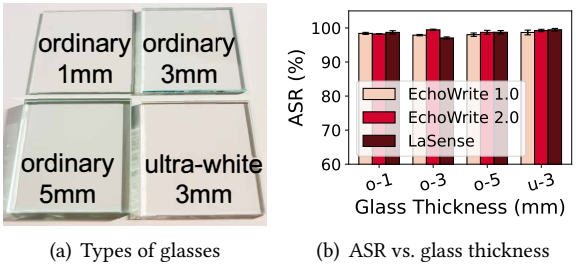


Fig. 35. Robustness to incidence angle.



(a) Types of glasses

(b) ASR vs. glass thickness

Fig. 36. Robustness of attacks through glass barriers.

ports with different sizes and mesh materials. Figure 31 shows close-up photos of the microphone ports, and Figure 32(b) summarizes the attack performance across devices, showing that injection remains feasible in all four cases.

Different smartphones as attack devices. We test four smartphones—Samsung S21 Ultra, Pixel 7Pro, Redmi K30S, and Redmi 14C—each driving the same 650 *nm*, 5 *mW* laser diode to inject attack signals into a Samsung S21 Ultra sensing smartphone. As shown in Figure 32(a), high-end smartphones like Samsung S21 Ultra and Pixel 7 Pro generate stronger and more stable attack signals, with induced acoustic signals showing around 20% higher energy than lower-end smartphones. However, even budget smartphones like Redmi K30S and 14C achieve an ASR of 92.50%, demonstrating the broad applicability of the proposed attack. Additionally, we found that integrating a simple, low-cost audio amplifier (e.g., LM386, $0.22) can further enhance the induced signal strength of budget smartphones.

Different smartphones as the victim. We use a Samsung S21 Ultra to drive the laser and inject attack signals into various smartphones. Figure 32(b) shows minimal variation in the attack performance due to the common use of MEMS microphones and optimized audio processing across devices. This confirms the broad applicability of our attack across different smartphones.

Laser diode diversity. We first evaluate the impact of laser wavelength while fixing the power at 5 *mW*. We observe minimal differences in signal energy across wavelengths, but the ASR for 980 *nm* lasers drops by 8% due to the greater difficulty of aiming fully invisible light. Despite this, it still achieves an ASR of 91.72% as shown in Figure 33(a), demonstrating the feasibility of using non-visible lasers for attacks. Then, we evaluate the impact of laser power by fixing the wavelength as 980 *nm* (invisible). Figure 33(b) shows that while the ASR slightly increases with power, 5 *mW* is already sufficient, proving that the proposed attack does not require high laser power.

Smartphone usage frequency. In our experiments, the Redmi 14C (a low-end model) and the Samsung S21 Ultra (a high-end flagship) are used most frequently as laser drivers, as they represent two realistic extremes of low-end and high-end attacker smartphones. Across all trials, the Redmi 14C and Samsung S21 Ultra each account for approximately 40% of the usage. The Pixel 7 Pro and Redmi K30S are each used in about 10% of the experiments as mid-range devices, further confirming that our attack is not tied to any specific phone model.

7.6 Evaluations in Different Environments

We use a 5 *mW*, 650 *nm* laser diode, driven by a Samsung S21 smartphone. We use a Redmi 14C as the sensing device. By default, we conduct fake result attacks on the CW-based sensing system using the same setup as in previous experiments.

Cross-building attack. We evaluate our attack under real-world conditions, where the attacker covertly targets a device from another building, penetrating glass windows while accounting for natural environmental interference. The setup, as shown in Figure 15, takes place on a campus. The target device is in a classroom on the first floor, while the attacker is positioned on the second floor of an adjacent building, 3 m above the ground and 20.2 m away from the target. As shown in Figure 34, our attack achieves an average ASR of 93.56%, demonstrating its effectiveness in long-range, cross-building scenarios.

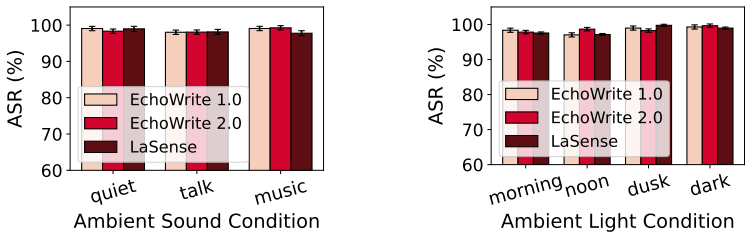
Robustness to device orientation. We evaluate how device orientation affects attack effectiveness by varying the incident angle of the laser beam relative to the target microphone. We define 0° as the case where the laser beam is perpendicular to the microphone opening (i.e., directly aimed at the microphone), and then increase the angular deviation from 0° to 90°. We use the default setup to evaluate the CW-based fake acoustic sensing result attack. Figure 35 shows the achieved ASR under different incident angles. We observe that the attack performance remains stable for small angular deviations (0–30°), where the ASR stays close to the best-case performance (e.g., $\geq$ 95%). As the deviation increases further, the ASR decreases noticeably; however, the attack can still succeed at larger angles, achieving approximately 68% at 40° and 62% at 50°. Overall, these results suggest that while a LoS path is required, the laser-based attack does not rely on perfect perpendicular alignment and can tolerate moderate device-orientation deviations in practice.

The effect of glass thickness. Similar to the above real-world attack scenarios, attackers transmit laser signals through windows and glass doors to realize a stealthy attack. We evaluate this by letting the laser pass through different types of glass, including ordinary green-tinted glass (1 mm, 3 mm, and 5 mm) and ultra-white glass (3 mm), as shown in Figure 36(a). The results in Figure 36(b) indicate that the attack ASR remains consistently high across different glass thicknesses. Notably, in our cross-building attack scenario (as shown in Figure 34), the classroom window penetrated is double-layered insulating acoustic glass with a thickness of 15 mm, effectively demonstrating the practicality of our attack. In summary, our tests span both single- and double-layer glass, reflecting common real-world use cases.

Observation under realistic user placement. We conduct a study in an open office environment with 20 participants, who naturally placed their sensing smartphones during normal use without any guidance on placement or mention of attacks. The office features non-transparent walls and standard normal-sized transparent windows. We consider an adversary located outside the office and assess whether a line-of-sight (LoS) injection path to the smartphone microphone exists. By examining only locations on the same floor level as the target office, we identify feasible LoS paths in 6 out of 20 natural device placements. We believe that with additional vantage points from different floors and neighboring buildings, line-of-sight paths would be available for even more devices. We further observe that placements such as downward-facing devices or microphones covered by surrounding surfaces, which fully block the LoS path to the microphone, are uncommon in natural placement scenarios. Moreover, acoustic sensing systems typically require users to avoid device placements that persistently block microphone openings, as such blockage can significantly degrade the sensing performance.

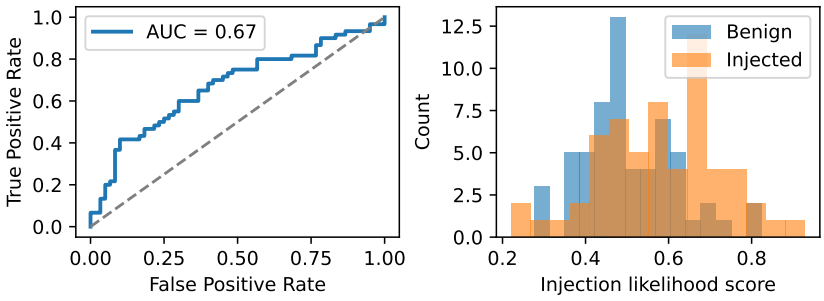
Robustness to Laser–Mic misalignment. Note that the offset tolerance depends on the laser spot size, and increasing the spot diameter improves tolerance to lateral misalignment. Based on our experiments, we observe that for a laser spot with radius r, the tolerated offset is slightly smaller than r. In our current implementation, at a distance of 10 m, a spot diameter of 42 mm can still deliver sufficient power for the attack, allowing a lateral offset of around 20 mm. Since the beam spot diameter is on the order of tens of millimeters, while a typical microphone port is approximately 1 mm in size, our attack can tolerate a non-trivial amount of lateral misalignment.

Ambient sound and light. In real-world environments, both ambient noise and lighting conditions vary, potentially affecting the performance of laser-induced attacks. We evaluate the attack ASR under different ambient sound conditions, including quiet environments, conversations, and background music. As shown in Figure 37(a), LightSaber consistently maintains a high ASR across all conditions, indicating that external sound does not significantly impact the attack. Similarly, we test the attack under various lighting conditions, including



(a) Impact of ambient sound (b) Impact of ambient light

Fig. 37. Robustness to ambient sound and light.



(a) ROC curve (b) Score distribution

Fig. 38. ECAPA-based detection of injected signals: (a) ROC curve; (b) distribution of detection scores for legitimate and injected segments.

morning, noon, dusk, and complete darkness. As shown in Figure 37(b), LightSaber remains effective across different lighting conditions, demonstrating that ambient light does not interfere with the attack.

8 POTENTIAL DEFENSES

Physical and hardware defenses. Although placing physical barriers—such as micro-perforated shields, diffraction films, or opaque covers—may appear to be a straightforward solution for blocking direct laser exposure to microphones, these methods often compromise acoustic performance and are impractical for compact or commercial devices. In addition, several hardware-based countermeasures have been proposed. For example, AIC [16] uses an extra speaker array to block inaudible attacks, and SurfingAttack [60] modifies speaker circuitry to resist interference. However, these methods often require hardware modifications or rely on non-portable, powered components, which limit their practicality in real-world deployment.

Dynamic sensing parameters. One strategy to mitigate the attack is dynamically altering key sensing parameters, such as signal duration, intervals, and bandwidth. Most of the current acoustic sensing systems often operate with fixed parameters, making them predictable and vulnerable to signal injection attacks. By applying randomized changes to these parameters, attackers will struggle to craft precisely aligned fake signals. However, this requires careful design to strike a balance between security enhancements and potential degradation in sensing performance.

Physical fingerprint authentication. Another potential defense is integrating physical fingerprint authentication before executing critical commands. Similar to biometric authentication for unlocking smartphones, acoustic sensing systems can require user-specific features, such as unique motion dynamics or personalized

gestures, to verify legitimate interactions. This approach enhances security by ensuring that only authorized users can trigger specific actions, making it difficult for attackers to inject fake sensing signals. Specifically, we apply the widely used ECAPA-TDNN model [10] as a binary audio-print detector. Under the CW-based failure attack setup, we collect two types of audio segments: (i) benign segments that contain only legitimate sensing signals, and (ii) attacked segments where the legitimate signals are mixed with the injected attack signals. These segments are labeled as legitimate and injected, respectively, and used to train ECAPA-TDNN. The model takes raw audio segments as input and outputs an "injection likelihood" score indicating how likely the segment contains injected signals. We evaluate the detector using the receiver operating characteristic (ROC) curve, which shows how well it can detect injected signals while avoiding false alarms on legitimate ones. We then use the Area Under the ROC Curve (AUC) as a single score that summarizes this performance: a higher AUC means the detector separates injected and legitimate signals more reliably. As shown in Figure 38, the detector achieves an AUC of 0.67, modestly better than random guessing (0.5). However, the two distributions still overlap heavily as detailed in Figure 38(b). Thus, this software-only physical fingerprinting provides partial protection rather than a complete defense; we believe that combining it with stronger front-end protections (e.g., optical filters) can yield more robust defenses.

Multi-microphone cross-verification. Most modern acoustic sensing devices use multiple microphones to improve signal reception. A potential defense against laser-induced attacks is the cross-verification of signals from multiple microphones. Since acoustic signals are transmitted and propagated omnidirectionally, nearby microphones will all receive them. Besides, multiple microphones can capture multi-dimensional motion information, providing further validation against attack signals [5]. However, an attacker can attempt to bypass this defense by using multiple lasers to inject dedicated signals into all microphones simultaneously. Notably, since smartphone audio outputs typically support two channels (left and right), our proposed smartphone-defined laser system can drive two laser diodes, enabling dual-point injection. Attacking devices with more microphones would require multiple smartphones and laser drivers, increasing the attack cost.

9 DISCUSSION AND FUTURE WORK

Drone-assisted attack. Laser-based remote attacks typically require a line-of-sight (LoS) path between the attacker and the target device. In some scenarios, off-the-shelf drones could be used as an optional way to temporarily gain LoS when ground paths are blocked (e.g., across a courtyard or toward a window) by carrying our lightweight smartphone-defined laser system (total weight $\leq$ 300 g). However, drone-based deployment faces practical constraints, including legal and regulatory restrictions, visual exposure, and challenges related to wind and stabilization. Thus, we consider drone assistance as a potential extension for specific scenarios, rather than a general or necessary assumption for our attack.

Other acoustic sensing techniques. LightSaber primarily focuses on designing attack signals for CW-based and FMCW-based systems, both of which are commonly used in commercial devices. We believe that by carefully crafting these attack signals, our approach can also be extended to other acoustic sensing techniques that use different signal types, such as Zadoff-Chu (ZC) sequences [41, 48], m-sequences [68], and GSM training sequences [63].

Attacks on MEMS sensors. The susceptibility of current microphones to our photoacoustic attack largely stems from their widespread use of MEMS technology. MEMS microphones rely on thin, lightweight, and highly compliant mechanical structures to achieve high sensitivity. These properties also make them particularly responsive to laser-induced acoustic excitation, leading to stronger perturbations than would be expected in sensors built from more rigid materials. Based on this observation, we find that the proposed attack can potentially extend to other MEMS-based sensors—such as accelerometers, gyroscopes, and parking radars—that adopt similar material and structural designs. By modulating laser signals to generate controlled acoustic waves, an adversary

could remotely perturb these sensors in a highly directional, long-range, and covert manner. Such capabilities introduce elevated risks to safety-critical systems (e.g., smart vehicles and drones), exceeding those posed by traditional sound-based attacks [13, 45]. In terms of alignment, the current attack does not rely on any explicit feedback. Instead, alignment is performed using visual and optical cues with manual adjustment. This manual alignment is generally sufficient when the sensing device is stationary or quasi-stationary, which is the typical setting for existing acoustic sensing scenarios. Nevertheless, we acknowledge that feedback-based automatic alignment mechanisms become critical for laser-based attacks targeting sensors under fast or frequent motion (e.g., vehicle-mounted LiDAR systems) [22], where continuous tracking and rapid re-alignment are required. Extending our attack to such dynamic sensing systems remains an interesting direction for future work.

Long-term detectability. We consider the long-term detectability of our attack from three practical perspectives. First, in terms of physical energy and heating, the lasers we use are relatively low power. Our experiments primarily use a 5 mW module, with the maximum tested power being 180 mW. In contrast, commercial laser engraving systems typically operate at optical powers ranging from 5 to 40 W [39], which is several orders of magnitude higher. In our default setup, the 5 mW laser also falls within typical consumer safety limits in the U.S. and is comparable to everyday laser pointers. Under these power levels, duty cycles, and distances, we did not observe noticeable surface heating, visible marks, or permanent damage on any device, suggesting that long-term thermal traces are very limited in our setting. Second, if an attack were sustained for a long time and a user became suspicious, defenders could deploy external tools (e.g., infrared cameras) to search for unusual heat spots or IR emission. However, in practice, the attacker can simply switch off the laser and disengage immediately; from a long distance, it is difficult to localize the emission source or identify the attacker in time. Third, if the victim intentionally changes the device position, the attacker can, in principle, re-aim and reacquire the target with the same aiming workflow.

Although we primarily employ lasers for signal modulation, our attack method is not inherently limited to laser sources. Other light-emitting devices, such as high-power LEDs, could also be used to modulate ultrasonic signals via the photoacoustic effect. These alternatives offer trade-offs: for instance, LEDs can relax the alignment precision requirements due to their wider emission angle, but they may significantly reduce the effective attack range compared to lasers. Exploring these trade-offs opens up new opportunities for practical, low-cost, and more flexible photoacoustic attack systems.

Photoacoustic effect vs. Photoelectric effect. When a laser illuminates the target, both photoacoustic and photoelectric effects may be induced. Although early studies [40] speculated that both mechanisms could contribute to the signal injection process, recent investigations [9] provide stronger evidence that the photoacoustic effect is the dominant physical mechanism in MEMS-based systems.

10 RELATED WORK

Attacks on wireless sensing systems. Previous works have revealed security vulnerabilities in wireless sensing systems. In RF-based sensing, researchers have used relay manipulation and interference injection to compromise Wi-Fi and mmWave sensing [19, 20, 25, 28, 69, 70]. For example, WiAdv [69] employs a full-duplex attack device to forward Wi-Fi signals, mimicking dynamic multi-path to deceive gesture recognition. MadRadar [19] uses a transceiver to capture and analyze radar signals, then transmits crafted signals for a black-box mmWave attack. Additionally, researchers [20, 70] exploit reconfigurable intelligent surfaces (RIS) to manipulate channels and spoof Wi-Fi sensing. However, attacks on acoustic sensing are rare because sound waves have a limited range and weaken through walls, making remote attacks difficult. Moreover, existing attacks on Wi-Fi and mmWave sensing systems typically assume that the attacker is close to the target and has access to the legitimate signal. In contrast, we relax this assumption and design a series of methods that enable remote acoustic attacks without

requiring access to legitimate signals. To the best of our knowledge, LightSaber is the first long-range, stealthy, and practical attack on acoustic sensing systems.

Photoacoustic applications. The photoacoustic effect is a physical phenomenon that converts light energy into sound energy and is widely used in medical imaging [6, 57] and gas sensors [18, 23, 34, 61]. In photoacoustic attacks, prior studies [40, 66] have explored using this effect to spoof human voice commands. However, prior studies [40, 66] mainly target voice command systems (e.g., Alexa, Siri) and cannot be directly applied to acoustic sensing systems. We summarize the key differences as follows:

(a) *Different attack goals.* Existing attacks primarily target VC systems, which passively receive acoustic signals for speech command recognition. In contrast, our attack targets acoustic sensing systems that actively transmit CW/FMCW signals to infer physical states from subtle, time-varying perturbations in the reflected signals for sensing such as gesture recognition and vital-sign monitoring.

(b) *Different attack signal designs.* Unlike "speech-on-laser" modulation used in VC attacks, which injects known and fixed speech signals without stringent timing constraints, compromising acoustic sensing systems requires crafting significantly more delicate attack signals. Specifically, the injected signals must (i) closely resemble legitimate CW/FMCW sensing waveforms and (ii) be timing-constrained by the target device's sensing timeline, without direct access to the original sensing signal or its exact timing. These requirements necessitate novel end-to-end signal design and timing strategies, going beyond prior VC-based laser injection attacks.

(c) *Ubiquitous, lightweight attack hardware.* Existing laser attacks typically rely on bulky and expensive laboratory equipment (e.g., function generators and bench-top laser drivers), which is less practical for real-world deployment. In contrast, we show that a commodity smartphone can serve as a ubiquitous and lightweight attack hardware by driving low-cost laser diodes via the smartphone's audio output, enabling a more realistic attack in everyday environments.

11 CONCLUSION

In this work, we propose LightSaber, the first system to exploit the "photoacoustic effect" for long-range, stealthy attacks on acoustic sensing using infrared lasers. We design signal generation pipelines tailored to CW-based and FMCW-based acoustic sensing. Furthermore, we introduce the concept of Smartphone-Defined Laser, allowing smartphones to drive laser diodes, making the attack feasible with widely available smartphones. We implement and evaluate LightSaber, demonstrating its long-range effectiveness, robustness, and practicality under diverse real-world scenarios. We believe our system exposes overlooked security vulnerabilities in acoustic sensing and paves the way for a new research trend in both attacks and protections for acoustic sensing systems.

Acknowledgment

This work was supported by the NTU SUG-NAP.

References

[1] Xiaomi sound pro. https://www.mi.com/speakers/sound-pro.

[2] Ugreen usb-c to 3.5mm headphone adapter. https://www.amazon.sg/UGREEN-Adapter-Headphone-Dongle-Compatible/dp/B082WG5VTK, 2026. Accessed: 2026-01-20.

[3] Amazon. Echo (4th gen) | with premium sound, smart home hub, and alexa. https://www.amazon.com/Echo-4th-Gen/dp/B07XKF5RM3. Accessed: 2025-01-27.

[4] Y. A. Badamasi. The working principle of an arduino. In *2014 11th international conference on electronics, computer and computation (ICECCO)*, pages 1–4. IEEE, 2014.

[5] Y. Bai, L. Lu, J. Cheng, J. Liu, Y. Chen, and J. Yu. Acoustic-based sensing and applications: A survey. *Computer Networks*, 181:107447, 2020.

[6] P. Beard. Biomedical photoacoustic imaging. *Interface focus*, 1(4):602–631, 2011.

[7] T. Chen, Y. Yang, X. Fan, X. Guo, J. Xiong, and L. Shangguan. Exploring the feasibility of remote cardiac auscultation using earphones. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*, pages 357–372, 2024.

[8] M. Cui, B. Xie, Q. Wang, and J. Xiong. Dancingant: Body-empowered wireless sensing utilizing pervasive radiations from powerline. In *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*, pages 1–15, 2023.

[9] B. Cyr, T. Sugawara, and K. Fu. Why lasers inject perceived sound into mems microphones: Indications and contraindications of photoacoustic and photoelectric effects. In *2021 IEEE Sensors*, pages 1–4. IEEE, 2021.

[10] B. Desplanques, J. Thienpondt, and K. Demuynck. Ecapa-tdnn: Emphasized channel attention, propagation and aggregation in tdnn based speaker verification. *arXiv preprint arXiv:2005.07143*, 2020.

[11] J. Ding, R. Chandra, R. Lal, and L. Tassiulas. Cost-effective soil carbon sensing with wi-fi and optical signals. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*, pages 1015–1029, 2024.

[12] Y. Feng, Y. Xie, D. Ganesan, and J. Xiong. Lte-based low-cost and low-power soil moisture sensing. In *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*, pages 421–434, 2022.

[13] M. Gao, L. Zhang, L. Shen, X. Zou, J. Han, F. Lin, and K. Ren. Kite: Exploring the practical threat from acoustic transduction attacks on inertial sensors. In *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*, pages 696–709, 2022.

[14] Google Nest. Google nest: Welcome to the helpful home, 2019.

[15] X. Guo, L. Tan, T. Chen, C. Gu, Y. Shu, S. He, Y. He, J. Chen, and L. Shangguan. Exploring biomagnetism for inclusive vital sign monitoring: Modeling and implementation. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*, pages 93–107, 2024.

[16] Y. He, J. Bian, X. Tong, Z. Qian, W. Zhu, X. Tian, and X. Wang. Canceling inaudible voice commands against voice control systems. In *The 25th Annual International Conference on Mobile Computing and Networking*, pages 1–15, 2019.

[17] A. Hordvik and H. Schlossberg. Photoacoustic technique for determining optical absorption coefficients in solids. *Applied optics*, 16(1):101–107, 1977.

[18] Y. Hu, S. Qiao, Y. He, Z. Lang, and Y. Ma. Quartz-enhanced photoacoustic-photothermal spectroscopy for trace gas sensing. *Optics Express*, 29(4):5121–5127, 2021.

[19] D. Hunt, K. Angell, Z. Qi, T. Chen, and M. Pajic. MadRadar: A black-box physical layer attack framework on mmwave automotive FMCW radars. In *Network and Distributed System Security Symposium (NDSS)*. Internet Society, 2024.

[20] C. Jiang, J. Yang, X. Li, Q. Li, X. Zhang, and J. Ren. Risiren: Wireless sensing system attacks via metasurface. In *Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security*, pages 3332–3345, 2024.

[21] H. Jiang, J. Zhang, X. Guo, and Y. He. Sense me on the ride: Accurate mobile sensing over a lora backscatter channel. In *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*, pages 125–137, 2021.

[22] Z. Jin, X. Ji, Y. Cheng, B. Yang, C. Yan, and W. Xu. Pla-lidar: Physical laser attacks against lidar-based 3d object detection in autonomous vehicle. In *2023 IEEE Symposium on Security and Privacy (SP)*, pages 1822–1839. IEEE, 2023.

[23] T. Kim, J. Kim, and X. Jiang. Aln ultrasound sensor for photoacoustic lamb wave detection in a high-temperature environment. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 65(8):1444–1451, 2018.

[24] N. Kühnapfel, S. Preußler, M. Noppel, T. Schneider, K. Rieck, and C. Wressnegger. Lasershark: Establishing fast, bidirectional communication into air-gapped systems. In *Proceedings of the 37th Annual Computer Security Applications Conference*, pages 796–811, 2021.

[25] C. Li, M. Xu, Y. Du, L. Liu, C. Shi, Y. Wang, H. Liu, and Y. Chen. Practical adversarial attack on wifi sensing through unnoticeable communication packet perturbation. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*, pages 373–387, 2024.

[26] D. Li, S. Cao, S. I. Lee, and J. Xiong. Experience: practical problems for acoustic sensing. In *Proceedings of the 28th Annual International Conference on Mobile Computing And Networking*, pages 381–390, 2022.

[27] D. Li, J. Liu, S. I. Lee, and J. Xiong. Lasense: Pushing the limits of fine-grained activity sensing using acoustic signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 6(1):1–27, 2022.

[28] J. Liu, Y. He, C. Xiao, J. Han, L. Cheng, and K. Ren. Physical-world attack towards wifi-based behavior recognition. In *IEEE INFOCOM 2022-IEEE Conference on Computer Communications*, pages 400–409. IEEE, 2022.

[29] J. Liu, D. Li, L. Wang, F. Zhang, and J. Xiong. Enabling contact-free acoustic sensing under device motion. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 6(3):1–27, 2022.

[30] W. Mao, Z. Zhang, L. Qiu, J. He, Y. Cui, and S. Yun. Indoor follow me drone. In *Proceedings of the 15th annual international conference on mobile systems, applications, and services*, pages 345–358, 2017.

[31] P. M. Morse and K. U. Ingard. *Theoretical acoustics*. Princeton university press, 1986.

[32] R. Nandakumar, S. Gollakota, and N. Watson. Contactless sleep apnea detection on smartphones. In *Proceedings of the 13th annual international conference on mobile systems, applications, and services*, pages 45–57, 2015.

[33] K. Qian, C. Wu, F. Xiao, Y. Zheng, Y. Zhang, Z. Yang, and Y. Liu. Acousticcardiogram: Monitoring heartbeats using acoustic signals on smart devices. In *IEEE INFOCOM 2018-IEEE conference on computer communications*, pages 1574–1582. IEEE, 2018.

[34] F. Qin, B. Li, L. Chen, T. Tang, Z. Wang, Y. Huang, Z. Zhang, Z. Zhang, and Y. Zheng. Influence mechanism of gas turbine blade thickness on photoacoustic signals to guide defect imaging detection. *IEEE Transactions on Instrumentation and Measurement*, 2023.
[35] W. Ruan, Q. Z. Sheng, L. Yang, T. Gu, P. Xu, and L. Shangguan. Audiogest: Enabling fine-grained hand gesture detection by decoding echo signal. In *Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing*, pages 474–485, 2016.
[36] P. Series. 1:'propagation data required for the design of terrestrial free-space optical links'. *Recommendation ITU-R*, 2012.
[37] J. Shenoy, Z. Liu, B. Tao, Z. Kabelac, and D. Vasisht. Rf-protect: privacy against device-free human tracking. In *Proceedings of the ACM SIGCOMM 2022 Conference*, pages 588–600, 2022.
[38] W. Song, J. Liu, and J. Han. Puppetmouse: Practical and contactless mouse manipulation attack via intentional electromagnetic interference injection. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 8(3):1–30, 2024.
[39] W. M. Steen and J. Mazumder. *Laser material processing*. springer science & business media, 2010.
[40] T. Sugawara, B. Cyr, S. Rampazzi, D. Genkin, and K. Fu. Light commands:{Laser-Based} audio injection attacks on {Voice-Controllable} systems. In *29th USENIX Security Symposium (USENIX Security 20)*, pages 2631–2648, 2020.
[41] K. Sun, T. Zhao, W. Wang, and L. Xie. Vskin: Sensing touch gestures on surfaces of mobile devices using acoustic signals. In *Proceedings of the 24th annual international conference on mobile computing and networking*, pages 591–605, 2018.
[42] X. Sun, J. Xiong, C. Feng, X. Li, J. Zhang, B. Li, D. Fang, and X. Chen. Gastag: A gas sensing paradigm using graphene-based tags. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*, pages 342–356, 2024.
[43] Texas Instruments. Lm386 low voltage audio power amplifier. https://www.ti.com/product/LM386, 2023. Accessed: 2026-01-20.
[44] Texas Instruments. Opa690 wideband, voltage-feedback operational amplifier with disable. https://www.ti.com/product/OPA690, 2023. Accessed: 2026-01-20.
[45] T. Trippel, O. Weisse, W. Xu, P. Honeyman, and K. Fu. Walnut: Waging doubt on the integrity of mems accelerometers with acoustic injection attacks. In *2017 IEEE European symposium on security and privacy (EuroS&P)*, pages 3–18. IEEE, 2017.
[46] Y.-C. Tung, D. Bui, and K. G. Shin. Cross-platform support for rapid development of mobile acoustic sensing applications. In *Proceedings of the 16th annual international conference on mobile systems, applications, and services*, pages 455–467, 2018.
[47] E. Upton and G. Halfacree. *Raspberry Pi user guide*. John Wiley & Sons, 2016.
[48] H. Wan, S. Shi, W. Cao, W. Wang, and G. Chen. Resptracker: Multi-user room-scale respiration tracking with commercial acoustic devices. In *IEEE INFOCOm 2021-IEEE conference on computer communications*, pages 1–10. IEEE, 2021.
[49] C. Wang, F. Lin, T. Liu, K. Zheng, Z. Wang, Z. Li, M.-C. Huang, W. Xu, and K. Ren. mmeve: eavesdropping on smartphone's earpiece via cots mmwave device. In *Proceedings of the 28th Annual International Conference on Mobile Computing And Networking*, pages 338–351, 2022.
[50] K. Wang, C. Shi, J. Cheng, Y. Wang, M. Xie, and Y. Chen. Solving the wifi sensing dilemma in reality leveraging conformal prediction. In *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*, pages 407–420, 2022.
[51] S. Wang, L. Zhong, Y. Fu, L. Chen, J. Ren, and Y. Zhang. Uface: Your smartphone can" hear" your facial expression! *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 8(1):1–27, 2024.
[52] K. Wu, Q. Yang, B. Yuan, Y. Zou, R. Ruby, and M. Li. Echowrite: An acoustic-based finger input system without training. *IEEE Transactions on Mobile Computing*, 20(5):1789–1803, 2020.
[53] B. Xie, M. Cui, D. Ganesan, X. Chen, and J. Xiong. Boosting the long range sensing potential of lora. In *Proceedings of the 21st Annual International Conference on Mobile Systems, Applications and Services*, pages 177–190, 2023.
[54] B. Xie, M. Cui, D. Ganesan, and J. Xiong. Wall matters: Rethinking the effect of wall for wireless sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 7(4):1–22, 2024.
[55] Y. Xie, X. Guo, Y. Wang, J. Cheng, and Y. Chen. Universal targeted adversarial attacks against mmwave-based human activity recognition. In *Network Security Empowered by Artificial Intelligence*, pages 177–211. Springer, 2024.
[56] J. Xiong, K. Sundaresan, and K. Jamieson. Tonetrack: Leveraging frequency-agile radios for time-based indoor wireless localization. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, pages 537–549, 2015.
[57] M. Xu and L. V. Wang. Photoacoustic imaging in biomedicine. *Review of scientific instruments*, 77(4), 2006.
[58] W. Xu, W. Song, J. Liu, Y. Liu, X. Cui, Y. Zheng, J. Han, X. Wang, and K. Ren. Mask does not matter: Anti-spoofing face authentication using mmwave without on-site registration. In *Proceedings of the 28th Annual International Conference on Mobile Computing and Networking*, pages 310–323, 2022.
[59] X. Xu, H. Gao, J. Yu, Y. Chen, Y. Zhu, G. Xue, and M. Li. Er: Early recognition of inattentive driving leveraging audio devices on smartphones. In *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*, pages 1–9. IEEE, 2017.
[60] Q. Yan, K. Liu, Q. Zhou, H. Guo, and N. Zhang. Surfingattack: Interactive hidden attack on voice assistants using ultrasonic guided waves. In *Network and Distributed Systems Security (NDSS) Symposium*, 2020.
[61] T. Yang, W. Chen, and P. Wang. A review of all-optical photoacoustic spectroscopy as a gas sensing method. *Applied spectroscopy reviews*, 56(2):143–170, 2021.
[62] Y. Yin, L. Xie, Z. Jiang, F. Xiao, J. Cao, and S. Lu. A systematic review of human activity recognition based on mobile devices: Overview, progress and trends. *IEEE Communications Surveys & Tutorials*, 2024.

[63] S. Yun, Y.-C. Chen, H. Zheng, L. Qiu, and W. Mao. Strata: Fine-grained acoustic-based device-free tracking. In *Proceedings of the 15th annual international conference on mobile systems, applications, and services*, pages 15–28, 2017.

[64] Y. Zeng, D. Wu, J. Xiong, E. Yi, R. Gao, and D. Zhang. Farsense: Pushing the range limit of wifi-based respiration sensing with csi ratio of two antennas. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(3):1–26, 2019.

[65] F. Zhang, Z. Wang, B. Jin, J. Xiong, and D. Zhang. Your smart speaker can" hear" your heartbeat! *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(4):1–24, 2020.

[66] G. Zhang, X. Ma, H. Zhang, Z. Xiang, X. Ji, Y. Yang, X. Cheng, and P. Hu. {LaserAdv}: Laser adversarial attacks on speech recognition systems. In *33rd USENIX Security Symposium (USENIX Security 24)*, pages 3945–3961, 2024.

[67] X. Zhang, Y. Zhang, Z. Shi, and T. Gu. mmfer: Millimetre-wave radar based facial expression recognition for multimedia iot applications. In *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*, pages 1–15, 2023.

[68] Z. Zhang, D. Chu, X. Chen, and T. Moscibroda. Swordfight: Enabling a new class of phone-to-phone action games on commodity phones. In *Proceedings of the 10th international conference on Mobile systems, applications, and services*, pages 1–14, 2012.

[69] Y. Zhou, H. Chen, C. Huang, and Q. Zhang. Wiadv: Practical and robust adversarial attack against wifi-based gesture recognition system. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 6(2):1–25, 2022.

[70] Y. Zhou, C. Li, H. Chen, and Q. Zhang. Ristealth: Practical and covert physical-layer attack against wifi-based intrusion detection via reconfigurable intelligent surface. In *Proceedings of the 21st ACM Conference on Embedded Networked Sensor Systems*, pages 195–208, 2023.

[71] Y. Zou, Z. Xiao, S. Hong, Z. Guo, and K. Wu. Echowrite 2.0: A lightweight zero-shot text-entry system based on acoustics. *IEEE Transactions on Human-Machine Systems*, 52(6):1313–1326, 2022.

A. Appendix

A.1 Hardware Components

Table 1. Hardware components used in the attack prototype.

Component	Model	Price
Audio adapter	UGREEN USB-C to 3.5 mm (with DAC) [2]	$7.73
Laser driver	TI OPA690-based driver [44]	$11.8
Audio amplifier	TI LM386 [43] (optional)	$0.22
Laser diode	650 nm / 980 nm (5–180 mW)	$0.6–5

A.2 Schematic of Laser Driver

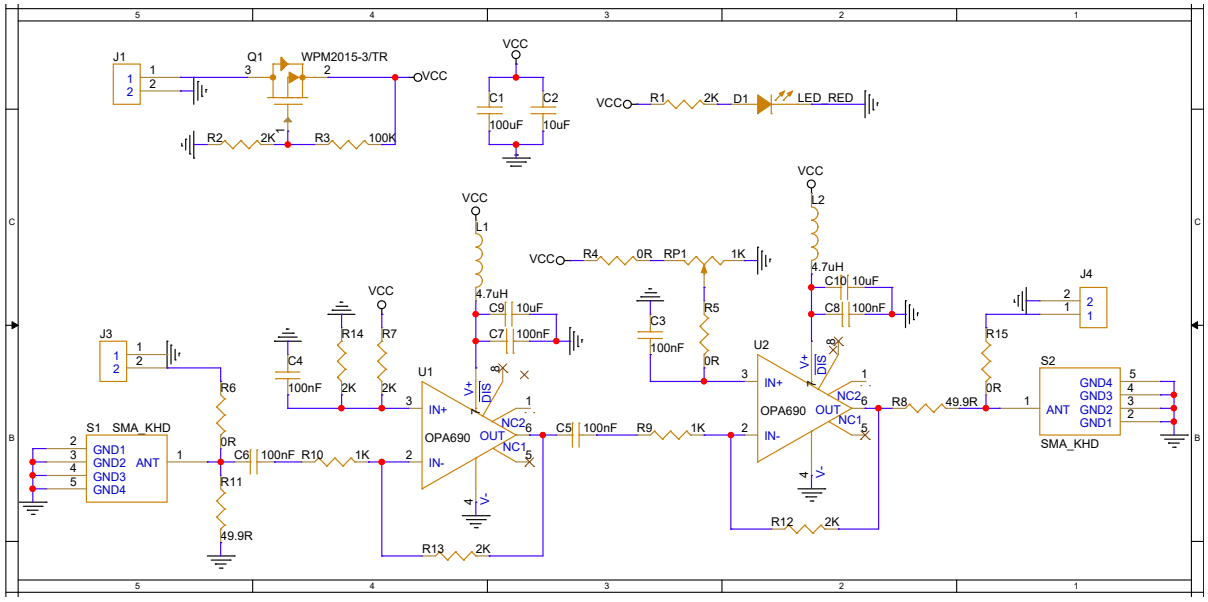


Fig. 39. Schematic of laser driver.