Ultrasonic Microphone Jammers

Silas Wang

***Abstract*—With the rise of technology, audio surveillance and acoustic attacks have had a profound impact on the development of modern society. Most notably, with the rise of voice assistants and the growing privacy implications surrounding those products. This paper methodically analyzes the design of the ultrasonic microphone jammer, as well as metrics to evaluate the jammer's effectiveness.**

***Index Terms*—Audio surveillance, Acoustic attacks, Microphone jamming**

I. Background

The ultrasonic microphone jammer exploits microphone nonlinearities to create a wall of sound inaudible to humans but sensitive to microphone amplifiers, masking any sound that a microphone would record. Ultrasonic frequencies that interact with each other are sent into the microphone amplifier, creating a tone within the audible hearing range beneath the anti-aliasing filter of a DAC [1].

Note that the discussion on ultrasonic microphone jamming involves the jamming of MEMS microphones specifically. MEMS microphones semiconductor-based and assembled using manufacturing techniques that parallel those of integrated circuits. As a result of their size, they frequently appear in small, handheld devices, as well as IoT devices such as Smart Toasters and Amazon Alexa. Incidentally, it is through these microphones in which information can be covertly stolen, due to their frequency in tech-related products, as well as their incredibly small size.

A. *Microphone Amplifiers*

The ultrasonic microphone jammer hinges on two properties of a microphone amplifer to mask sound: harmonic distortion and intermodulation distortion. The former describes the tendency for a signal to introduce overtones, while the latter describes a system that generates the sums and differences of multiple signals that enter a system. A microphone amplifier can be generalized by the following Taylor Series:

$$S_{out} = \sum_{i=1}^{\infty} A_i (S)^i \tag{1}$$

Where A_i is a gain value that changes the phase and amplitude of the input frequencies and S is an arbitrary input sound. Since microphone amplifiers typically exhibit a linear frequency response (ie $S_{out} \approx A_1 S$) within the audible spectrum, a replica of an audible input sound can be recorded through an amplifier and played back with without distorting the original signal [1]. However, as the frequency of the input audio f_S increases beyond 20kHz, microphone amplifiers will behave like the summation described in Equation (1). More specifically, any $i > 2$ will have a relatively weak A_i weight component, implying the generalization for Equation (1) when $f \geq 20$ kHz can be described as the following linear and nonlinear component:

$$S_{out} = A_1 S + A_2 S^2 \tag{2}$$

Suppose $S = S_1 + S_2$, where the input sound S is the sum of two arbitrary sounds S_1 and S_2. We then have the following equation.

$$S_{out} = A_1 (S_1 + S_2) + A_2 (S_1 + S_2)^2 \tag{3}$$

If we let $S_1 = \cos\left(2\pi f_{S_1} t\right)$ and $S_2 = \cos\left(2\pi f_{S_2} t\right)$, Equation (3) now becomes

$$\begin{aligned} S_{out} = {} & A_1 \left[\cos\left(2\pi f_{S_1} t\right) + \cos\left(2\pi f_{S_2} t\right)\right] \\ & + A_2 \left[\cos\left(2\pi f_{S_1} t\right) + \cos\left(2\pi f_{S_2} t\right)\right]^2 \end{aligned} \tag{4}$$

Expanding the quadratic component will give us

$$\cos^2\left(2\pi f_{S_1} t\right) + 2\cos\left(2\pi f_{S_1}\right)\cos\left(2\pi f_{S_2} t\right) + \cos^2\left(2\pi f_{S_2} t\right)$$

which with the power reduction and product to sum trigonometry identities results in

$$\begin{aligned} & \left[\frac{1}{2} + \frac{\cos\left(2\pi 2 f_{S_1}\right) t}{2}\right] + \boldsymbol{\cos\left(2\pi f_{S_1} - f_{S_2}\right) t} \\ & + \cos\left(2\pi f_{S_1} + f_{S_2}\right) + \left[\frac{1}{2} + \frac{\cos\left(2\pi 2 f_{S_2}\right) t}{2}\right] \end{aligned} \tag{5}$$

Through the bracketed terms in Equation (5), we see the terms $2f_{S_1}$ and $2f_{S_2}$, the second harmonic of S_1 and S_2 respectively, thus showing the existence of harmonic distortion within a microphone amplifier. Additionally, notice that through the product to sum identity, we derive $\cos\left(2\pi f_{S_1} - 2\pi f_{S_2} t\right)$ (bolded term) and $\cos\left(2\pi f_{S_1} + 2\pi f_{S_2} t\right)$, proving the existence of intermodulation distortion. If we let $f_{S_1} = 50\text{kHz}$ and $f_{S_2} = 40\text{kHz}$, we will see that $\cos(2\pi(50\text{kHz} - 40\text{kHz})t) = \cos(2\pi(10\text{kHz})t)$ will appear from Equation (3).

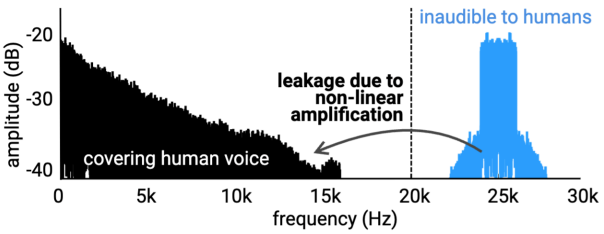

Fig. 1. A diagram demonstrating intermodulation distortion with ultrasonic signals. Source: [2]

Note that in Equation (4), the only term that will pass through the Nyquist filter will be $\cos(2\pi 10\text{kHz})$ (the bolded

term), effectively allowing for the creation of audible tones using inaudible sound waves and microphone nonlinearities.

Unlike studio-grade microphones, MEMS microphones typically contain an Automatic Gain Control (AGC) circuit, as their usage in consumer-grade electronics demand effortless control of a wide dynamic range. This property can also be taken advantage of with the ultrasonic microphone jammer, as a high power audible intermodulated signal that enters the amplifier will subsequently lower the rest of the input signal. If the signal-to-jam is substantially quieter than the high power intermodulated signal, the signal-to-jam will also be reduced significantly to compensate for the input signal [1].

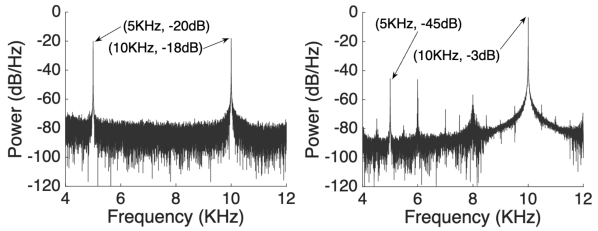

Fig. 2. Two graphs demonstrating the concept of Automatic Gain Control (AGC). The left graph shows a 5kHz at −20dB and a 10kHz tone at −18dB. The right graph shows the same 5kHz tone at −45dB as a result of the 10kHz tone increasing to −3dB. Source: [1]

B. *Jamming Signals*

Given $S_{out} = A_1(S_1 + S_2) + A_2(S_1 + S_2)^2$, we seek to find S_1 and S_2 such that S_{out} effectively masks the audible spectrum. If $S_1 = \cos\left(2\pi f_{S_1} t\right)$ where $f_{S_1} \geq 25\text{kHz}$, S_2 must be a signal that masks a large bandwidth of the audible spectrum through intermodulation distortion when interacting with S_1. We outline three different techniques in which a jamming signal can be created to cast a "Shadow Signal" [1], [3].

a) ***Static Tones:*** A static tone signal assumes that S_2 is another periodic tone with a fixed frequency f_{S_2}. As described in Equation (5), the interaction of two static tones S_1 and S_2 creates an audible shadow signal with frequency $f_{S_1} - f_{S_2}$. Note that S_2 can also be written as the sum of multiple sinusoidal signals, which can be described as

$$S_2 = \sum_{i=1}^{j} \cos\left(2\pi f_{S_1} + f_i t\right) \tag{6}$$

where f_i describes the ith bias frequency to be summed. While this type of signal is the simplest to conceptualize, the resulting complexity of the jamming signal–and by extension the signal-to-noise ratio–is not adequate to mask the sounds recorded by a microphone [3].

b) ***Frequency Modulated Tones:*** >An ultrasonic sinusoidal singal S_1 can be frequency modulated by a lower frequency $f_{S_2} < f_{S_1}$ to ensure that the audible $S_1 - S_2$ term is dynamically changing. If we assume the carrier to be S_1 and S_2 to be the modulator, we have an adjustable bandwidth to adequately block a significant portion of the audible hearing spectrum through the difference between S_1 and S_2. Some examples of S_2 incloude a sawtooth wave, a sine wave, and a sample-and-hold based waveform. The following equations describe this mathematically:

$$S_{SAW} = \cos\left(2\pi\left(f_{S_{SAW}} + (kt \bmod m)\frac{f_n}{m-1}t\right)\right) \tag{7}$$

where $(kt \bmod m)$ is a linear ramp with slope k from 0 to $m-1$, which gets converted to a frequency scale through $\frac{f_n}{m-1}$ [3].

$$S_{S\&H} = \cos(2\pi(f_{S\&H} + \text{rand}(N)t)) \tag{8}$$

where $(f_{S\&H})$ is the sample and hold rate and $\text{rand}(N)$ is a random frequency bias value [3].

$$S_{SINE} = \cos(2\pi f_{SINE} + \beta\cos(2\pi f_{FM})) \tag{9}$$

where f_{FM} is the frequency of the sine wave that is modulating the S_1 tone and β is the amount of modulation applied to S_1 [1].

c) ***White Gaussian Noise:*** An ultrasonic signal S_2 can be considered to be white noise, in which the sound energy is evenly distributed across the ultrasonic spectrum. The resulting sound generated by $S_1 - S_2$ will also reflect the characteristics of white noise within the audible spectrum. This can be described with the equation

$$S_2 = wgn\left(f_{S_1}, f_{S_1} + f_b\right) \tag{10}$$

where f_b is the bandwidth of noise [3].

II. Ultrasonic Microphone Jammer Design

This section brieflty goes over a block-diagram structure of an ultrasonic microphone jammer and some consequences that mnay arise with the system that this paper has been outlining.

The implementation of an ultrasonic microphone jammer prototype does not involve many unique components. In principle, the microphone jammer only consists of a power supply, signal generator, audio amplifier, and an array of piezoelectric ultrasonic transducers. Another common addition to the devices mentioned is a microcontroller, which enables for the development of an embedded system. The figures below showcases an example of a simple ultrasonic microphone jammer:

Variations to this design aim to address some design problems that arise from the ultrasonic microphone jammer. We discuss the properties of directionality, selectivity, noise reduction, and the ringing effect.

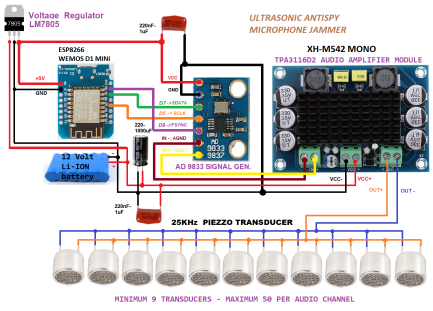

Fig. 3. A microcontroller-based ultrasonic microphone jammer using an ESP8266 microcontroller, an AD9833 signal generator, a TPA3116D2 audio amplifier, and an array of ultrasonic transducers. Source: [4]

A. *Directionality*

The wavelength of a sound wave shares a relationship with its behavior as it travels through a medium. For instance, lower frequencies tend to not reflect off surfaces due to their size, while higher frequencies tend to be more directional and behave like rays of light. Because the wavelength of an ultrasonic frequency is incredibly small, the ultrasonic microphone jammer becomes incredibly directional if the transducer array is arranged on a 2D plane. This design is quite impractical, as it would require for the ultrasonic jammer to be pointed directly at the microphones that must be jammed; making it ineffective against hidden microphones, or if the microphones-to-protect is not stationary.

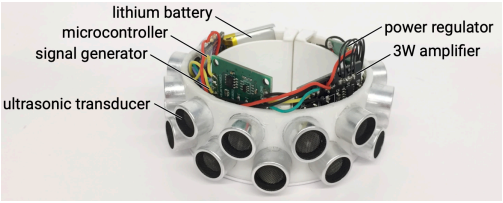

Fig. 4. A wearable microphone jammer featuring ultrasonic transducers and a bracelet design. Source: [2]

A promising solution to this was designed by *Chen et. al* in 2020 through the use of a wearable bracelet with a diameter of 9cm, 23 ultrasonic transducers attached to the outer shell, and various components (signal generator, amplifier, microprocessor, lithium ion battery) within the ring's inner enclosure. Because this design utilizes a ring-like shape, the device is wearable for humans and does not need to be repositioned. Additionally, the blind spots within a typical transducer-based jammer will be mitigated through the movements that the human speaker will make with their hands as they move around. But because this is a wearable devic, the size of the device will impact part selection and assembly. The following figures showcase select measurements and simulations of their device [2].

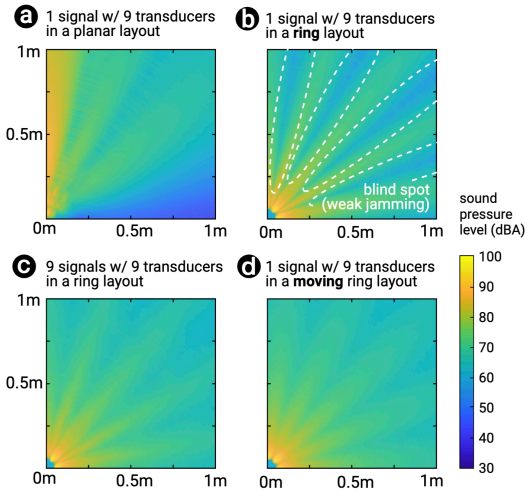

Fig. 5. A simulation of 9 out-of-phase transducers in various configurations, signal amounts, and rotational movements. Source: [2], [5]

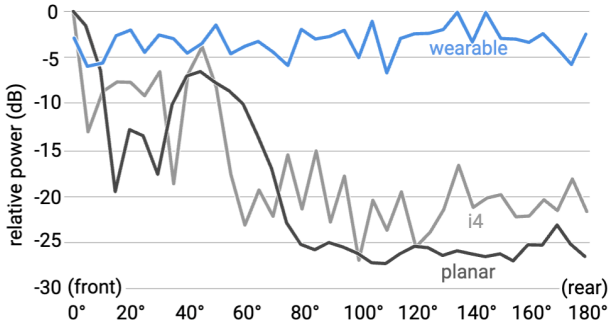

Fig. 6. A measurement of the wearable microphone jammer's power relative to the maximum power of each transducer 1 meter away from the device. Source: [2]

Similar research has also been conducted by *Shen et. al* with their *JamSys* paper, where they construct an algorithm similar to the harmony search algorithm to minimize blindspots given an area to protect with the jammer and a number of transducers [6].

B. *Selectivity*

Assuming the usage of an effective microphone jammer, we discuss the issue of selectivity. That is, the ability for an ultrasonic microphone jammer to be able to discriminate its jamming abilites on a device-by-device basis. For instance, if a microphone jammer was placed next to a person that is currently on a phone call, the jammer would be able to block a spy covertly recording the conversation, but also the phone on call. A selective jammer would allow authorized devices within a private network to record high quality audio, but prevent an unauthorized adversary who is covertly recording the conversation.

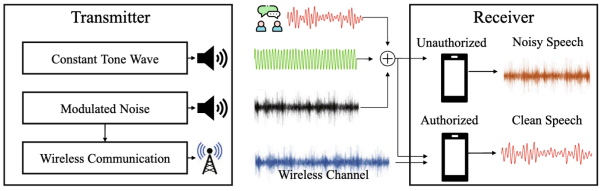

Fig. 7. A block diagram of a selective ultrasonic microphone jammer design proposed by *Chen et. al (2021)*.

A design described by *Chen et. al (2021)* utilizes a wireless network and a gyroscope-based receiver. The noise generated by the jammer is transmitted through a multi-channel wireless communications system that utilizes WiFi, Bluetooth, and a speaker-gyroscope noise transmission system based off of a design proposed by *Gao et. al (2020)*. Since most receiver smartphones are equipped with gyroscopes, the gyroscope-based communications can be used without any extra hardware additions. The noise will be received through the gyroscope as it detects the changes in sound vibrations. Software on authorized devices would then be able to decode the noise using the Normalized Least Mean

Squares algorithm and receive the information transmitted through the gyroscope, as well as switch between each communication channel to prevent information from getting sniffed [7], [8].

C. *Ringing Effect*

Select research papers have reported a particular "ringing" effect with certain models of consumer-grade piezoelectric ultrasonic transistors. Certain frequencies resonate within the body of the transducer, creating an audible ringing effect that decays for longer than the other parts of the intermodulated signal; a nonlinear side effect that compromises the inaudibility of a microphone jammer [1]. As a result of this, we must filter the outgoing ultrasonic signal before it reaches the microphone amplifier to ensure that no audible ringing effect exists, mitigating the nonlinearities of the transdutcer.

We describe this nonlinearity through representing the transducer through a generic impulse responsoe and convolving the signal from the microphone jammer with the impulse response. Suppose h is an impulse response representing the linear time-invariant system of the ultrasonic transducer. h is then

$$h = k_0\delta(t) + k_1\delta(t-1) \tag{11}$$

where k_i represents the amount of signal at a given moment in time i for any $i \in \mathbb{Z}$. We obtain the ultrasonic transmitter signal S_{UT} by creating a jamming signal using the frequency modulation technique described in Section I.B.b.

If we let $S_{UT} = \cos(2\pi f_c t + \beta\cos(2\pi f_m t))$ be the frequency modulated jamming signal where f_c is the carrier frequency and f_m is the modulator frequency, the resulting behavior as the jamming signl S_{UT} enters the ultrasonic transducer h can then be described as the convolution of $S_{UT} * h$ where

$$\begin{aligned} S_{UT} * h &= \cos(2\pi f_c t + \beta\cos(2\pi f_m t)) \\ &\quad * k_0\delta(t) + k_1\delta(t-1) \\ &= k_0[\cos(2\pi f_c t + \beta\cos(2\pi f_m t))] \\ &+k_1[\cos(2\pi f_c(t-1) + \beta\cos(2\pi f_m(t-1)))] \end{aligned} \tag{12}$$

It is given that the behavior of microphones are very similar to speakers–and by extension transducers. This means that the Taylor Series that was used to approximate a microphone amplifier in Section I.A can also be used to approximate the nonlinear behavior of a transducer. Since we know that any $f_c > 20$kHz, $S_c * h$ is inaudible as it travels through the microphone amplifier. But since the output of a microphone amplifier also contains a nonlinear component, meaning $S_{out} = A_1(S_c * h) + A_2(S_c * h)^2$, through intermodulation distortion, we have the following:

$$\begin{aligned} (S_c * h)^2 \Longrightarrow k_0k_1\cos([f_c t + \beta\cos(f_m t)] \\ -[f_c(t-1) + \beta\cos(f_m(t-1))]) + \text{other terms} \end{aligned} \tag{13}$$

Implying that through the very same nonlinear principle that ultrasonic microphone jammers used to create shadow signals, the behavior of the transducer's output signal also creates a shadow tone, creating an audible ringing sound as it travels through the air.

To filter this ringing tone from the jamming signal S_{UT}, we remove the ringing tone before the signal reaces the transducer. Through covolution, this can be described as $S_{UTm} = h^{-1} * [S_{UT} * h]$, where S_{UT} is the ultrasonic jamming signal and $S_{UT} * h$ represents the signal with the nonlinearities of the transducer.

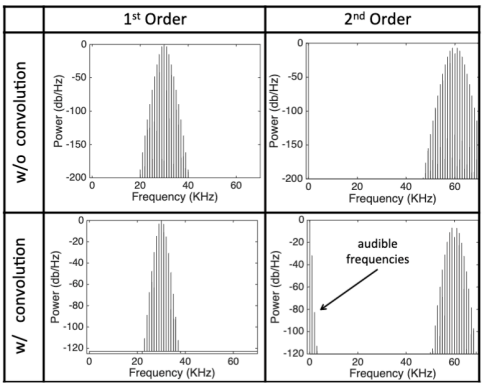

Fig. 8. A visualization of the nonlinearities of the ultrasonic transducer, creating additional audible frequencies before it enters the microphone amplifier. Source: [1]

By forcing the transmitter to output $S_{UT} * h^{-1}$, we can filter out the ringing tone before it reaches the microphone amplifier. This can be computed by calculating the k_i coefficients for resonating frequencies. Because it can be guaranteed that the transducer's k values will not change over time, these k values can be precalculated and stored, reducing the amount of computation for the implementation of the inverse filter [1].

III. Evaluation

This section discusses the effectiveness of the ultrasonic microphone jammer as a whole. That is, the ability for an inaudible wall of sound to effectively mask sonic information that would enter a microphone. We introduce this section by discussing select security implications, and then provide an elaborate framework developed by *Gao et. al (2024)* on a potential system to accurately measure the efficacy of ultrasonic microphone jammer designs.

A. *Security*

The entire premise of an ultrasonic microphone jammer is to prevent an adversary from stealing information. Prior research has used multiple metrics to quantify this process, such as comparing the amount of words that a speech-to-text algorithm recognizes when fed raw audio or jammed audio, or asking volunteers to quantify speech clarity of a jammed signal compared to a signal applied with noise reduction algorithms. The failure of these measurements lie within a few overlooked properties by which humans can evaluate information [3].

To begin, there are three different factors that affect the potency of an ultrasonic microphone jammer: the device itself, the competency of the eavesdropper in separating noise from a signal and interpreting the signal, and the

amount of time it takes said adversary to unveil private information [3].

While many ultrasonic microphones focus on metrics such as word recognition, it is also important to recognize that many other types of information can be extracted from a recording strictly from the ambience. For instance, a bug next to a printer may imply the victim is in an office, and depending on the amount of audio the adversary has, the adversary may be able to formulate the victim's daily routine, the number of people around the bug, their identities, etc.

Additionally, proof-of-concepts detailing a machine listening system that correctly identifies 50-60% of handwritten words exists [9]. Similar research has also been conducted with keylogging, the contents of a paper as it is being printed, and keystrokes [2], [10].

Assuming the scenario in which an effective ultrasonic microphone jammer is in use, there are a few different methods of noise reduction methods that can be applied to extract information from a jammed signal.

a) ***Blind Source Separation:*** Known as the "Cocktail Party Algorithm" of sound isolation, Blind Source Separation is a potent algorithm for de-noising a sound without having prior knowledge of the contents within the audio. We assume a vector of observed mixed signals X to be the product of all actual independent sounds A and the amount of each independent sound W. To find a vector $\hat{S}$ such that $\hat{S} = XW$, we perform Independent Component Analysis (ICA) to obtain an approximation of W^{-1}. Since the jamming signals that were discussed in Section I.B involved creating a jamming signal that is independent to the signal to mask, BSS is incredibly effective in removing the noise from a jammed signal. Note that for this algorithm to work, the number of observed signals X must be greater than or equal to the number of actual signals A [3].

b) ***Filters & Spectrum Analysis:*** Filters such as the Low Pass, Band Pass, Band Stop, and High Pass Filters can be used to eliminate any noisy characteristics of the signal, provided that the noisy regions of the signal can be identified. Through the usage of spectrum analysis algorithms such as the Short Time Fourier Transform (STFT) or the Discrete Wavelet Transformation (DWT), a spectrogram can be created to further analyze the specific regions that should be filtered. The former, performs the Fourier transform over a fixed window of time to produce a 2D spectrogram, while the latter separates the signal into multiple frequency regions and quantifying the amount of signal in each frequency region. Algorithms like Adaptive Notch Filering allow for a precise way to de-noise a signal [3].

c) ***Localization:*** Research using ultrasonic sound for cybersecurity points has shown the ability to locatae where the ultrasound was being played from by studying the attenuation rate of ultrasonic waves [11]. Another technique known as beamforming utilizes a parabolic array of phase-misaligned microphones, and delay compensating each out of phase microphone to locate the angle of arrival of the sound source [3]. The combination of these two techniques can provide allow an adversary to approximate the given location of a soutnd source without any prior knowledge.

d) ***Coupling Effect:*** Based on the system that was described for sound generation in earlier sections (Section I.A and Section I.B), the creation of audible shadow signals using ultrasonic frequencies requires a static S_1 tone by which the audible coefficient $S_1 - S_2$ can be consistently derived, allowing for the control of the bandwidth of noise and the center frequency through the use of jamming signal techniques. A byproduct of this is that if the adversary is knowledgable about the system design of an ultrasonic microphone jammer, through spectrum analysis, the adversary can locate the central frequency and algorithmically determine the bandwidth of noise that the microphone jammer is broadcasting, even if multiple ultrasonic tones are used to create the jamming signal (Equation (6)).

B. Addressing Noise Removal

To reduce an adversary's ability to recover information from a jammed signal, the complexity of the jammed signal must be altered such that the jamming signal and the signal-to-mask must share spectral characteristics, or have some spectral coherence.

a) ***Multiband Jamming Signal*** *:* Recall Section I.B.b, where the idea of creating a large bandwidth of noise through modulating a static ultrasonic tone S_1 with a lower frequency S_2 was discussed. The idea of a multiband jamming signal involves partitioning the ultrasonic domain into n smaller bands where each band contains an ultrasonic tone S_{1_n} that gets frequency modulated with a S_{2n}. As a result of this, number of shadow signals in the audible spectrum drastically increase alongside the jamming signal's bandwidth, leading to more intervals where the two signals are spectrally coherent, and fewer moments in time where a guard tone can be discovered [3].

For example, if we let $n = 3$ for instance, the number of intermodulation signals we can directly derive from three frequency modulated signals S_{FM1}, S_{FM2}, and S_{FM3} is $\binom{3}{2} = 3$, with $S_{FM1} - S_{FM2}$, $S_{FM1} - S_{FM3}$, and $S_{FM2} - S_{FM3}$. Higher-order intermodulation distortion terms such as $2S_{FM1} - S_{FM2}$ will also appear as a result of deriving higher-order coefficients of the Taylor Series introduced in Section I.A.

b) ***Active Noise Cancellation*** *:* An idea proposed by *Gao et. al (2024)* involves utilizing two additional microphones and a signal processor module to actively cancel out the signal-to-protect that travels through the air by sending the inverse of the signal-to-protect into the ultrasonic domain. The first microphone is permitted to record by the ultrasonic noise jammer's network, and records the audio-to-protect. This signal gets inverted and modulated into the ultrasonic domain, while the second one records the ultrasonic signal being sent from the ultrasonic transducer, and dynamically adjusts the inverse signal generation based on the past error rate [3].

Combined with a multiband jamming signal method to generate coherent sound, they claim that this system can successfully defend against virtually any method of noise

cancellation, as the ultrasonic jamming signal will also contain the inverse of the signal-to-protect, effectively canceling any of the sound being protected in the room [3].

C. *Framework for Evaluation*

Experiments with ultrasonic microphones to evaluat their designs with metrics focused on word recognition and audio quality. Some other metrics also include intelligibility after noise reduction, and word recognition using speech recognition systems like Google Now. While researchers were able to show that an ultrasonic microphone jammer system can retain low word recognition rates upon assessing jamming signal quality, the metrics presented are not representative of a real scenario where an adversary is attempting to recover information from a jammed signal.

For instance, an eavesdropper may employ a plethora of tactics (BSS, Spectrum Analysis, Beamforming, etc.) to remove noise from the signal. But even if the adversary doesn't successfully recover the clean source audio, they can semantically deduce what certain words mean and various information from the ambient noise alone. Additionally, no metric exists documenting the effectiveness of multiple humans and speech recognition algorithms working in tandem to uncover words in a noisy signal [2], [3], [7].

Researchers from *Gao et. al (2024)* proposed a detailed three-pronged perspective on evaluating the effectiveness of microphone jammers, through intensity, semantic comprehension, and collaborative recognition. Their performance metric is thus described as

$$S_{total} = \boldsymbol{W} \cdot [S_{\text{SNR}},\ S_{\text{In}},\ S_{\text{CRR}}]^T$$

where $\boldsymbol{W}$ is a three-dimensional weight vector, S_{SNR} is the intensity score, S_{In} is the intelligibility score, and S_{CRR} is the collaborative recognition score. The weights were determined by experimental surveys where volunteers were asked to rank audio surveillance scenarios on a scale of 1 (no threat) to 7 (high threat) as well as their attitudes on data leaks through ultrasonic microphone jammers. Through this method, *Gao et. al* determined that the default weight vector is $\boldsymbol{W} = [0.3337\ 0.3609\ 0.3053]$ [3].

Note that this paper only introduces the framework provided by *Gao et. al (2024)* as a metric for evaluating the effectiveness of an ultrasonic microphone jammer. The experimental methods by which the data can be measured is out of the scope of this paper.

a) ***Intensity:*** The intensity metric S_{SNR} evaluates the effectiveness of the microphone jammer's ability to conceal ambient information. It can be determined by the product of two three-dimensional vectors, where

$$S_{SNR} = \boldsymbol{w_{SNR}} \cdot \left[\text{SNR}\ \text{SNR}_{seg}\ \text{fwSNR}\right]^T$$

where $\boldsymbol{w_{SNR}}$ is a three-dimensional weight vector which through a series of experimental trials is $\boldsymbol{w_{SNR}} = [\mathbf{0.34\ 0.22\ 0.44}]$ [3].

The primary metrics for S_{SNR} utilize three different ways of measuring the signal-to-noise ratio of a given signal. SNR for instance, is widely used to measure the amount of noise in a signal, while SNR_{seg} and fwSNR account for the signal to noise ratio over extended periods of time and across frequency bands. The final metric S_{SNR} is a normalized value on the interval [0, 1], where larger values imply better ambient noise protection [3].

b) ***Intelligibility:*** The intelligibility metric S_{In} is a measurement of human comprehension and intelligibility. It is a normalized value on the interval [0, 1] where larger values indicate better signal intelligibility and is determined by the product of the following:

$$S_{\text{In}} = \boldsymbol{w_{\text{In}}} \cdot [-\text{WSS}\ -\text{LLR}\ \text{NCM}\ \text{CSII}\ \text{PESQ}\ \text{STOI}]^T$$

where through the analysis of speech intelligibility, the default values of $\boldsymbol{w_{\text{In}}}$ is $[0.06\ 0.13\ 0.20\ 0.21\ 0.18\ 0.22]$. The final metric S_{In} This is achieved through measuring propagation distortion, additive distortion, and algorithmic artifacts, each using two metrics [3].

The propagation distortion component measures the change in a signal as it travels through a medium. WSS, or the Weighted Spectral Slope describes a change in the spectral envelope (the amplitude and frequency of a signal at a point in time). A large change in WSS indicates that the signal has been significantly altered. The Log-

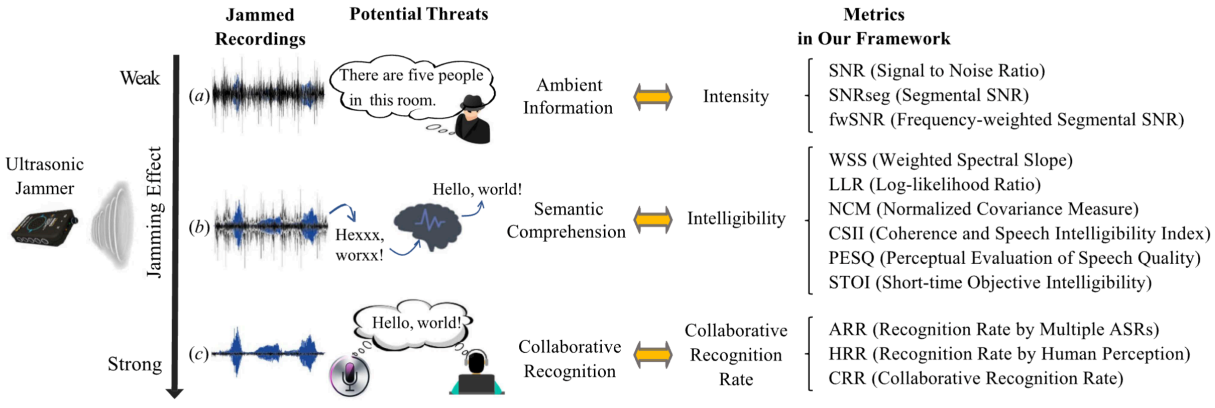

Fig. 9. An overview of the evaluation framework proposed by *Gao et. al (2024)* that measures the amount of ambient information, intelligibility, and word recognition from humans and speech algorithms. Source: [3]

FIVE REPRESENTATIVE UMJS AND THEIR OVERALL PERFORMANCES

Representative UMJ	Produced Noise Category	Without Noise Elimination				With Noise Elimination			
		S_{total}	S_{SNR}	S_{In}	S_{CRR}	S_{total}	S_{SNR}	S_{In}	S_{CRR}
Wearable Jammer [5]	[0,1] kHz DSFN hopping per 0.45 ms	0.86	0.84	0.76	1.00	0.41	0.41	0.49	0.32
Patronus [6]	[85,255] Hz DSFN hopping per 0.2 s	0.83	0.82	0.74	0.93	0.28	0.32	0.47	0.02
MicShiled [7]	4 kHz bandwidth WGN	0.87	0.82	0.81	1.00	0.37	0.31	0.47	0.33
Backdoor [8]	8 kHz bandwidth WGN	0.88	0.85	0.79	1.00	0.48	0.46	0.58	0.41
Selective Jammer [9]	three 4 kHz WGN covering a 12 kHz band	0.93	0.90	0.89	1.00	0.56	0.48	0.58	0.61

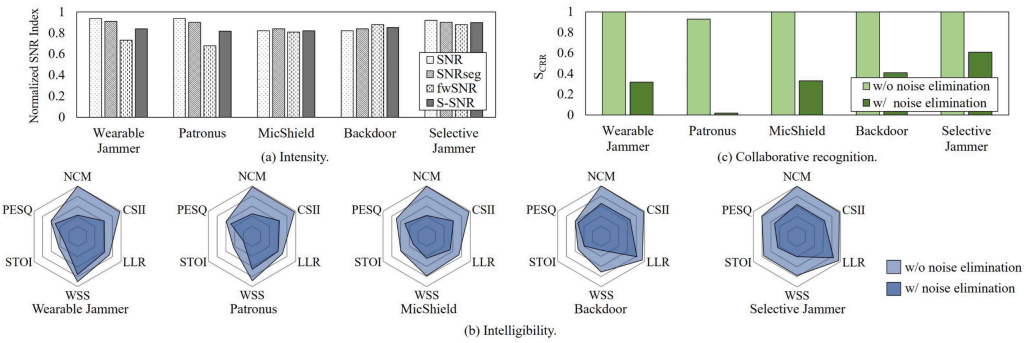

Fig. 10. Performance of five representative UMIs and their vulnerability to adversarial noise elimination. Note that DSFN is equivalent to the frequency modulated jamming signal discussed at Section I.B.b and WGN is equivalent to white noise. Source: [3]

Likelihood Ratio LRR on the other hand, is the logarithm of the clean sound's amplitude with the distorted sound's amplitude, summed across an interval of time. Larger LRR values indicate low intelligibility. Because these values are negatively correlated, the opposites are taken [3].

The additive noise distortion component measures the statistical characteristics of the clean and distorted signal in the frequency domain. NCM, or the Normalized Covariance Measure involves computing the linear relationship of the clean and distorted signal for each frequency band. This value gets normalized on the interval [−1, 1], which then is taken into a weighted sum based on the frequency bands that are most important for intelligibility [3].

The Coherence and Speech Intelligibility Index (CSII) is a weighted average of the clean signal's power (the phase and amplitude of a signal at a point in time) and distorted signal's power. Just like NCM, the weighted average is determined by the importance of each frequency band to intelligibility.

Algorithmic artifact distortion is the impact that external noise, reverberation, and digital audio artifacts have on a given signal. Perceptual Evaluation of Speech Quality (PESQ) is more tailored to digital artifacts and is taken over longer windows of time, while the Short Term Objective Intelligibility (STOI) measure focuses on the shorter acoustic cues that make it effective at quantifying the impact of reverb and room noise on intelligibility. Higher PSEQ and STOI implies better intelligibility [3].

c) ***Collaborative Recognition***: The Collaborative Recognition metric S_{CRR} is a measurement of the amount of words that a human can recover from a jammed signal using speech recognition algorithms and humans. This metric is described as the the following sum:

$$\text{CRR} = \text{HRR} + \text{ARR}$$

where HRR represents the percentage of words recognized by all human volunteers, and ARR represents the percentage of all words recognized by all speech recognition algorithms that were used. This can be rewritten as

$$\text{CRR} = \frac{\text{human recognized words}}{\text{all words}} + \frac{\text{algorithm recognized words}}{\text{all words}}$$

By nature, this sum is negatively correlated to the effectiveness of a ymicrophone jammer, as larger values imply that more words are recognized and the signal is not effectively jammed. Since the other metrics are also positively correlated to the jammer's effectiveness, the opposite is taken. This means the normalized final metric on the interval [0, 1] is $S_{\text{CRR}} = (1 - \text{CRR})$, where higher values imply less words are recognized by humans and speech recognition algorithms [3].

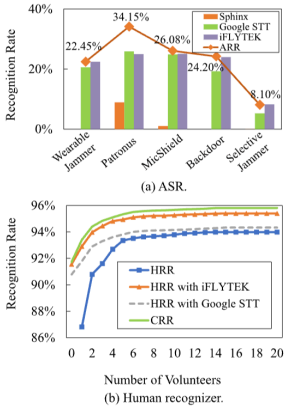

Fig. 11. The ARR and CRR of five ultrasonic microphone jammer designs assessed by 3 speech recognition algorithms and 20 human volunteers. Source: [3]

D. *Other Considerations*

The following sections documents topics relating to the ultrasonic microphone jammer that were not discussed in detail within this paper, as well as research related to the ultrasound-based system that was described in this paper.

a) ***Jammer and Microphone Placement*** *:* Ultrasonic microphone jammers using planar transducer arrays are only capable of jamming sounds in a small cone directly in front of the transducer array. This can become incredibly problematic for hidden microphones behind some kind of material, even if the jammer is effective. Additionally, most experiments with an ultrasonic microphone jammer involve testing the microphone at a very short distance ($d < 1$m), making the scalability of this system a point of discussion. *Chen et. al (2020)* measures this using a word error rate metric measuring the number of incorrectly recognized words. Topics relating to ultrasonic microphone jammer placement, blind spots, and evaluation of hidden microphones have not been meticulously explored in this paper [2].

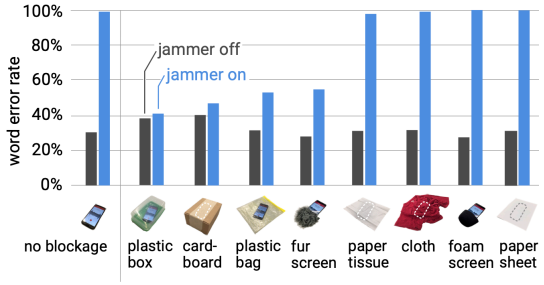

Fig. 12. A measurement of incorrectly recognized words recorded by MEMS microphones behind certain materials using the wearable mcrophone jammer design by *Chen et. al (2020)*

b) ***Loud Ultrasonic Sounds*** *:* According to the U.S. Occupational Safety and Health Administration (OSHA), ultrasonic subharmonics–such as the ones created by IMD–are dangerous at above 105 dB_{SPL} [2]. Additionally, Havana Syndrome, an illness featuring side effects including severe headaches, dizziness, blurred vision, tinnitus, has been reported and disputed by multiple organizations including the Department of Defense, Central Intelligence Agency, and the National Institute of Health. A theory of cause includes concentrated beams of energy, such as those generated by ultrasonic transducer arrays. While it is possible that ultrasonic microphone jammers induce health-related side effects through its highly concentrated beams of sound energy, no research directly confirms this [12]. Inevitably, the intensity of the ultrasonic signal and its relationship with the ultrasonic microphone jammer has not been discussed at length.

c) ***DolphinAttack****:* Another way of using ultrasonic sounds is to inaudibly send voice commands to voice assistants such as Amazon Alexa or Siri. *Zhang et. al (2017)* has shown that it is possible to control voice assistants by modulating a text-to-speech voice command into the ultrasonic domain. Through intermodulation distortion within the microphones of the voice assistants, the voice command is then recognized. Note that the recognition rate is also based on the command type, as keywords such as "weather" and "airplane" have a higher success rate, whereas keywords such as "call" and "open" require other words for the commands to be successfully recognized.

IV. Conclusion

The ultrasonic microphone jammer is a device that cleverly utilizes microphone nonlinearities to mask sound. Through its various design features and proposed workarounds to the properties it exploits, the ultrasonic microphone jammer is a unique option for consunmers to address the growing concern of cybersecurity and audio surveillance. While there is research that exists that show the effectiveness of the device, there are also major design implications that could effectively nullify the effectiveness of an ultrasonic microphone jammer.

References

[1] N. Roy, H. Hassanieh, and R. R. Choudhury, "Backdoor: Making Microphones Hear Inaudible Sounds," in *Proceedings of the 15th Annual International Conference on Mobile Systemds, Applications, and Services*, 2017, pp. 2–14. doi: 10.1145/3081333.30813.

[2] C. Yuxin *et al.*, "Wearable Microphone Jamming," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, in CHI '20. Honolulu, HI, USA: ACM, 2020, pp. 1–11. doi: 10.1145/3313831.3376304.

[3] M. Gao *et al.*, "A Resilience Evaluation Framework on Ultrasonic Microphone Jammers," *IEEE Transactions on Mobile Computing*, vol. 23, no. 2, pp. 1914–1929, 2024, doi: 10.1109/TMC.2023.3244581.

[4] A. Loboda, "Antispy Jammer." Accessed: Aug. 20, 2025. [Online]. Available: https://github.com/mcore1976/antispy-jammer

[5] "Wearable Microphone Jamming." Accessed: Aug. 20, 2025. [Online]. Available: https://github.com/y-x-c/wearable-microphone-jamming

[6] H. Shen, W. Zhang, H. Fang, Z. Ma, and N. Yu, "JamSys: Coverage Optimization of a Microphone Jamming System Based on Ultrasounds," *IEEE Access*, vol. 7, pp. 67483–67496, 2019, doi: 10.1109/ACCESS.2019.2918261.

[7] Y. Chen, M. Gao, Y. Liu, and Y. Zheng, "Implement of a secure selective ultrasonic microphone jammer," *CCF Transactions on Pervasive Computing and Interaction*, vol. 3, no. 4, pp. 367–377, 2021, doi: 10.1007/s42486-021-00074-2.

[8] M. Gao *et al.*, "Deaf-Aid: Mobile IoT Communication Exploiting Stealthy Speaker-to-Gyroscope Channel," in *The 26th Annual International Conference on Mobile Computing and Networking (MobiCom '20)*, in MobiCom '20. London, United Kingdom: Association for Computing Machinery, 2020, p. 13. doi: 10.1145/3372224.3419210.

[9] T. Yu, H. Jin, and K. Nahrstedt, "Audio based Eavesdropping of Handwriting via Mobile Devices," pp. 444–455, 2016, doi: 10.1145/2971648.2971765.

[10] "Acoustic Cryptanalysis." Accessed: Aug. 20, 2025. [Online]. Available: https://en.wikipedia.org/wiki/Acoustic_cryptanalysis

[11] X. Ji, G. Zhang, X. Li, G. Qu, X. Cheng, and W. Xu, "Detecting Inaudible Voice Commands via Acoustic Attenuation by Multi-channel Microphones," *IEEE Transactions on Dependable and Secure Computing*, no. , pp. 1–17, 2024, doi: 10.1109/TDSC.2024.3355117.

[12] "Havana Syndrome." Accessed: Aug. 20, 2025. [Online]. Available: https://en.wikipedia.org/wiki/Havana_syndrome