EyeHearYou: Probing Location Identification via Occluded Smartphone Cameras and Ultrasound

Nina Shamsi
College of Engineering
Northeastern University
Boston, USA
nshamsi@ece.neu.edu

Yan Long
College of Engineering
Northeastern University
Boston, USA
y.long@northeastern.edu

Kevin Fu
College of Engineering
Northeastern University
Boston, USA
k.fu@northeastern.edu

***Abstract*—This paper explores how to localize a device equipped solely with camera sensors by leveraging the unintended response of occluded camera hardware to transmitted ultrasound—specifically, determining with high probability which ultrasound transmission pattern was injected during image capture, based on distinctive ultrasound signals unwittingly detected by the image sensor. Prior device location identification methods require the use of dedicated hardware or protocols, e.g., microphone arrays or GPS. We envision a new potential mechanism by leveraging ubiquitous camera hardware on mobile and IoT devices to receive acoustic signals produced by ultrasonic localization beacons, even when the camera is occluded. We discover that ultrasonic signals affect gyroscopes integral to modern camera sensors' stabilization hardware and induce distinct destabilization signals in the dark images captured by occluded cameras. This work provides theoretical analysis, simulation modeling, and experimental evidence of how this optical acoustic side channel creates different noise patterns in camera images when the camera is subjected to different ultrasound stimuli. Our evaluation with 119 videos captured by a smartphone camera over multiple days shows success in detecting whether the smartphone is near an ultrasonic transmitter that can be associated with different locations.**

***Index Terms*—Optical Acoustic Side Channel, CMOS Image Sensors, Dark Signal, Device Localization, Smartphone Privacy**

I. Introduction

Privacy-oriented guidelines for mobile security often recommend judiciously granting camera and microphone access to user-installed applications, affecting user behavior [1], where some consumers may use lens covers and tape as an occlusion to prevent being observed by unintended recipients. Our work shows that an attacker can still learn information regarding the device location through an optical acoustic side channel even if the camera lens is occluded. As mobile and IoT devices are becoming ubiquitous and integral to a user's daily life, location privacy has become a major concern due to the fact that various information channels could leak the location of a device and its user. For example, global positioning system (GPS), often used for location-based services provided by smartphones, inevitably opens the door for applications to monitor user locations [2], [3]. In addition, microphones on mobile and wearable devices are exploited to receive location-specific acoustic signals generated by audio transmitters [4], [5], [6]. Given these threats, previous research has invented methods to filter out fine-grained location information from their sensor data [7], [8], or educated users to simply disable GPS and microphones on their devices. Despite these precautions against GPS and microphones which are known to be able to leak location information, this paper investigates a new, complementary threat model that asks the question: *How may adversaries possibly identify the location of a device by analyzing the camera images taken by the device, even if the camera lens is occluded?* Location privacy leakage due to unwitting sharing of camera photos and videos that contain identifiable landmarks in the image scenes is well known [9], [10], [11]. It is foreseeable that location identification threats through cameras will increase as cameras become one of the most widely deployed types of sensors [12], [13]. However, this work investigates how to achieve location identification with a dark image scene captured by an occluded camera by leveraging ultrasonic beacon signals in the ambient acoustic environment to create distinct features in dark images that seem to contain no usable information. Specifically, ultrasonic signals induce distinct destabilization patterns depending on the image stabilization mechanism used by camera sensors, e.g., optical image stabilization (OIS), electronic image stabilization (EIS), or both, by affecting the mechanical behavior of microelectromechanical systems (MEMS) gyroscopes integrated in these camera sensors. This leads to discernible spatial variance in dark signals of an image. Dark signals are thermally induced signals in image sensors that are dominated by image sensor noise sources due to low levels of image

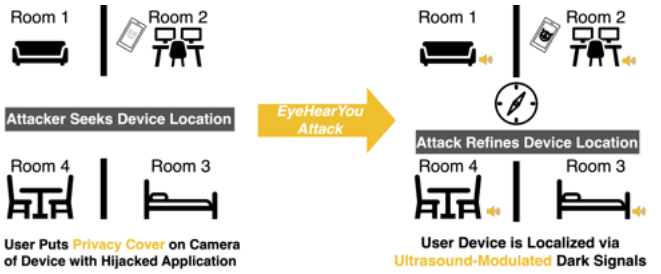

Fig. 1. **The Threat Model**. Using a dark signal optical acoustic side channel, a target mobile device can be located in a GPS-denied environment using only an occluded camera of a mobile device and ultrasonic transmitters. The target mobile device front and back cameras are covered with privacy guards.

sensor irradiation [14], [15], [16], such as when a camera lens is occluded. This work discovers that even with occluded camera lenses, the ultrasonic modulated signals differ from ambient acoustic noise signals, enabling the development of a signal correlation-based indicator function to predict device proximity to an ultrasonic transmitter, underscoring potentially leveraging occluded cameras for localization.

To probe the characteristics, and provide preliminary existence proof, of this optical acoustic side channel, our work develops a simulation model that aims to validate how the ultrasonic attack signal, at the resonant frequency of the gyroscope, modulates the camera response. In addition, we perform experiments with the camera of a commercial-off-of-the-shelf (COTS) smartphone and an ultrasonic speaker transmitting sinusoidal pulse and sweep waveforms at a distance of up to 0.28 m from the smartphone. Our simulation and experiments demonstrate strong correlations between the ultrasound signals around the camera hardware and the spatial variance of dark images captured by the occluded smartphone camera. Based on these observations, we develop a low-resource signal indicator to identify whether the occluded camera is near an ultrasonic transmitter that can be associated with a known location. The signal indicator holds a given signal as a baseline by computing the maximum correlation coefficient of the baseline signal against itself, and then it compares the maximum correlation coefficient of an incoming input signal against the baseline maximum correlation coefficient.

Our results show that the signal indicator could reliably distinguish between ultrasound-modulated signals and ambient acoustic noise signals across different ultrasonic transmission patterns and image stabilization settings. We evaluate the signal indicator using a MATLAB simulation on a dataset consisting of 119 videos captured by the occluded camera in various environmental settings and observe a near 99.9% accuracy in detecting whether there is an ultrasound signal around the occluded camera for a given camera configuration. In addition, we analyze the impact of key camera configuration parameters such as optical zoom, exposure time, rolling shutter rate, and image sensor size, among others. Finally, we discuss the limitations of our simulations and short-distance experiments, and how future research could leverage the analysis methodology provided in this work to design feasibility studies in more diverse scenarios. To summarize, our work has the following contributions:

- A threat model of using open and occluded cameras to receive ultrasonic signals for device location identification. The threat model completes existing GPS and microphone-based approaches and enables more ubiquitous localization capability in camera-only IoT devices.
- A methodology for acquiring ultrasound information from a dark image scene captured by occluded cameras. Our analysis of how ultrasound modulates image stabilization patterns presents an exploitable optical acoustic side channel.
- Simulation and experimental results. Our results provide preliminary evidence for the existence of such optical acoustic side channels and a foundation for future implementation to build upon.

II. Background & Related Work

Optical Acoustic Side Channels. Sound is a mechanical wave that can impart energy to any surface it strikes. The kinetic energy of sound waves can cause slight deviations on the surfaces of objects, which can be observed as pixel displacements using lenses to extract motion vectors and reconstruct the original sound [17], [18], [19]. Images acquired using mobile devices can be used to reconstruct sound from motion vectors of pixel displacements in images, where the pixel displacement is imparted in sequential image frames due to the kinetic energy of sound waves moving the springs holding the camera lens [20]. Pixel displacements in images caused by sound energy are amplified by the rolling shutter of image sensors because the shutter mechanism exposes and reads out sequential pixel rows from top to bottom. Note that sound reconstruction via the rolling shutter effect is possible in part due to the dynamic range of pixel intensities in an irradiated sensor image, i.e., an image taken with camera exposure [14], but a unirradiated sensor image will lack dynamic range as most intensity values will be nil; therefore, using images lacking dynamic range in pixel intensities is challenging for optical acoustic side channels.

Image Stabilization. Camera stabilization methods are useful for optical acoustic side channels [21], [20], [22]. Unwanted device movement while taking an image or capturing a video can cause motion artifacts, such as blurs, in the captured media. Optical image stabilization (OIS) involves the compensatory movement of a lens in the opposite direction to an image sensor in order to correct for motion blurs, often using micro-mechanical electronic systems (MEMS), such as gyroscopes [23], [24]. Electronic image stabilization (EIS) is a software-only counterpart of OIS, and corrects motion artifacts using signal processing [25]. In this paper, we configure different image stabilization mechanisms for a camera to determine their effect on the resultant dark signal camera responses.

Sound-Based Localization. Sound-based localization determines the location of a target using the direction of arrival (DOA) of sound waves and the distance between the transmitters and the receivers [26], [27], [28]. Measures used in sound-based localization compute time of arrival (TOA), time difference of arrival (TDOA), or time of flight (TOF) of the transmitted signal, providing high accuracy measurements while incurring an equipment cost, thus limiting scalability [27]. Ultrasound-based device localization techniques use sound frequencies near or greater than 20 kHz to localize a target, but have variable accuracy due to time-based measurements [29], [30], [31] [32]. Recent sound-based localization techniques improve on problems related to multi-path interference [29], or improving 2D accuracy [30] since pioneering work such as Cricket [32] and DOLPHIN [28]. However, accuracy gains from neural network-based methods are hindered by the need for training data, and computational complexity of the

methods [31]. We develop a low-resource signal processing-based method for device localization. Our approach does not require sound direction of arrival estimation, or time synchronization between the transmitter and image sensor. Our results remain consistent across different optical and hybrid zoom settings using two different image sensors, and different sets of camera and transmitter parameters, indicating that the observed phenomena are intrinsic properties of the optical system.

III. Threat Model

The EyeHearYou threat model is centered on an optical acoustic dark signals side channel in mobile device cameras where the amplitude of mean pixel intensities of an occluded camera can be modulated using an external driving force. The threat model transmitter is an ultrasound source, and the transceiver is an occluded camera of the target mobile device. The occluded camera is a transceiver because it both receives the ultrasound signal, and transmits a signal to indicate device proximity to the transmitter.

Adversary Motivation. The adversary wants remote device localization in a GPS-denied environment, and does not have access to an indoor positioning system, e.g., WiFi, or dedicated surveillance equipment. The adversary also lacks direct (line of sight) or indirect scene information of the target or its environment. The adversary can access the target mobile device camera, but when the adversary remotely accesses the camera, the position of the device results in low light, or occluded, conditions at either the front, rear, or both cameras.

Device Characteristics. The adversary performs the attack via a user-installed application which grants camera access; user-installed applications, such as QR code or PDF scanners, are a well-known attack vector for gaining access to smartphone cameras [33], [34]. User studies have shown that most mobile device users are more concerned about granting microphone access than camera access to mobile device applications [1], and as such user-installed mobile device applications are increasingly denied microphone access [35]. Therefore, we assume a scenario in which the adversary targets a single device via a user-installed application, and the user-installed application has denied microphone access. The target device is any commercial-off-the-shelf (COTS) mobile device with gyroscope-based image stabilization, which includes most modern smartphones and tablets [36], [37], [38].

Adversary Capabilities. The adversary has remote access to the camera of a target mobile device, and will acquire a signal indicator (SI) from a video recording alerting the adversary to the target device being within range of a transmitter. Once the adversary has a location of interest (LOI) in mind, an attack is carried out as described in Figure 2. The adversary can create a set of known nodes η by placing ultrasonic transmitters within the LOI, and enable on-device video processing to remotely acquire an SI, where each node η' transmits a characteristic ultrasonic frequency within the range of the device gyroscope and movable lens resonant frequency. We assume that the adversary will create a network of nodes to cover an LOI area such that the target device is always within range of a transmitter. By doing video processing on-device and acquiring the SI, the adversary precludes video upload latency and bandwidth limitations. The adversary will likely turn on the video recording periodically, or as needed, so as not to consume too much target device resources.

IV. Problem Formulation

In this section, the occluded camera problem is formulated, and the constraints determining the practicality and scalability of the ultrasound transmitter network are defined.

1) Camera Block and Response Intensity: We define an occluded camera as a mobile device camera that has an occlusion on both the rear and front cameras. The universal set T_θ of target device T device positions, $t_\theta \in \{t_{\theta_1}, \ldots, t_{\theta_k}\}$ for K positions results in a signal μ_y from the camera where the linear response of the camera is for a normalized pixel intensity $x \in [0, 1]$ [14] modified by the device position t_θ. If the adversary can access the camera of a target mobile device, and the device is not occluded simultaneously at both the front and rear cameras, such as while the target device is in use, then either the front or rear camera can be used to surreptitiously make note of the target's indoor location, and no further method of inquiry may be required for target surveillance. We consider the threat model for $T_{\theta.B} \subset T_\theta$ where both the front and back cameras are simultaneously occluded for prolonged periods of time, such as when the target mobile device owner is ambulatory or stationary for 10 secs or longer in a GPS-denied environment. Studying the camera response under this entire set of device positions is out-of-scope for this paper, and accordingly we set t_θ as either open ($t_{\theta.O}$) or occluded ($t_{\theta.B}$). Some examples of how an occluded camera occurs naturally under normal use of a smartphone include, for example, placing a device a) in a bag, b) in a pocket, c) facedown on a desk with a privacy-oriented camera cover for the rear camera, or d) on its back inside a wallet case, which occludes the front camera.

For an occluded camera, the signal acquired by the adversary will be characterized by a $t_{\theta.B}$ such that $\mu_{y(x,t_{\theta.B})} << \mu_{y(x,t_\theta)}$, which we will refer to as μ_{blk} and μ_y, respectively. The mean intensity amplitude is modulated by an ultrasound signal at the resonant frequency of the device gyroscope and movable camera lens, and a set of configurable channel functions $F_c = \{f_{c_1}(.), \ldots, f_{cC}(.)\}$ for C functions under adversary control, where a given channel function f_c affects either the transmitted attack signal, or the modulated response of the transceiver. In Table I we present a summary of channel functions evaluated in this paper for both the transmitter and the transceiver, and describe their functionality for the threat model.

2) Device Localization: Given a set of known ultrasonic transmitter nodes η with a node η', we define the device localization problem with an upper bound l_{max} of distance for which a receiver can detect a transmitter signal in a given

[1]Frequencies used for evaluations withheld for responsible disclosure.

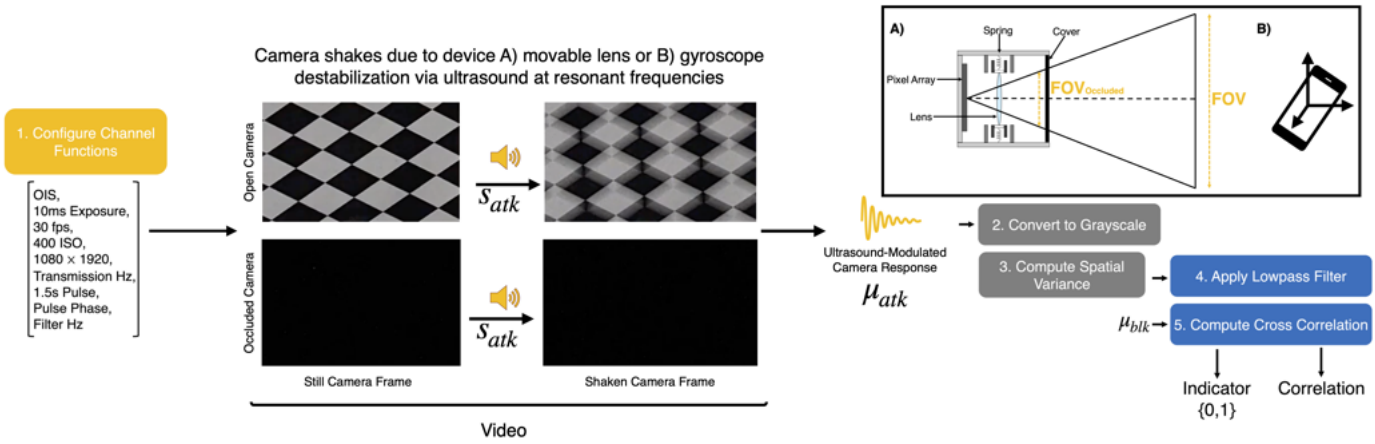

Fig. 2. **The Attack System Design.** The adversary configures channel functions for the ultrasound signal transmitter and the user-installed mobile application, which acts as a transceiver. The occluded camera signal amplitude is modulated by ultrasound (s_{atk}), and the adversary uses Steps 1-5 to localize a device using camera responses (μ_{atk}) from unique ultrasound patterns for each location. The occluded ambient acoustic noise (μ_{blk}) is acquired beforehand, and also processed using steps 2-4. The open camera response is shown for comparison. The underlying mechanism for the camera response which produces the ultrasound-modulated signals is destabilization of mechanisms used to correct camera blurs or motions, such as the device movable lens, or device gyroscope. A) An occluded camera field of view, $FOV_{Occluded}$, has neither depth nor spatial information, and the resultant camera response gives a unirradiated sensor image. The springs are part of the movable lens structure, and used for mechanical camera stabilization. B) Gyroscope-based camera stabilization is used by both OIS and EIS mechanisms.

TABLE I
THREAT MODEL CHANNEL FUNCTIONS

Channel Function	Parameters	Functionality	Controls
OIS	On/Off	Enable/Disable	Camera
EIS	On/Off	Enable/Disable	Camera
Exposure (ms)	{10,20,40,80,100}	Set	Camera
Frame Rate (fps)	30	Set	Camera
ISO	400	Set	Camera
Zoom	{1X, 3X, 12X}	Set	Camera
Frame Size (Pixels)	1080 × 1920	Set	Camera
Pulse Frequency (Hz)	{ ω_1 , . . . , ω_N }[1]	Set	Transmitter
Pulse Interval (s)	{1.5, 8}	Set	Transmitter
Pulse Pattern	{Pulse, Sweep}	Set	Transmitter
Pulse Start Phase (deg)	0	Set	Transmitter
Filter Range (Hz)	{ ω_1 , ω_2 }	Set	SI

system, and the system is viable for all distances $\leq l_{max}$. Each node η' transmits a signal as defined by the adversary, and there is no communication between the nodes. Let the ultrasonic signal modulating an image frame f be a sine wave over a time period t with amplitude A, phase φ, and frequency ω, $s_{atk} = sin(\varphi + \pi\omega t)A$. The attack sequence involves collecting an ambient acoustic noise signal from the target mobile device before pulsing all transmitters, and using a look-up table to identify the location using the signal correlation.

Signal Indicator. The adversary's goal is to configure a subset of channel functions to differentiate between μ_{blk} and μ_{atk}, the ultrasound modulated signal from an occluded camera, by using the signal indicator (SI),

$$SI(\mu_{in}) = \begin{cases} 1, & \text{if } R(\mu_{in}, \mu_{blk}) \neq R(\mu_{blk}, \mu_{blk}), \\ 0, & \text{if } R(\mu_{in}, \mu_{blk}) = R(\mu_{blk}, \mu_{blk}). \end{cases} \quad (1)$$

The SI compares the correlation between a given camera response, μ_{in}, and an existing ambient acoustic noise baseline signal, μ_{blk}, as described in Equation 1. The ambient baseline is fixed, and is acquired by the adversary before beginning the ultrasound transmission. In this paper we show that the camera destabilization frequency response to ultrasound at its resonant frequency creates minute movements of the camera lens such that the movement of the lens creates a camera response where the spatial variance of the pixel intensities can be filtered to a signal which is correlated differently to the ambient acoustic noise signal in that same setting.

Attack Range. The adversary can use the resonant frequencies of the device and camera gyroscopes via one of two modes, pulses or sweeps, which broadens the adversary's node network. For example, if a given device has multiple resonant frequencies, then the lower bound of the adversary's node network is $\omega \times \varphi \times 2$(modes), with a theoretical upper limit available through phase modulation of the signal [39], [40], [41]. As such, an adversary can create a network of nodes in a given location of interest, where each node is spaced by the maximum distance l_{max}. Note current ultrasonic transmitters have a 4 m to 11 m range [42], [43].

V. OPTICAL ACOUSTIC SIDE CHANNEL WITHOUT SCENE INFORMATION

In this section we describe the functional mapping of the ultrasonic signal from the transmitter, s_{atk}, to the SI of the observed occluded camera response, μ_{atk} : $s_{atk} \times (G_c \subset F_c) \rightarrow SI$, where G_c is a set of channel functions for the attack. The adversary is interested in acquiring a camera response μ_{atk} which has pixel intensities characterized by image sensor dark signals created during low light conditions from its position t_θ, and a subset of channel functions G_c for the transmitter and transceiver. The camera response under ultrasonic acoustic pressure is distinct from the image sensor dark signals created during low light conditions and ambient sound, μ_{blk}. The mean intensity of the camera response when

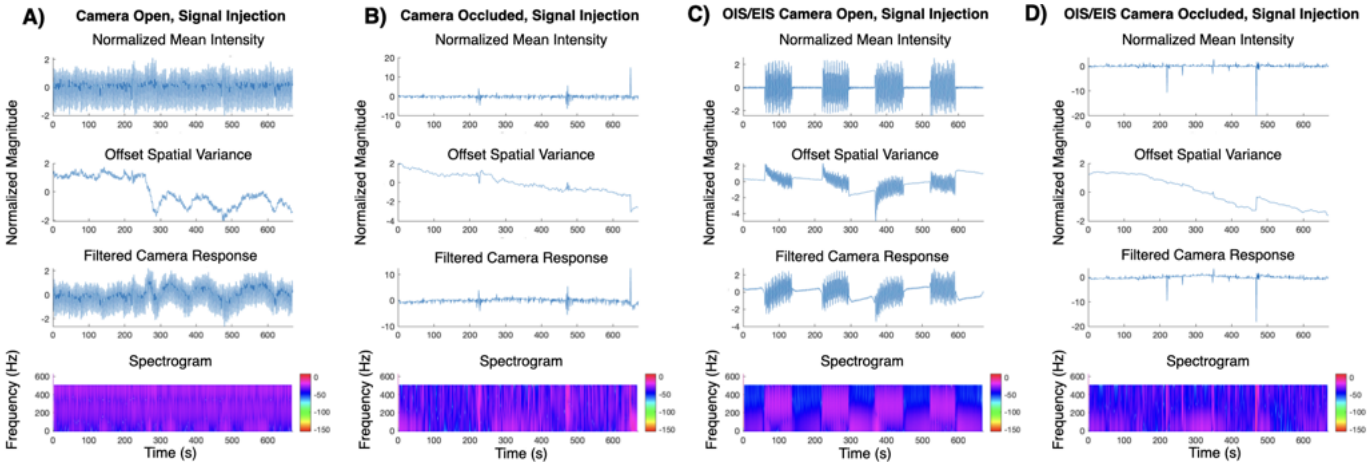

Fig. 3. **Open and Occluded Ultrasound-Modulated Camera Responses.** Normalized mean intensity for ultrasound signal induced camera responses does not vary as much as spatial variance for those same signals, as seen in B) and D). The filtered response is computed as the difference between the normalized mean intensity and the offset spatial variance, and then used as an input for the signal indicator function. Camera responses shown here are from videos recorded at 10 ms exposure at 30 fps. The y-axis scale between different camera responses is not normalized to provide a sense of the range of variation.

the camera is open and experiencing ambient sound is simply μ_y. We simulate the camera response under open and occluded ultrasound modulated conditions to understand how to develop a signal indicator which can differentiate between real-world ambient and ultrasound modulated camera responses.

A. Camera Ultrasound Response Model

Dark Signals. A camera works by irradiating an image sensor over exposure time, t_{exp}. The amount of subsequent charge generation resulting from the photoelectric effect following image sensor irradiation is dependent in part on light intensity and the number of active pixels, M $\times$ N, in a pixel array. Following charge collection and conversion over a number of steps in the sensor circuit, the resultant digital output signal μ_y from an image sensor is assumed to increase linearly with radiant exposure. However low light conditions, such as when the camera lens is occluded, result in lower amounts of photogenerated charge, and in this case the camera response is primarily characterized by dark current and image sensor noise sources. Dark current is charge randomly generated by electrons discharged from a pixel without sensor irradiation due to the thermal energy of the image sensor, and starts accumulating as soon as exposure begins [44], [16]. An occluded camera signal is in part produced by both dark current and an offset, the dark signal value at zero exposure which guides its fixed pattern noise [14]. Dark signal is amplified following longer collection and integration, such as by increasing t_{exp} directly, or by setting an exposure value by modifying the camera ISO. The camera ISO is a standardized scale for modifying the camera's exposure value [45]. Blocking a camera lens results in a diminished field of view, as seen in Figure 2A, and the new projection is an M $\times$ N dark frame, *f*, lacking both a dynamic range of pixel intensities and spatial information. The goal of the threat model is to modulate the limited information in the frame *f*.

Electronic Rolling Shutter Transceiver for Ultrasonic Signals. Whereas prior works have shown the amplification of pixel displacements in irradiated sensor images due to sound energy [46], [47], [48], [49], we find that the image sensor rolling shutter also encodes the amplitude variance of an ultrasonic signal, s_{atk}, as spatial information on a unirradiated sensor image frame. Assuming that the mean intensity of an open camera without noise sources remains stable overtime, then adding a sound which shifts the lens produces pixel deviations overtime in accordance with the lens movement, such that the mean camera response for f becomes $\mu_{atk} = sin(\varphi + pi\omega\varsigma)A$, where ς is $V\tau r$. The time series V $\left\{n \cdot \frac{1}{\lfloor \frac{r}{M} \rfloor} \mid n \in \mathbb{Z}, 0 \leq n \cdot \frac{1}{\lfloor \frac{r}{M} \rfloor} < 2\right\}$, where n is an integer, represents the change in mean pixel intensity for a given sequence of frames. The electronic rolling shutter readout, r, is the rate at which the shutter sequentially reads M rows, and the interval τ is the time between successive ultrasonic pulses in seconds. The camera response is subject to random and fixed sources of noise, which we include as,

$$\mu_{atk} = sin(\varphi + \pi\omega\varsigma)A + NF(\phi(m) + P^{\kappa}(\lambda)) \qquad (2)$$

where Gaussian (ϕ) and Poisson (P) noise can be parameterized using different values of m and λ to set the mean of their respective distributions. The NF is a noise factor which can be modified in place of independently configuring the parameters for the noise functions, and κ modifies the Poisson contribution of dark current shot noise, which may increase with exposure.

Given that the mean intensity of a signal includes sources of random and fixed noise, we compute the spatial variance of the signal to reduce the noise, and amplify the pixel intensity deviations. The gray values in two images taken at the same exposure will vary slightly because of temporal noise, but the pixel intensity offset of the image sensor ensures that the nonuniformity remains stationary across two different frames

as fixed pattern noise [14]. Consequently, the temporal noise variance across two images taken at the same exposure can be expressed as spatial variance with a pixel offset size O as,

$$s^2 = \frac{1}{MN}\left(\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} f_1[m]_{\delta_M}[n]_{\delta_N} f_2[m]_{\delta_M}[n]_{\delta_N}\right) - \mu_{f_1}\mu_{f_2} \tag{3}$$

where $\delta_M = M - O$ and $\delta_N = N - O$. Note that pixel intensity offset intrinsic to the image sensor is different from the pixel offset size O we use for computing the offset spatial variance. In general, a pixel from a occluded video frame may have intensities $m_i \geq 0, n_j \geq 0$ because of the dark signal pixel intensity offset and nonuniformity, various image sensor noise sources, and some light always incident to the sensor. In summary, mean intensity is the camera response which includes sources of both random and fixed noise, e.g., from readout [15]. We compute the spatial variance of each subsequent frame with a pixel offset size O determining the range of pixel rows and columns selected to average the signal response, which produces a L $\times$ 1 time series. The filtered camera response is acquired as the difference of the normalized mean intensity and offset spatial variance, and is used the input for the signal indicator function.

B. Attack Feasibility

In this section, we provide evidence as existence proof for the underlying mechanism of an optical acoustic side channel attack using image sensor dark signals guided by two research questions,

1) **RQ 1:** How can information still be captured from an optical system with an occluded camera lens? Our hypothesis is that dark signal pixel intensities produce a modulated camera response based on the camera destabilization from ultrasound. We validate this by using the μ_{atk} (Equation 2) modulation model to simulate camera response signals and compare them to signals acquired from open and occluded cameras.
2) **RQ 2:** How can occluded camera information be utilized in a device localization threat model? We hypothesize, that depending on the set of channel functions, the signal produced under ambient and ultrasound conditions will be uncorrelated, and thus using an ambient acoustic noise signal as a baseline will enable developing an indicator function to determine device proximity to a transmitter. To validate this hypothesis, we compute the correlation between the ultrasound-modulated camera response and the baseline ambient acoustic noise signal for a given set of channel functions.

Ultrasonic Modulation of Unirradiated Pixel Intensities. Our first goal is to find a clear camera response to utilize in a threat model as occluded camera signals can be noisy. Figure 3 compares the spatial variance, as defined in Equation 3, and the mean intensity of signals acquired from image sensors under both occluded and open lens conditions, with ultrasound. Figure 9, included in the Appendix, shows the same camera responses, but collected under ambient acoustic noise conditions. For the existence proof tests, a Samsung smartphone with an occluded camera lens is placed 0.5 cm from an ultrasonic speaker on a tripod, while a sine wave with a peak-to-peak voltage of 20 V is pulsed every 8 seconds; we increase the distance to 28 cm (0.28 m) for evaluations. We also record ambient recordings, where no ultrasound is played. We observe that the occluded camera signal lacks sufficient dynamic range, such that the normalized mean intensity of ultrasound modulated responses, μ_{atk}, does not vary over time relative to the normalized mean intensity of occluded ambient recordings, μ_{blk}, as seen in Figure 3 B and D.

While the occluded camera signal lacks dynamic range, the lens movement induced by the ultrasonic signal creates areas of varying pixel intensities within a frame. As a result, computing the spatial variance of the signal makes the modulation of the ultrasonic signal more apparent. Figure 3 shows greater variation in the spatial variance of μ_{atk} compared to μ_{blk}. When comparing the occluded camera scenarios, Figure 3 B and D, to the open camera scenarios, Figure 3 A and C, we observe regions of lens movement in the occluded ultrasound cases, coupled with shifts in frequencies in corresponding time points in the spectrograms. This suggests that the optical stabilization mechanisms are vulnerable to providing information about the acoustic signal, even when the camera lens is occluded. Previous research on optical acoustic side channels has typically relied on visual access to the scene for exploitation. However, this result indicates that visual scene access is not necessary for exploiting ultrasonic signals and dark frame signal analysis.

Ultrasound Modulated Camera Response vs. Intrinsic Camera Noise. The ultrasound signal in the occluded camera can be observed by comparing the modulated occluded and ambient occluded camera responses in the frequency domain. While the ambient occluded camera responses are dominated by readout noise, which occurs at periodic intervals, the ultrasound modulated filtered responses are relatively greater in magnitude, as seen in Figure 3, and have varying frequencies at intervals outside of readout noise time. Additionally, comparing the mean intensity of modulated occluded vs. ambient occluded signals shows that ultrasound can stabilize the signal to remove readout noise, as seen in Figure 5.

Camera Response Simulation. We develop a simulation of the open and occluded camera responses to determine how each optical system parameter is responsible for a given camera response. From our experiments, we know that when a camera is open and ultrasound is played, the lens jumps vigorously, which is a known effect of the gyroscope frequency response to sound at its resonant frequency [22], [50], [51]. In Figure 4, a μ_{atk} signal modulated by ultrasound and $G_c = \{$OIS, EIS, 10 ms exposure, 30 fps, 400 ISO, 1080 $\times$ 1920 image size, a transmission frequency, 8 secs pulse interval, and a filter frequency$\}$ is shown. The signal model, shown in Figure 4, is the modulated open camera response to destabilization without noise sources, where the peaks of the signal arise from destabilization of the optical system upon receiving ultrasound. The interconnecting arms between the

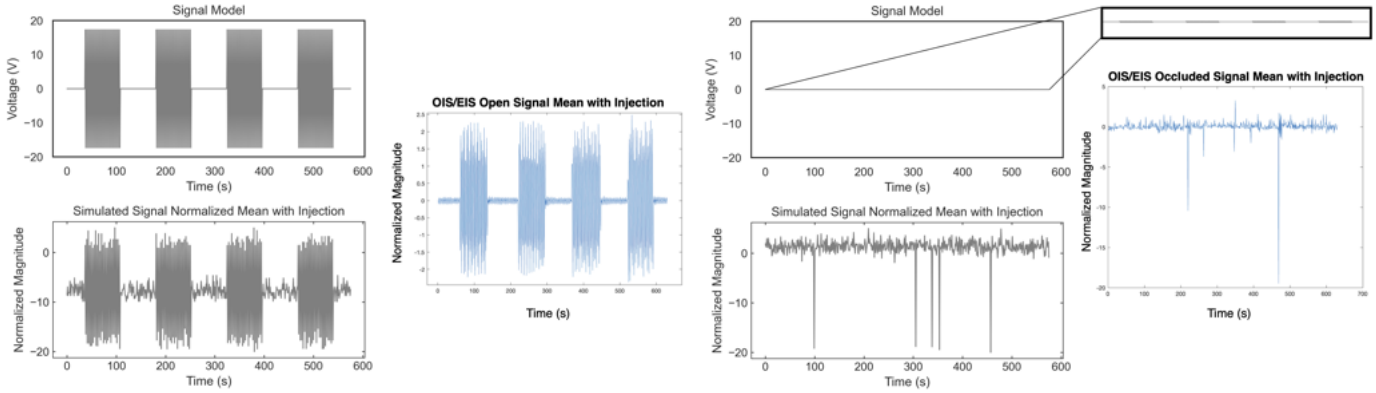

Fig. 4. **Simulated vs. Real Signals.** The μ_{atk} signal obtained from open and occluded OIS/EIS camera responses are simulated (in gray) and shown against their real counterparts (in blue). The real camera response shows when the camera shakes due to the received ultrasound, where the interlocking arms indicate when the sound is off. The simulated signal is produced by configuring rolling shutter rate r and ultrasound interval pulse τ, and then by adding noise to the signal model, as shown in Equation 2. The amplitude of the occluded signal model is very small in comparison, and shown in the zoom-in inset. For the occluded simulated camera response, the contribution of the Poisson noise was multiplied with -1 to achieve the downward spikes.

signal peaks are when ultrasound transmission is turned off. Removing the sensor readout frequency, r, and pulse interval τ produces a less stable looking signal which deviates from the camera response; this is shown in Figure 10 for a camera response over a longer time, given in the Appendix. Changing the noise parameters of Equation 2 changes the correlation of the simulated μ_{atk} to the real-world μ_{atk}, which is likely because dark current shot noise in the image sensor arises from a Poisson process [15], and all gyroscopes have Gaussian noise [52]. We then model the occluded camera response as a signal with a relatively lower amplitude as we do not expect nearly the same amplitude of pixel shifts as the open camera response to sound. Adding noise sources to the simulated occluded model produces a camera response which matches a real-world occluded camera response, for example as shown by comparing a simulated μ_{atk} normalized mean response to real-world μ_{atk} normalized mean response. Therefore, low-level pixel intensities in an occluded camera can be modulated using ultrasound in the absence of image sensor irradiation.

Movable Lens vs. Gyroscope Destabilization. There are at least two optical stabilization mechanisms an adversary can exploit in this threat model. This is confirmed by comparing occluded camera recordings to open camera recordings for both ultrasonic signals (μ_{atk}) and ambient recording (μ_{blk}) scenarios while configuring the image stabilization method. Figure 9 shows the ambient open and occluded camera responses, and is given in the Appendix. For ambient recodings, the experiment procedure is the same as described for the occluded camera video capture, except in this case ultrasound is not played while recording. Normally, when camera stabilization is enabled and ultrasound is played, a camera shake can be visually observed on screen when the camera is open, and as expected a camera shake does not appear when camera stabilization is disabled and ultrasound is played. However, it appears that lens destabilization from the ultrasound is apparent in image analysis from occluded camera recordings when both OIS and EIS are disabled, even though a camera shake does not visibly show up when the camera is open and OIS/EIS is disabled during ultrasound transmission video recording. We reason that a secondary mechanism is enabling lens movement in this case, which is likely the movable lens of the device camera. But it is unclear if the lens movement is due to a second camera gyroscope, or simply due to the movement of the lens springs. Additional analysis using gyroscope noise metrics is included in Appendix Section C.

TABLE II
MAXIMUM CORRELATION COEFFICIENT FOR A SUBSET OF OCCLUDED SIGNALS (0.5 CM, S2, PULSE)

Trial	Max Coeff.	Signal In	Baseline Signal	SI
1	158	{10ms,None,Signal}	{10ms,None,Ambient}	1
2	181	{10ms,EIS,Signal}	{10ms,EIS,Ambient}	1
3	108	{20ms,None,Signal}	{20ms,None,Ambient}	1
4	93	{20ms,EIS,Signal}	{20ms,EIS,Ambient}	1
5	82	{40ms,None,Signal}	{40ms,None,Ambient}	1
6	121	{40ms,EIS,Signal}	{40ms,EIS,Ambient}	1
7	119	{80ms,None,Signal}	{80ms,None,Ambient}	1
8	86	{80ms,EIS,Signal}	{80ms,EIS,Ambient}	1

C. Signal Indicator for Modulated Camera Responses

To use ultrasound modulated dark signals for a device localization attack, we develop a signal indicator to differentiate between an ultrasound modulated camera response vs. an ambient camera response. When sources of kinetic energy, such as sound, transfer that energy to a microelectromechanical sensor upon impact, the vibrational frequency of the sensor matches the frequency of the driving source of kinetic energy, but the sensor's new vibrational frequency has a lag relative to the frequency of the sound source [53]. If the occluded camera pixel intensities do get modulated by ultrasound signal, due to either the gyroscope or the movable lens springs being struck by ultrasound, then the camera response signal will be correlated to the attack signal. Cross-correlation is an operation which can determine the displacement of one time series relative to another, and is often used as a measure of

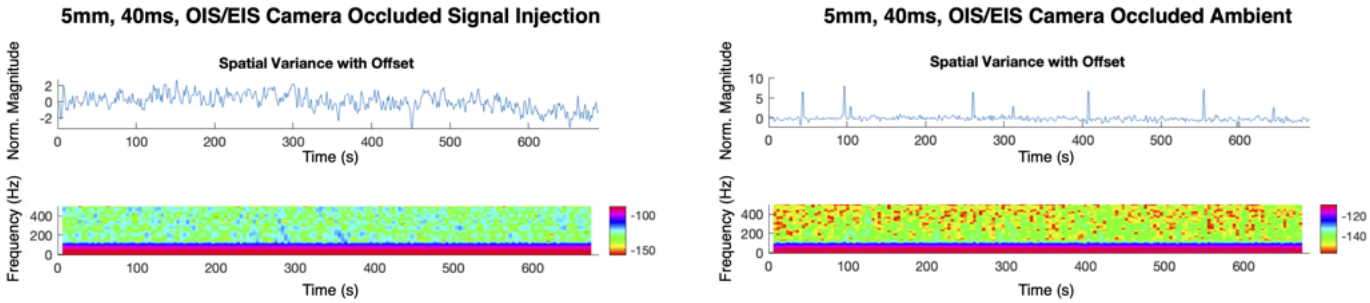

Fig. 5. **OIS/EIS Spatial Variance and Spectrograms for 40 ms Signals.** While signal spikes do not show up clearly over the time-axis for the modulated signal, the spectrograms of the modulated and ambient acoustic noise signals show different frequency content. The ambient acoustic noise signal shows readout spikes, but readout noise is suppressed in the ultrasound modulated signal, though its signal amplitude is lower. Ultrasound modulated signals seem to experience an increase or decrease in amplitude, likely due to the phase of the incoming sound wave.

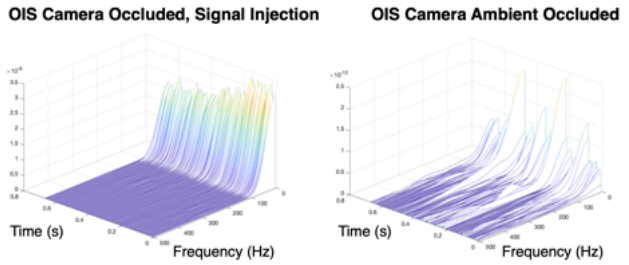

Fig. 6. **3D View of 40 ms exposure OIS Signal Frequencies.** Visualizing the signals using a three-dimensional view of the frequency domain helps show the extent of their modulation from ultrasound. The amplitude of the occluded ambient acoustic noise signal is lower, and shows readout spikes.

signal similarity. The existence proof experiments, Figures 3 and 5, that two camera responses acquired under the same conditions, but one modulated with ultrasound pulses, have different frequency spectra. This explains why two ambient acoustic noise signals are more correlated versus signals where one is ambient and the other is modulated by ultrasound, as shown in Table II; additional tables in Appendix Section A.

VI. Evaluations

In this section, we describe the experimental methodology to confirm our findings from the existence proof analysis of RQ1 and RQ2. This is done by implementing the methodology as outlined in Section IV-2 to enable the adversary to estimate the cross-correlation of transceiver signals and a baseline ambient acoustic noise signal for a given set of channel functions. Real-world camera responses modulated by real-world s_{atk} are evaluated by configuring the channel functions shown in Table I.

A. Experimental Methodology

Generating Ultrasound. Ultrasound is generated using a DG5072 function generator (2-channel, 70 MHz) to produce a sine wave with a peak-to-peak amplitude of 20 V and start phase of $0\,\mathrm{deg}$, and transmitted through a Vifa omni-directional ultrasonic speaker (Part 60409). A portable single-channel ultrasonic power amplifier is used for the speaker, powered by an external AC power supply. A digital laser

Fig. 7. **Occluded and Open Camera Setup.** Occluded camera is on the left, open camera is on the right. An on-device application is used to record the experienced lack of sensor irradiation by the rear camera. Lux is a unit of sensor illumination, and a lux of ≈ 0 implies the sensor is not receiving enough light to change the reading for the on-device application.

measure helps confirm distance between the ultrasonic speaker and the device, where the device is held in place using a tabletop tripod at 0.5 cm (0.005 m) and 28 cm (0.28 m) from the ultrasonic speaker.

Rear Camera Blocking. To emulate low light conditions and encourage dark signal generation, the rear camera of a target device is covered with aluminum foil and secured in place with duct tape. By using an on-device light meter application, the experienced irradiation of the main device rear camera is recorded. The camera blocking is shown in Figure 7, along with the in-app sensor irradiation metrics. Note that some light will always irradiate an image sensor through small gaps between the contact surfaces of a cover and the device, or pinhole-sized gaps in the weaves of fabric threads. The evaluation experiments are performed on two sensors of the rear camera of a Samsung Galaxy S20 5G using three different combinations for optical and hybrid zoom, as indicated in

TABLE III
ZOOM SETTING AND DEVICE IMAGE SENSOR

Zoom	Sensors	Type
1X	*S2*	Optical
3X	*S2*	Hybrid
12X	*S1*, *S2*	Hybrid

Table III. While not used for evaluations, we try to reproduce the ultrasonic attack on the iPhone 15 Pro Max (2023).

Data Collection. The exposure and image stabilization configuration of the target device is varied using Mobile AR Sensor (MARS) Logger [54], or the default camera. The ISO and frame rate values are held constant at 400 and 30 fps, respectively. The open rear camera is focused on an image pattern, or a plain white wall. For open or occluded camera recordings with ultrasound, the signal is pulsed at 1.5 secs or 8 secs intervals (Pulse Transmission). A 5 secs linear sweep is generated using the function generator with start and end frequencies to capture at least two resonant frequencies of the device gyroscope (Sweep Transmission). Our classification accuracy is on the entire dataset, unless stated otherwise (n=119), and videos are collected over multiple days.

Signal Processing and Classification. We developed a program in MATLAB to simulate and evaluate the signal indicator. The camera response magnitude is normalized, and the offset spatial variance is filtered using a lowpass filter. The filtered spatial variance is then subtracted from the normalized camera response, μ_{atk}. We then compute the cross-correlation of the aligned μ_{atk} and baseline ambient acoustic noise signals. The signal indicator is simulated by holding a given ambient signal for a camera configuration as a baseline, and then passing in an input ultrasound-modulated signal.

B. Results

Ultrasound-Modulated Camera Responses. We evaluate the ultrasound vs. ambient response signal indicator using different combinations of optical zoom configurations to determine if the ultrasound modulated camera response is an intrinsic property of the optical system, or if it instead depends on the set of channel functions used. The image sensors, with their respective pixel pitch sizes in parentheses, are 12MP IMX555 (1.8 μm) and 64MP S5KGW2 (0.8 μm), where MP stands for megapixels, are denoted as *S1*, *S2*, respectively. The 12MP IMX555 (*S1*) is used for all experiments in subsection V-B. Given the variation in readout frequencies across different image sensors, and their total M $\times$ N pixel areas, random image sensor noise could potentially interfere with ultrasonic signal modulation. However, the signal indicator remains robust across the different optical zooms, at both distances as seen in Table IV, likely because the readout frequency of the image sensor compensates for image noise sources during pixel intensity modulation. While it is true that image sensors with smaller pixel pitches may experience higher levels of flicker noise [55], [56], the correlation between μ_{atk} and the baseline ambient acoustic noise signal remains distinct even at a distance of 0.28 m, suggesting that the signal modulation persists regardless of the distances tested. Given that currently available commercial ultrasonic transmitters have a range of 4 m to 11 m [42], [43], these results indicate that the threat model can be effectively executed across larger image sensor areas and with smaller pixel pitches, and across different distances, but a feasibility study with real-world device localization parameters will better confirm these findings. In total 119 videos of $\approx 15-32s$ are analyzed, and a subset of the results are presented in Tables II and IV, and Tables VI and VII in the Appendix. Increasing the exposure seems detrimental to receiving the transmitted signal as the frequency spectra of the ultrasound-modulated occluded 80 ms and 160 ms exposure time responses is similar to that of occluded ambient camera responses. This is most likely due to increased sensor irradiation clearing out the received or incoming signal.

TABLE IV
MAXIMUM CORRELATION COEFFICIENT FOR DIFFERENT OPTICAL ZOOMS AND ULTRASONIC TRANSMISSION PATTERNS

Trial	Max Coeff.	Signal In	Baseline Signal	SI
1	119	{1X,None,5cm,Sweep,Signal}	{1X,None,5cm,Sweep,Ambient}	1
2	669	{1X,None,5cm,Sweep,Ambient}	{1X,None,5cm,Sweep,Ambient}	0
3	99	{3X,None,5cm,Sweep,Signal}	{3X,None,5cm,Sweep,Ambient}	1
4	669	{3X,None,5cm,Sweep,Ambient}	{3X,None,5cm,Sweep,Ambient}	0
5	90	{12X,None,5cm,Sweep,Signal}	{12X,None,5cm,Sweep,Ambient}	1
6	669	{12X,None,5cm,Sweep,Ambient}	{12X,None,5cm,Sweep,Ambient}	0
7	578	{1X,None,28cm,Pulse,Signal}	{1X,None,28cm,Pulse,Ambient}	1
8	668	{1X,None,28cm,Pulse,Ambient}	{1X,None,28cm,Pulse,Ambient}	0

Camera Responses from Different Speaker Directions. The results indicate that ultrasound modulated dark signals can behave as carrier waves. But it is unclear if the results would hold the same when a shorter pulse is used, or if the ultrasound modulated camera responses vary enough depending on the direction of the sound source. One way to test for this is to use a shorter pulse duration while varying the location of the sound source. A brief evaluation showed that a shorter 1.5 secs pulse produced at 0.5 cm distance from three different directions, 1) behind the device, 2) near the camera edge, and 3) in front of the device, showed that the ultrasound-modulated dark signal response could be produced from all three locations; Figure 11 shows this in the Appendix.

Ultrasonic Signal Patterns for Location Identification. Given a large enough location of interest (LOI), the adversary needs as many transmitters as there are areas to cover, where the signal attack range can be determined by the adversary as discussed in Section IV-2. The MEMS gyroscope in the device we used for our experiments is resonant at multiple ultrasonic frequencies. We evaluated two ultrasonic signal patterns without any optical stabilization configurations to validate if the ultrasonic modulated signal could differentiate enough because of the attack signal ultrasound pattern, and if that difference would reflect in the signal correlation. The ultrasonic patterns were produced by the following settings: 1) {1X,None,Pulse}, and 2){1X,None,Sweep}, which are listed as trials 1 and 7, respectively, in Table IV. The look-up table for the signal indicator may look like Table IV, from which the adversary can identify different locations by comparing the maximum

correlation coefficients for different ultrasound transmission patterns for the same camera configurations. Based on our limited dataset, our device localization simulation accuracy is 99.9%, but it depends on the fixed ambient signal baseline prior to providing the ultrasound-modulated signal.

VII. Discussion

Here we provide guidance on protecting against dark signal ultrasound-mediated device localization attacks.

No Guarantee of Privacy. This work demonstrates that blocking a smartphone camera and denying microphone access does not eliminate the risk of exploitation through an optical acoustic side channel. A key observation is that producing an ultrasound-modulated signal requires an uncovered mobile device edge near the camera, even if the front and back camera lens are occluded. Therefore, mobile devices may be protected from ultrasonic attacks by using a protective cover which encases all sides of the device. We note that the iPhone 15 Pro Max (2023) did not yield any indication of being susceptible to an ultrasonic attack for frequencies tested between 18-120 kHz, likely due to its device body material, or the location of the gyroscope within the device.

Dark Signal Optical Acoustic Side Channel. The central contribution of this paper is the discovery that an optical acoustic side channel does not require scene information to function as a threat, and can operate using low-level image sensor irradiation. We use device localization as a case study where the ultrasonic transmitter is an external device, however we observe that the ultrasonic signal can be recorded via the device microphone when microphone permissions are granted. Therefore, we note the possibility of using videos recorded on a target mobile device as sources of ultrasound signals to carry out either device localization, or attacks such as SurfingAttack on voice assistants [57]. A possible defense against these attacks would be for device manufacturers to enable noise reduction algorithms by default, ensuring that pixel intensities in unirradiated sensor images are less exploitable in such threat models.

VIII. Limitations and Future Work

Device Localization. The performance of our signal indicator relies on the attack sequence, i.e., collecting an ambient acoustic noise signal as baseline before activating all transmitters, and this performance is based on a simulation in MATLAB, as per the scope of this existence proof study. We use one speaker and a stationary target for our existence proof work, but a feasibility study could use a grid of ultrasonic transmitters with mobile and stationary targets. A future real world implementation can validate larger speaker-camera distances using specialized high-power speakers. Extensive previous works have shown how simply increasing the output power of speakers could effectively increase the range of acoustic interference on sensors hardware [20], [22], [50], [58]. More diverse types of devices equipped with cameras besides smartphones, such as IoT home cameras, smart screens, etc., can be tested following the methodology of this work. While we collected 119 different video samples for different permutations of each channel configuration, each channel configuration is only one video, though 15-32s, and sampled at 30 fps. As such, the reliability of the signal indicator to identify a location based on the signal correlation could be further evaluated in a feasibility study. Note that MEMS gyroscopes can wear out from prolonged ultrasound exposure, but that does not preclude multiple attack vectors which exploit them [59], [60]. Note that for all methods which target a device, the target device location is not necessarily the target device owner location, which an adversary may be more interested in than the target device location.

Dark Signal Modulation. Dark signals exist in all image sensors [15], [44], [61], and do not require anything external for generation. In a sense, the physical camera block is a filter, and its effects can also be modeled via software for an open camera. Such signal modulation can help an adversary control the optical system output using understanding of its semiconductor-level functionality, and perhaps add or remove image information, including watermarking [62], from cameras in real time even when the camera lens is not occluded. Our paper uses ultrasound as a dark signal modulation technique, but other methods for dark signal modulation may use RF, including microwave hyperthermia [63], [64], or simulating extrapolated dark signals over increased exposure times. While CMOS image sensor communication uses light [65], our dark signal approach opens possibilities to transform image sensor noise into usable information for communication by demodulating the transmitted signal.

IX. Conclusion

This study provides existence proof of an optical acoustic side channel that does not rely on direct or indirect scene information, and can be exploited using dark signals intrinsic to an image sensor. MEMS gyroscopes and movable lens mechanisms appear vulnerable to ultrasonic exploitation, even with occluded lenses, and such destabilization may enable device localization attack. Our findings challenge the assumption that emerging threat models must rely on advanced machine learning or deep learning techniques, which are resource intensive relative to our methodology. We found that ultrasound-modulated signals could be observed across different optical and hybrid zoom settings, indicating that the camera response is produced by intrinsic optical system parameters, such as the rolling shutter rate and image sensor size, as confirmed by a simulation model. Future research remains to confirm the feasibility of device localization with a dark signal optical acoustic side channel attack.

Acknowledgments

The authors appreciate the insights and remarks from our reviewers. This work was supported in part by the National Science Foundation Industry-University Cooperative Research Centers Program under grant IUCRC-1916762, and by the Center for Hardware and Embedded System Security and Trust (CHEST) Industry Fund.

REFERENCES

[1] M. Furini, S. Mirri, M. Montangero, and C. Prandi, "Privacy Perception when Using Smartphone Applications," *Mobile Networks and Applications*, vol. 25, pp. 1055–1061, June 2020.

[2] A. M. V. V. Sai and Y. Li, "A survey on privacy issues in mobile social networks," *IEEE Access*, vol. 8, pp. 130906–130921, 2020.

[3] H. Li, H. Zhu, S. Du, X. Liang, and X. Shen, "Privacy leakage of location sharing in mobile social networks: Attacks and defense," *IEEE Transactions on Dependable and Secure Computing*, vol. 15, no. 4, pp. 646–660, 2016.

[4] G. Guo, R. Chen, K. Yan, Z. Li, L. Qian, S. Xu, X. Niu, and L. Chen, "Large-scale indoor localization solution for pervasive smartphones using corrected acoustic signals and data-driven pdr," *IEEE Internet of Things Journal*, vol. 10, pp. 15338–15349, 2023.

[5] N. M. C. Tiglao, M. I. Alipio, R. D. Cruz, F. S. Bokhari, S. Rauf, and S. A. Khan, "Smartphone-based indoor localization techniques: State-of-the-art and classification," *Measurement*, 2021.

[6] K. Liu, X. Liu, and X. Li, "Guoguo: Enabling fine-grained smartphone localization via acoustic anchors," *IEEE transactions on mobile computing*, vol. 15, no. 5, pp. 1144–1156, 2015.

[7] J. Kang, D. Steiert, D. Lin, and Y. Fu, "Movewithme: Location privacy preservation for smartphone users," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 711–724, 2020.

[8] A. Konstantinidis, G. Chatzimilioudis, D. Zeinalipour-Yazti, P. Mpeis, N. Pelekis, and Y. Theodoridis, "Privacy-preserving indoor localization on smartphones," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 11, pp. 3042–3055, 2015.

[9] J. Morris, S. Newman, K. Palaniappan, J. Fan, and D. Lin, ""do you know you are tracked by photos that you didn't take": large-scale location-aware multi-party image privacy protection," *IEEE Transactions on Dependable and Secure Computing*, vol. 20, no. 1, pp. 301–312, 2021.

[10] F. Li, Z. Sun, A. Li, B. Niu, H. Li, and G. Cao, "Hideme: Privacy-preserving photo sharing on social networks," in *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*, pp. 154–162, IEEE, 2019.

[11] P. Ilia, I. Polakis, E. Athanasopoulos, F. Maggi, and S. Ioannidis, "Face/off: Preventing privacy leakage from photos in social networks," in *Proceedings of the 22nd ACM SIGSAC Conference on computer and communications security*, pp. 781–792, 2015.

[12] C. W. Chen, "Internet of video things: Next-generation iot with visual sensors," *IEEE Internet of Things Journal*, vol. 7, pp. 6676–6685, 2020.

[13] A. Mohan, K. Gauen, Y.-H. Lu, W. W. Li, and X. Chen, "Internet of video things in 2030: A world with many cameras," in *2017 IEEE international symposium on circuits and systems (ISCAS)*, pp. 1–4, IEEE, 2017.

[14] "Emva standard 1288 standard for characterization of image sensors and cameras," 2021.

[15] M. Konnik and J. Welsh, "High-level numerical simulations of noise in CCD and CMOS photosensors: Review and tutorial," Dec. 2014.

[16] R. L. Baer, "A model for dark current characterization and simulation," in *Electronic imaging*, 2006.

[17] "The visual microphone: Passive recovery of sound from video: ACM Transactions on Graphics: Vol 33, No 4."

[18] G. Zhu, X.-R. Yao, Z. Sun, P. Qiu, C. Wang, G.-J. Zhai, and Q.-G. Zhao, "A high-speed imaging method based on compressive sensing for sound extraction using a low-speed camera," *Sensors (Basel, Switzerland)*, vol. 18, 2018.

[19] D. Zhang, J. Guo, Y. Jin, and C. Zhu, "Efficient subtle motion detection from high-speed video for sound recovery and vibration analysis using singular value decomposition-based approach," *Optical Engineering*, vol. 56, 2017.

[20] Y. Long, P. Naghavi, B. Kojusner, K. Butler, S. Rampazzi, and K. Fu, "Side Eye: Characterizing the Limits of POV Acoustic Eavesdropping from Smartphone Cameras with Rolling Shutters and Movable Lenses," in *2023 IEEE Symposium on Security and Privacy (SP)*, (San Francisco, CA, USA), pp. 1857–1874, IEEE, May 2023.

[21] W. Zhu, X. Ji, Y. Cheng, S. Zhang, and W. Xu, "Tpatch: A triggered physical adversarial patch," in *USENIX Security Symposium*, 2023.

[22] X. Ji, Y. Cheng, Y. Zhang, K. Wang, C. Yan, W. Xu, and K. Fu, "Poltergeist: Acoustic Adversarial Machine Learning against Cameras and Computer Vision," in *2021 IEEE Symposium on Security and Privacy (SP)*, (San Francisco, CA, USA), pp. 160–175, IEEE, May 2021.

[23] M. R. e Souza, H. de Almeida Maia, and H. Pedrini, "Survey on digital video stabilization: Concepts, methods, and challenges," *ACM Computing Surveys (CSUR)*, vol. 55, pp. 1 – 37, 2022.

[24] F. L. Rosa, M. Virzi, F. Bonaccorso, and M. Branciforte, "Optical image stabilization (ois)," 2015.

[25] J. Xiong, "Overview of the electronic image stabilization technology," *Optics and Precision Engineering*, 2001.

[26] R. S. Naser, M. C. Lam, F. Qamar, and B. B. Zaidan, "Smartphone-based indoor localization systems: A systematic literature review," *Electronics*, 2023.

[27] G. M. Mendoza-Silva, J. Torres-Sospedra, and J. Huerta, "A meta-review of indoor positioning systems," *Sensors (Basel, Switzerland)*, vol. 19, 2019.

[28] Y. Fukuju, M. Minami, H. Morikawa, and T. Aoyama, "Dolphin: an autonomous indoor positioning system in ubiquitous computing environment," *Proceedings IEEE Workshop on Software Technologies for Future Embedded Systems. WSTFES 2003*, pp. 53–56, 2003.

[29] M. T. Chew, F. Alam, M. Legg, and G. S. Gupta, "Accurate ultrasound indoor localization using spring-relaxation technique," *Electronics*, vol. 10, p. 1290, 2021.

[30] A. L. Cretu-Sîrcu, H. Schiøler, J. P. Cederholm, I. Sîrcu, A. Schjørring, I. R. Larrad, G. Berardinelli, and O. Madsen, "Evaluation and comparison of ultrasonic and uwb technology for indoor localization in an industrial environment," *Sensors (Basel, Switzerland)*, vol. 22, 2022.

[31] T.-C. Yang, A. Cabani, and H. Chafouk, "A survey of recent indoor localization scenarios and methodologies," *Sensors (Basel, Switzerland)*, vol. 21, 2021.

[32] B. Priyantha, "The cricket indoor location system," 2005.

[33] M. A. Ferrag, L. Maglaras, A. Derhab, and H. Janicke, "Authentication schemes for smart mobile devices: Threat models, countermeasures, and open research issues," *Telecommunication Systems*, vol. 73, pp. 317–348, Feb. 2020.

[34] L. Wu, X. Du, and X. Fu, "Security Threats to Mobile Multimedia Applications: Camera-Based Attacks on Mobile Phones," *Communications Magazine, IEEE*, vol. 52, pp. 80–87, Mar. 2014.

[35] W. Cao, C. Xia, S. T. Peddinti, D. Lie, N. Taft, and L. M. Austin, "A Large Scale Study of User Behavior, Expectations and Engagement with Android Permissions,"

[36] F. Han, L. Xie, Y. Yin, H. Zhang, G. Chen, and S. Lu, "Video stabilization for camera shoot in mobile devices via inertial-visual state tracking," *IEEE Transactions on Mobile Computing*, vol. 20, pp. 1714–1729, 2021.

[37] H. Ovrén and P.-E. Forssén, "Gyroscope-based video stabilisation with auto-calibration," *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2090–2097, 2015.

[38] C.-R. Lee, J. H. Yoon, M. Park, and K. jin Yoon, "Gyroscope-aided relative pose estimation for rolling shutter cameras," *ArXiv*, vol. abs/1904.06770, 2019.

[39] J. Guedey, Y. Deval, H. Lapuyade, and F. Rivet, "Binary phase-shift keying for ultrasonic intra-body area networks," in *2020 IEEE MTT-S International Microwave Biomedical Conference (IMBioC)*, pp. 1–3, 2020.

[40] W. Jiang and W. M. D. Wright, "Indoor airborne ultrasonic wireless communication using ofdm methods," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 64, no. 9, pp. 1345–1353, 2017.

[41] C. Li, D. A. Hutchins, and R. J. Green, "Short-range ultrasonic communications in air using quadrature modulation," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 56, no. 10, pp. 2060–2072, 2009.

[42] "Rosemounttm 3101, 3102, and 3105 level transmitters.,"

[43] "Hawkultra one ultrasonic level transmitter.,"

[44] L. Li, M. Li, Z. Zhang, and Z.-L. Huang, "Assessing low-light cameras with photon transfer curve method," *Journal of Innovative Optical Health Sciences*, vol. 09, p. 1630008, May 2016.

[45] S. R. Teli, S. Zvanovec, and Z. Ghassemlooy, "The first tests of smartphone camera exposure effect on optical camera communication links," in *2019 15th International Conference on Telecommunications (ConTEL)*, pp. 1–6, 2019.

[46] Y. Fuse, Y. Yasumi, and T. Takiguchi, "Sound recovery using vibration modes of the object in a video," *2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pp. 2027–2031, 2018.

[47] Y. Yasumi and T. Takiguchi, "Visual sound recovery using momentary phase variations," 2017.
[48] M. Soledad, "Analyzing vibrating objects from video," 2016.
[49] J. Quisberth, Z. Wang, and H. Nguyen, "Acquisition of audio information from silent high speed video," 2016.
[50] Y. Son, H. Shin, D. Kim, Y.-S. Park, J. Noh, K. Choi, J. Choi, and Y. Kim, "Rocking drones with intentional sound noise on gyroscopic sensors," in *USENIX Security Symposium*, 2015.
[51] Y. Tu, Z. Lin, I. Lee, and X. S. Hei, "Injected and delivered: Fabricating implicit control over actuation systems by spoofing inertial sensors," in *USENIX Security Symposium*, 2018.
[52] C. Shen, J. Li, X. Zhang, Y. Shi, J. Tang, H. Cao, and J. Liu, "A Noise Reduction Method for Dual-Mass Micro-Electromechanical Gyroscopes Based on Sample Entropy Empirical Mode Decomposition and Time-Frequency Peak Filtering," *Sensors (Basel, Switzerland)*, vol. 16, p. 796, May 2016.
[53] M.-H. Bao, *Basic mechanics of beam and diaphragm structures*, p. 23–88. Elsevier, 2000.
[54] "OSUPCVLab/mobile-ar-sensor-logger." OSU Photogrammetric Computer Vision Lab., Nov. 2024.
[55] P. Martin-Gonthier and P. Magnan, "Cmos image sensor noise analysis through noise power spectral density including undersampling effect due to readout sequence," *IEEE Transactions on Electron Devices*, vol. 61, no. 8, pp. 2834–2842, 2014.
[56] P. B. Catrysse and B. A. Wandell, "Roadmap for cmos image sensors: Moore meets planck and sommerfeld," in *IS&T/SPIE Electronic Imaging*, 2005.
[57] Q. Yan, K. Liu, Q. Zhou, H. Guo, and N. Zhang, "Surfingattack: Interactive hidden attack on voice assistants using ultrasonic guided waves," *Proceedings 2020 Network and Distributed System Security Symposium*, 2020.
[58] 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)*, pp. 3–18, IEEE, 2017.
[59] T. Liu, Z. Hong, and H. Chen, "A traceability localization method of acoustic attack source for mems gyroscope," *IEEE Embedded Systems Letters*, vol. 15, pp. 13–16, 2023.
[60] S. Khazaaleh, G. Korres, M. Eid, M. S. Rasras, and M. F. Daqaq, "Vulnerability of mems gyroscopes to targeted acoustic attacks," *IEEE Access*, vol. 7, pp. 89534–89543, 2019.
[61] P. Martin-Gonthier, R. Molina, P. Cervantes, and P. Magnan, "Analysis and optimization of noise response for low-noise cmos image sensors," in *10th IEEE International NEWCAS Conference*, pp. 513–516, 2012.
[62] Y. Zhao, B. Liu, M. Ding, B. Liu, T. Zhu, and X. Yu, "Proactive deepfake defence via identity watermarking," in *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 4591–4600, 2023.
[63] H. Xiong, J. Xie, Y. Liu, B. Wang, D. Xiao, and H. Zhang, "Microwave hyperthermia technology based on near-field focused metasurfaces: Design and implementation," *Advanced Functional Materials*, Aug. 2024.
[64] X. He, W. Geyi, and S. Wang, "Optimal design of focused arrays for microwave-induced hyperthermia," *IET Microwaves, Antennas & Propagation*, vol. 9, no. 14, pp. 1605–1611, 2015.
[65] R. Huang and T. Yamazato, "A review on image sensor communication and its applications to vehicles," *Photonics*, vol. 10, no. 6, 2023.
[66] D. V. Fedasyuk and T. Marusenkova, "Mems gyroscopes' noise simulation algorithm," in *International Conference on Computer Science and Information Technologies*, 2019.
[67] E. Tatar, T. Mukherjee, and G. K. Fedder, "Stress effects and compensation of bias drift in a mems vibratory-rate gyroscope," *Journal of Microelectromechanical Systems*, vol. 26, no. 3, pp. 569–579, 2017.

TABLE V
GYROSCOPE NOISE METRICS.

Device State	Bias Instability (deg/hr) ↑			Angle Random Walk (deg/$\sqrt{hr}$) ↑		
	x	*y*	*z*	*x*	*y*	*z*
None Signal Occluded	6.0401e-05	5.8815e-04	1.4290e-04	1.0119	16.8445	0.0235
None 25kHz Occluded	3.6118e-05	8.8339e-05	7.7392e-05	0.0015	0.0016	0.0020
None Ambient Occluded	7.4342e-05	5.9864e-05	2.0927e-04	0.0020	0.0017	0.0022
OIS Signal Occluded	0.0014	0.0028	0.0012	0.3650	8.2023	0.0120
OIS 25kHz Occluded	5.5501e-05	1.6790e-04	5.5764e-04	0.0017	0.0018	0.0021
OIS Ambient Occluded	2.6487e-04	6.0405e-04	2.4543e-04	0.0016	0.0016	0.0021

TABLE VI
MAXIMUM CORRELATION COEFFICIENT FOR A SUBSET OF OCCLUDED ULTRASOUND MODULATED SIGNALS (0.5 CM, S2, PULSE)

Trial	Max Coeff.	Signal In	Baseline Signal	SI
1	669	{10 ms,OIS}	{10 ms,OIS}	0
2	668	{10 ms,OIS/EIS}	{10 ms,OIS/EIS}	0
3	668	{20 ms,OIS}	{0 ms,OIS}	0
4	669	{20 ms,OIS/EIS}	{20 ms,OIS/EIS}	0
5	669	{40 ms,OIS}	{40 ms,OIS}	0
6	669	{40 ms,OIS/EIS}	{40 ms,OIS/EIS}	0
7	669	{160 ms,OIS}	{160 ms,OIS}	0
8	669	{160 ms,OIS/EIS}	{160 ms,OIS/EIS}	0

APPENDIX

A. Evaluating the Signal Indicator.

The maximum of the correlation coefficient vector provides a way to compare the correlation of any two given signals. When the signal indicator receives the same signal as the baseline, as seen in Table IV, the maximum coefficient is relatively higher than when it receives a different signal vs. the baseline, as seen in Table II. The frequency spectra of the occluded signal and occluded ambient conditions, as shown in Figure 6, answers in part why this may be the case. The other reason is that enabling different camera options encodes each signal differently, which seems to be the case as the normalized mean intensity of different signals in Figure 3 varies even for ambient conditions. Setting a camera to OIS/EIS makes it easier to modulate pixel intensities, as seen for the open camera signal injection responses in Figure 3, likely because the combination of hardware and software image stabilization amplifies the incoming signal.

B. Open and Occluded Camera Experiments.

By comparing Figures 3 and 9, it is apparent that the offset spatial variance is better at differentiating between occluded ambient (μ_{blk}) and occluded signal camera (μ_{atk}) responses. In general, spatial variance makes it easier to distinguish between ambient and ultrasound-modulated responses. But the normalized mean intensity signal was used to simulate the camera response because of the clear visual difference in pixel intensities due to the camera shake induced by ultrasound. Cross correlation for open and occluded camera responses was determined using the simulated and real-world signals, as shown in Figure 10; a range of parameter values was tested for each of the parameters shown.

TABLE VII
MAXIMUM CORRELATION COEFFICIENT FOR A SUBSET OF OCCLUDED AMBIENT ACOUSTIC NOISE SIGNALS (0.5 CM, S2)

Trial	Max Coeff.	Signal In	Baseline Signal	SI
1	668	{10 ms,OIS}	{10 ms,OIS}	0
2	668	{10 ms,OIS/EIS}	{10 ms,OIS/EIS}	0
3	669	{20 ms,OIS}	{0 ms,OIS}	0
4	669	{20 ms,OIS/EIS}	{20 ms,OIS/EIS}	0
5	669	{40 ms,OIS}	{40 ms,OIS}	0
6	669	{40 ms,OIS/EIS}	{40 ms,OIS/EIS}	0
7	668	{160 ms,OIS}	{160 ms,OIS}	0
8	669	{160 ms,OIS/EIS}	{160 ms,OIS/EIS}	0

C. Ultrasound-Modulated Occluded Signals via Gyroscope Destabilization.

The gyroscope frequency response and the corresponding camera response, under signal injection and ambient conditions, is shown in Figure 8. As mentioned in Section V-B, either the device gyroscope or the camera movable lens destabilization can be used for an optical acoustic side channel, creating additional ultrasound targets for the adversary. This effect is also confirmed by comparing the noise metrics of frequencies outside of both the camera movable lens and device gyroscope resonant frequencies in Table V, which shows that the 25 kHz ultrasonic signal gyroscope response is the same as that in ambient conditions. The bias drift of a gyroscope is attributed to flicker noise, which contribute to random walk noise in the angle readings from a MEMS gyroscope [66], [67]. Ultrasound at the resonant frequency of a gyroscope results in greater bias instability, as seen in the OIS only condition in Table V, but the bias instability of ambient sound, OIS/EIS off, and 25 kHz frequency conditions are all similar, indicating that the source of the pixel intensity amplitude modulation is the movable lens structure of the camera. Unsurprisingly, the angle of random walk for the signal occluded condition without any image stabilization, None Signal Occluded in Table V, shows the most random walk drift due to signal modulation.

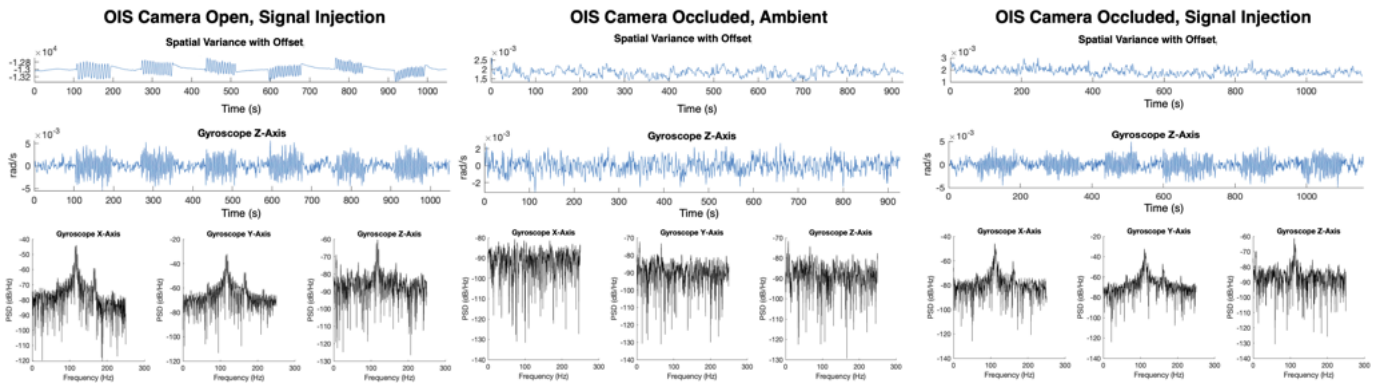

Fig. 8. **Gyroscope Frequency Response** The gyroscope frequency response confirms that the lens movement is resultant from the gyroscope destabilization from ultrasound at its resonant frequency. The gyroscope noise metrics in Table V are useful in identifying a set of channel functions for producing the strongest occluded camera responses. Simply using IMU (inertial measurement unit) data for device localization is insufficient when a low-resource method is desired, and can be challenging since it depends on walking measurements [26].

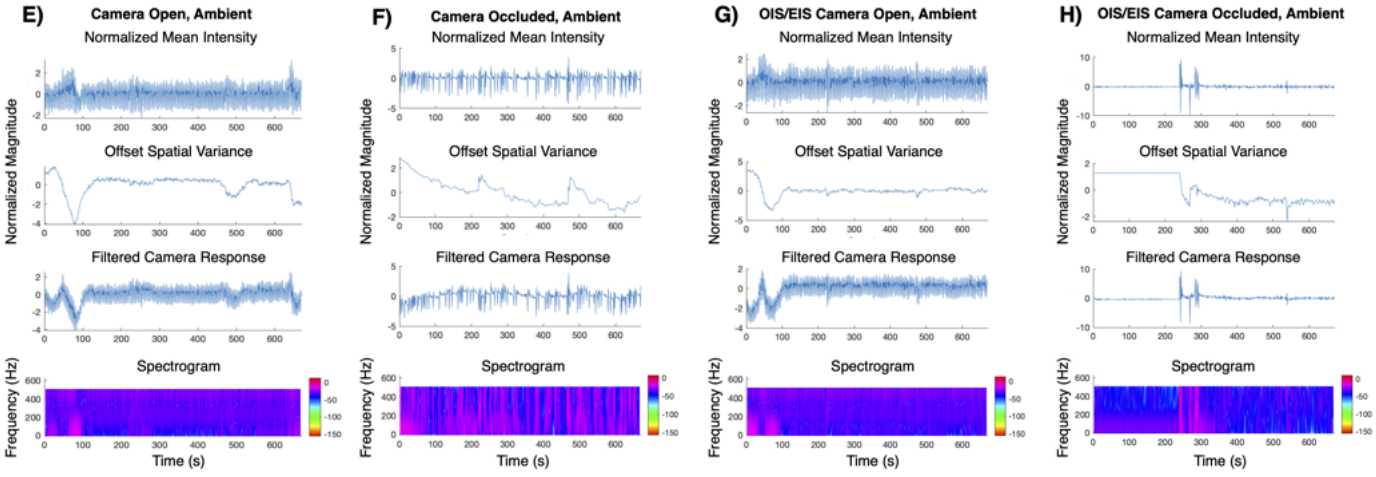

Fig. 9. **Ambient Camera Response.** The spectrograms for the ultrasound signal injections cases, as shown in Figure 3, have different frequency content relative to their ambient counterparts. The ambient occluded camera offset spatial variances responses are noisier relative to the ultrasound-modulated occluded responses. Camera responses shown here are from videos recorded at 10 ms exposure at 30 fps.

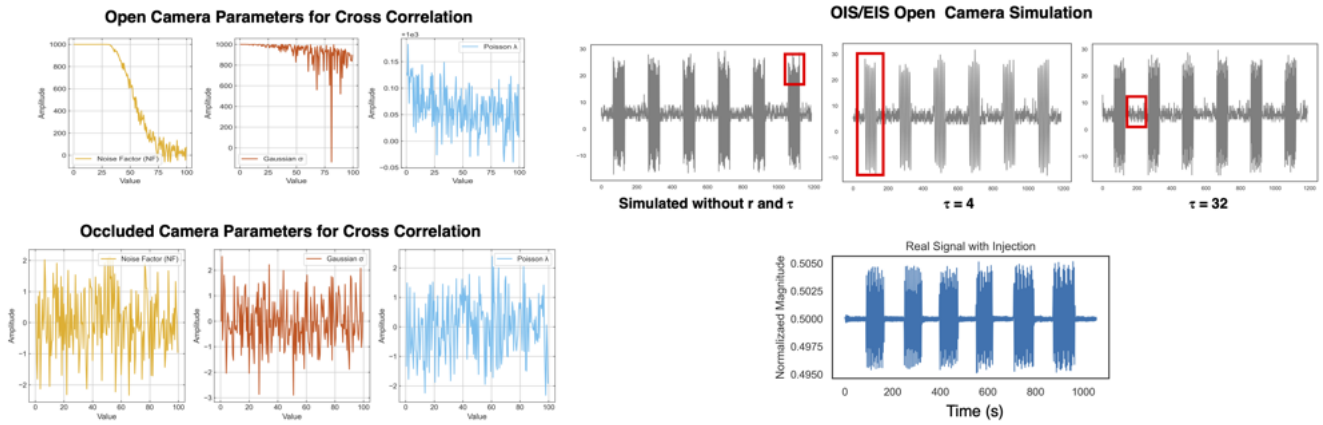

Fig. 10. **Simulation Correlation to Real-World Ultrasound Modulated Camera Responses.** Configuring the parameters of Equation 2 changes how the simulated signal correlates to the real-world ultrasound modulated signal, and strong positive correlation parameters can be found for both the Open Camera OIS/EIS and occluded Camera OIS cases. The Occluded Camera cross-correlation is not as high, and almost appears Gaussian, than the Open Camera simulation because of the way the noise parameters change the peaks of the signal, as seen in Figure 4, implying that there are additional sources of signal variance which may need to be added to the model. The simulation helps confirm our understanding of how the image sensor responds to ultrasound modulation, which opens possibilities mentioned in Section VII. The red boxes in the simulated signals (gray) show how each aspect of the signal can differ from the real-world (blue) camera response.

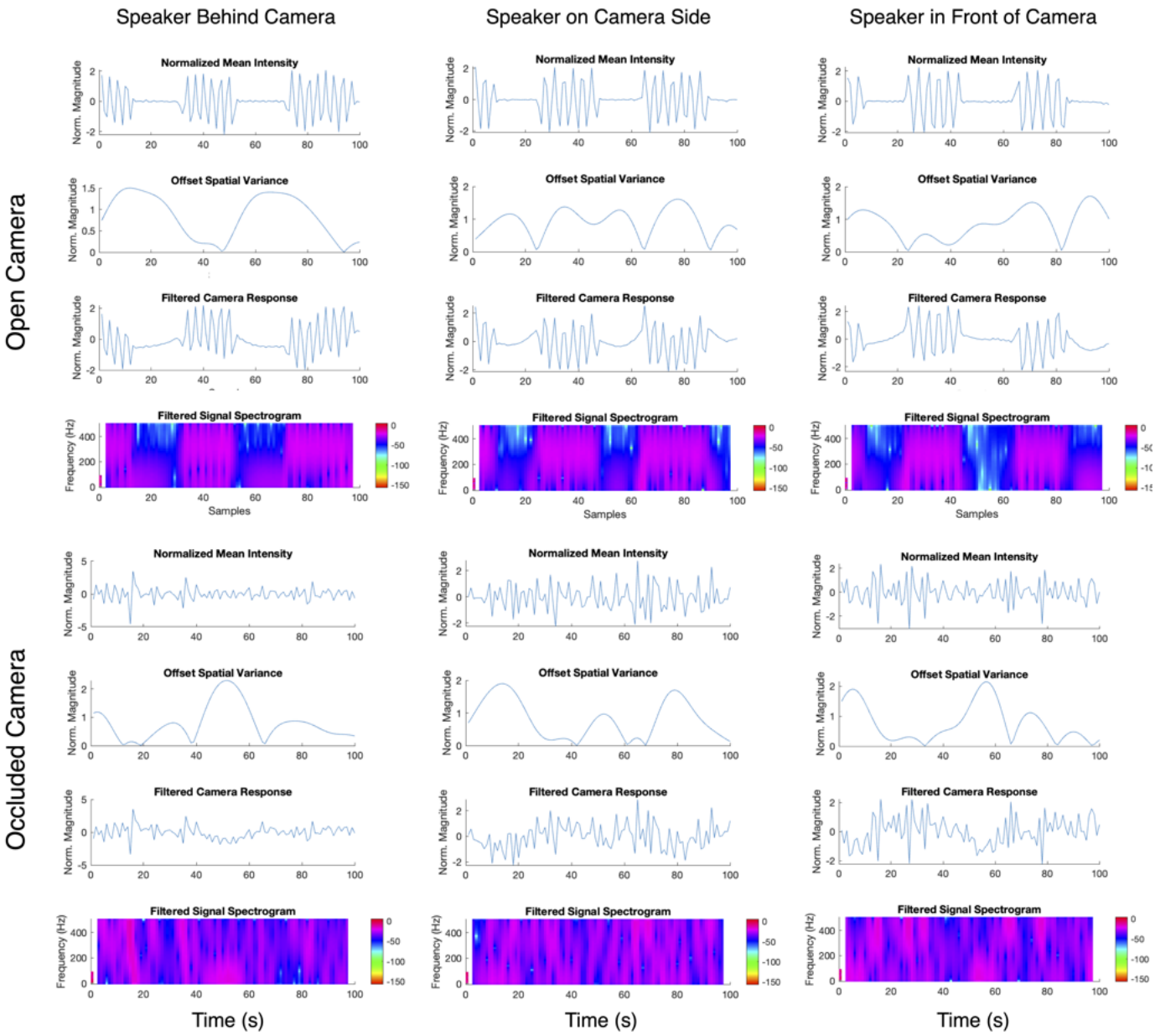

Fig. 11. **Ultrasound-Modulated Camera Responses for Different Speaker Directions.** Recorded at 1.5 secs pulse intervals, these relatively shorter camera responses demonstrate that an ultrasound-modulated signal does not depend on the direction of the speaker. This implies that the adversary does not need to rely on a specific speaker location to target and localize a device, as long as the device is in range. Additional discussion on device distances is mentioned in Section VIII.