Understanding the Effectiveness of Ultrasonic Microphone Jammer

YUXIN CHEN* and HUIYING LI*, University of Chicago
STEVEN NAGELS, University of Chicago
ZHIJING LI, UC Santa Barbara
PEDRO LOPES, University of Chicago
BEN Y. ZHAO, University of Chicago
HAITAO ZHENG, University of Chicago

Recent works have explained the principle of using ultrasonic transmissions to jam nearby microphones. These signals are inaudible to nearby users, but leverage "hardware nonlinearity" to induce a jamming signal inside microphones that disrupts voice recordings. This has great implications on audio privacy protection.

In this work, we gain a deeper understanding on the effectiveness of ultrasonic jammer under *practical scenarios*, with the goal of disabling both visible and hidden microphones in the surrounding area. We first experiment with existing jammer designs (both commercial products and that proposed by recent papers), and find that they all offer limited angular coverage, and can only target microphones in a particular direction. We overcome this limitation by building a circular transducer array as a wearable bracelet. It emits ultrasonic signals simultaneously from many directions, targeting surrounding microphones without needing to point at any. More importantly, as the bracelet moves with the wearer, its motion increases jamming coverage and diminishes blind spots (the fundamental problem facing any transducer array). We evaluate the jammer bracelet under practical scenarios, confirming that it can effectively disrupt visible and hidden microphones in the surrounding areas, preventing recognition of recorded speech. We also identify limitations and areas for improvement.

CCS Concepts: • **Human-centered computing → Personal digital assistants**; *Natural language interfaces*; Sound-based input / output;

Additional Key Words and Phrases: Wearable computing; microphone jamming; voice assistants; privacy; ultrasonic transmission;

ACM Reference Format:
Yuxin Chen, Huiying Li, Steven Nagels, Zhijing Li, Pedro Lopes, Ben Y. Zhao, and Haitao Zheng. 2019. Understanding the Effectiveness of Ultrasonic Microphone Jammer. 1, 1 (April 2019), 17 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Despite the initial excitement around voice-based smart devices for the home and office, consumers are becoming increasingly nervous with the fact that these smart devices are, by default, *always* listening, recording, and possibly saving sensitive personal information they hear [9, 35, 38, 49]. Take home digital assistants as an example. From the outside, they appear to only respond to designated wake-up words (*e.g.*"Alexa" and "Hey Google"). However, their implementation requires them to listen continuously to detect these wake-up words. It has been shown that these devices can monitor and record all voices, sounds and conversations in real time, either maliciously [50],

[*] Both authors contributed equally to this research.

Authors' addresses: Yuxin Chen, yxchen@cs.uchicago.edu; Huiying Li, huiyingli@cs.uchicago.edu, University of Chicago; Steven Nagels, stevennagels@cs.uchicago.edu, University of Chicago; Zhijing Li, zhijing@cs.ucsb.edu, UC Santa Barbara; Pedro Lopes, pedrolopes@cs.uchicago.edu, University of Chicago; Ben Y. Zhao, ravenben@cs.uchicago.edu, University of Chicago; Haitao Zheng, htzheng@cs.uchicago.edu, University of Chicago.

XXXX-XXXX/2019/4-ART $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

by misconfiguration [9], or after compromise by attackers [46]. Leaked audio data can be processed to extract confidential information [26, 27, 50], track user activity [21], count speakers [51], or even extract handwriting content [52]. These negative implications on users' security and privacy are significant and unacceptable.

Clearly, it is important to build tools that can protect users against the potential compromise or misuse of microphones in the age of voice-enabled smart-devices. Recent work along this line [42] shows that ultrasonic microphone jammers can emit an ultrasonic wave that prevents commodity microphones from recording human speech. While these ultrasonic signals are imperceptible to human ears, they leak into the audible spectrum after being captured by commodity microphones, producing a jamming signal inside the microphone circuit to disrupt voice recordings. The leakage is caused by an inherent, nonlinear property of microphone hardware. Not only have researchers built low-cost prototypes using off-the-shelf ultrasonic speakers [42], but also ultrasonic jammers are currently even commercially available to the public.

In this work, we seek a deeper understanding of this approach by studying the effectiveness of ultrasonic microphone jammers under practical scenarios. Current studies [42, 43, 54] focus on disabling a known microphone device by pointing the jammer at it. In contrast, our work considers broader and more complex everyday scenarios. We explore (1) jamming both visible and hidden microphones in an area, (2) strategies to minimize blind spots in coverage of current ultrasonic devices, and (3) jamming under realistic scenarios where either the human speaker or the microphones are moving.

Our work is organized into three phases.

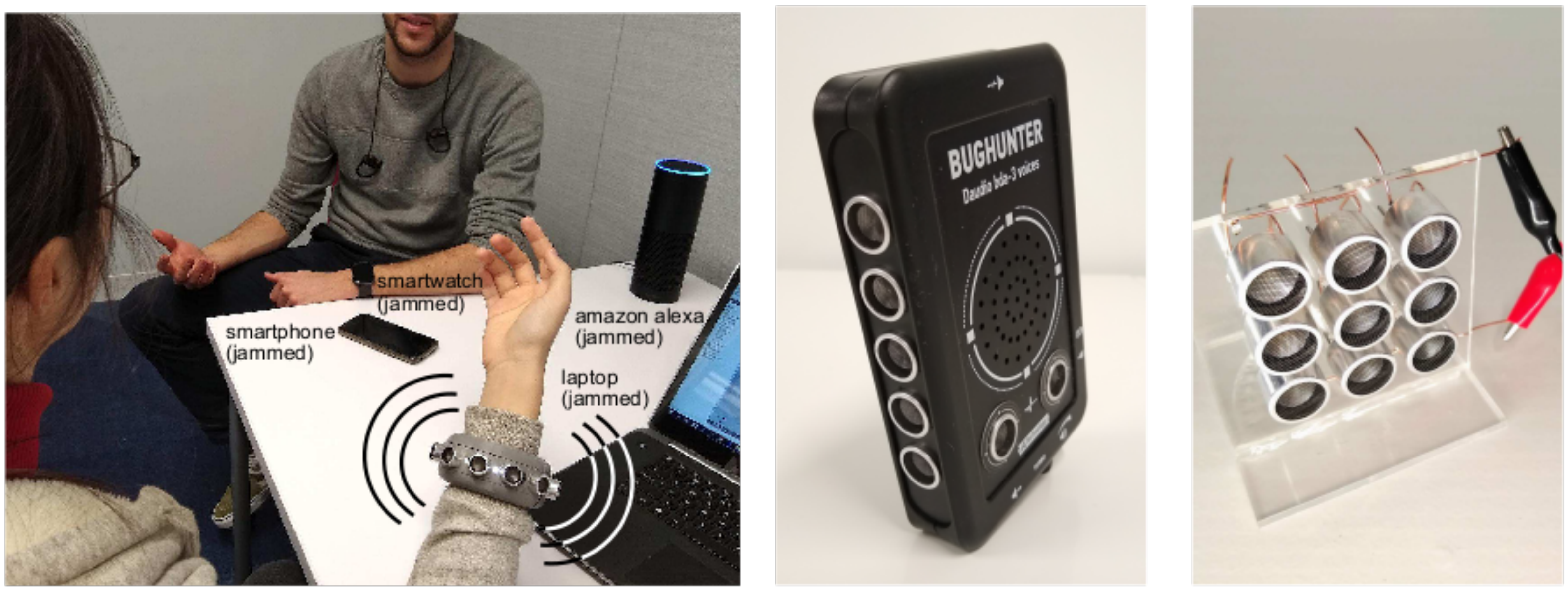
First, we experiment with two of today's ultrasonic jammer platforms, including a commercial product (Figure 1(b)) and a prototype suggested by a recent research paper [42] (Figure 1(c)). We test both jammers, and find that they offer only directional jamming with limited angular coverage and produce blind spots within the covered directions. This is caused by the inherent directionality of commodity ultrasonic transducers and the use of transducer arrays. Since the "speakers" used in these devices operate beyond the audible range, they are denoted with "ultrasonic transducer," and we will refer to these simply as "transducers."

In the second phase, we expand angular coverage by placing multiple transducers on a wearable bracelet, which simultaneously emits ultrasound in many directions. Thus the bracelet jammer mimics an *omni-directional* jammer using inherently *directional* off-the-shelf ultrasonic transducers. More importantly, as the wearable jammer moves with the wearer, normal motion by the user effectively increases coverage and dramatically reduces coverage blind spots (the fundamental problem facing any speaker array). We implemented our design into a self-contained, wearable, jamming bracelet that we depict in Figure 1(a) and later in Figure 12.

Finally, we study the effectiveness of this new wearable ultrasonic jammer in natural settings that have not been considered by prior work, including scenarios with multiple microphone devices, hidden (or covered) microphone devices, multiple users engaged in conversations, and even users in motion while talking. We use these experiments to validate the effectiveness of the our jammer design, and to identify its limitations and areas for improvement.

Our Contributions.

- Understanding and expanding the angular coverage of ultrasonic microphone jammers.
- Increasing jamming coverage through the design of a wearable jammer, which in addition leverages the user's gestures to further reduce blind spots.
- Systematic evaluation in life-like scenarios to validate effectiveness and identify practical limitations of ultrasonic jammers.



(a) Our wearable ultrasonic jammer moves with the user

(b) Commercial ultrasonic jammer from i4, $799

(c) *Backdoor* 3×3 jammer [42] using ultrasonic transducers

Fig. 1. We demonstrate that ultrasound jammers can be more effective if made wearable (a), instead of the current approach, which is to use stationary emitters (b,c). Our prototype is a bracelet that jams surrounding microphones using ultrasound, leveraging the known effect of the microphone's non linearity [42, 43, 54]. We designed and validated the effectiveness of this wearable jamming bracelet in comparison to an (b) existing commercial and (c) state of the art ultrasound jammers. We found that our approach offers omni-directional jamming, increases coverage, removes undesired blind spots, and requires less power than commercial jammers.

2 BACKGROUND AND RELATED WORK

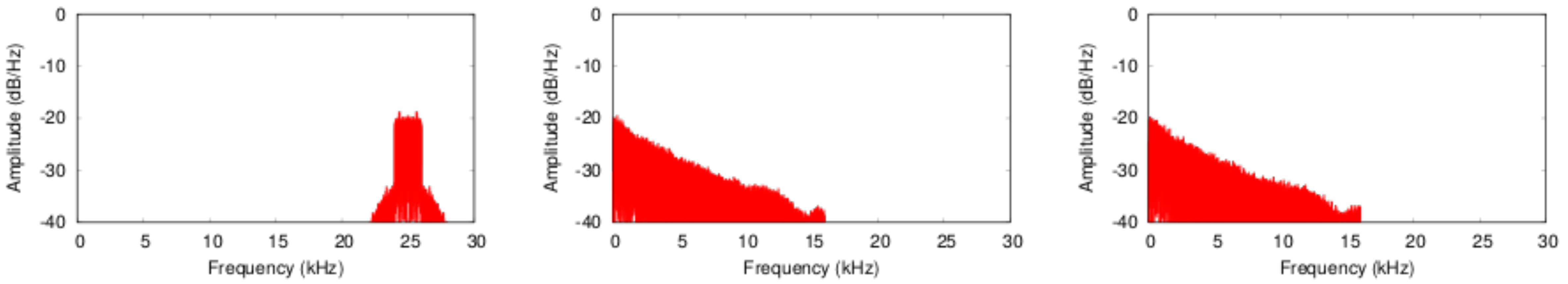
As background, we describe the underlying principle behind ultrasonic microphone jammers and summarize prior work in this area. We also briefly discuss prior work that leverages ultrasonic signals for both sensing and communication to contextualize the usage of ultrasound in HCI.

2.1 Principles of "Silent" Microphone Jamming using Ultrasonic Signals

Recent work demonstrated the feasibility of using ultrasonic transducers to disable nearby microphones. Such jamming is "silent" since ultrasound is inaudible[1] to most humans. Such jamming is possible because ultrasonic signals, after being captured by commodity microphones (MEMS microphones), leak into the audible spectrum and produce a jamming signal *inside* the microphone circuit. This leakage is caused by *hardware non-linearity*, an inherent property of commodity microphone devices [16]. This leakage adds so much audible noise on the microphone circuitry that it effectively renders voice recordings unusable.

Non-linearity in Microphone Hardware. Linearity in microphones refers to its ability to generate an electrical output proportional to the amplitude of the sound input. While electronic components such as amplifiers are carefully designed to be linear over as wide a frequency range as possible, linear recording devices do not exist in practice. Any device, such as a microphone, exhibits non-linearity in some frequency bands. This non-linearity in microphones was originally discovered by musicians and leveraged for sound synthesis [33]. Only more recently has it generated serious impact on the mobile and security communities, given the pervasiveness of microphones in digital voice assistants and smartphones [42, 43, 46, 54].

[1] Ultrasound is sound waves of frequencies above the upper bound of human hearing (20kHz).



(a) Ultrasonic jamming signal

(b) Microphone recording of (a)

(c) Recording of (a) + human speech

Fig. 2. Samples of (a) source ultrasonic jamming signals, (b) after being captured by the microphone; (c) after being captured by the microphone together with human speech.

A commodity microphone consists of four components: a transducer, an amplifier, a low-pass filter, and an analog-to-digital converter (ADC). The low-pass filter has a cut-off frequency of 20KHz (human audible range is [20Hz,20kHz]) to support the ADC. One can represent the microphone input signal S_{in} and output signal S_{out} as follows:

$$S_{out} = \sum_{i=1}^{\infty} A_i S_{in}^i = A_1 S_{in} + A_2 S_{in}^2 + A_3 S_{in}^3 + \ldots \tag{1}$$

where the 2nd term $A_2 S_{in}^2$ and the subsequent terms reflect the non-linear behavior of the microphone hardware.

The process of ultrasonic jamming is simple: the signal generator produces a carefully crafted jamming signal in the ultrasonic band, passes it to the amplifier and then to the ultrasonic transducer. When captured by nearby microphones, the jamming signal leaks into the audible band, and distorts any recordings, particularly those of human voices. Figure 2 shows example traces of the source ultrasonic jamming signal (as amplitude modulated white noise) and as it is captured by the microphone both without and with the presence of human speech. We see that ultrasonic signals (centered around 25kHz) produce microphone recordings that cover up signals of human speech.

2.2 Related Work

Leveraging Microphone Non-linearity. Recently, researchers have leveraged microphone non-linearity as a potential tool for setting up hidden communication channels, disabling microphones, or as an adversarial avenue for injecting hidden voice commands.

A series of projects leveraged this property to attack digital voice assistants [43, 46, 54]. Here, an adversary can play (arbitrary) voice commands modulated in the ultrasonic range and leverage the non-linearity of microphones in home digital assistants (*e.g.* Amazon echo) to force the target device to decode them as normal voice commands. Since the original ultrasonic command is inaudible, the attacker can successfully issue commands without being detected (*i.e.*, heard) by nearby users.

Recent work by Nirupam et. al. [42] leverages non-linearity to build inaudible communication among devices and to jam microphones. The *Backdoor* device utilizes a jamming signal based on either amplitude modulation (AM) or frequency modulation (FM). *Backdoor* is tested in a limited set of experiments, (*e.g.* the jammer pointing to a single microphone) to validate the design. In parallel, there are already commercial products that use ultrasonics for microphone jamming, although all of them are bulky (0.38kg–5kg) and pricey ($799–$6900) [3–5, 15].

Our work is inspired by these existing works on microphone non-linearity, particularly *Backdoor* [42]. However, we dive deeper into this line of research, to examine the effectiveness of ultrasonic microphone jammers under practical scenarios, with the goal of disabling both visible and hidden microphones in the user's surroundings.

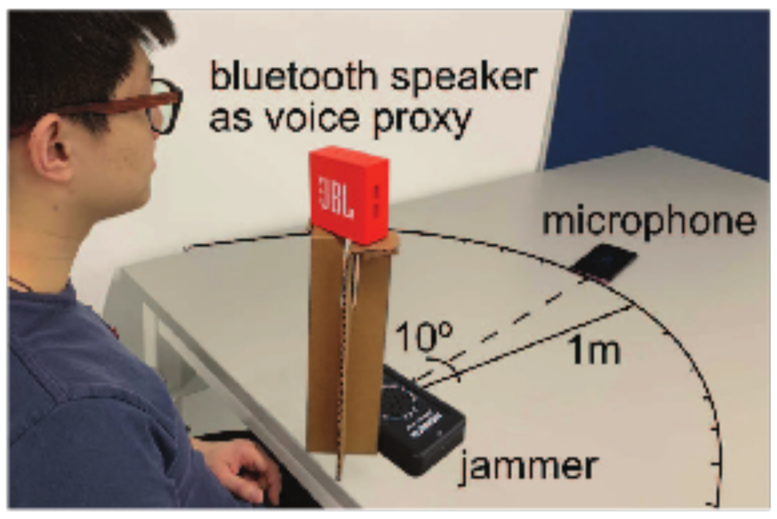
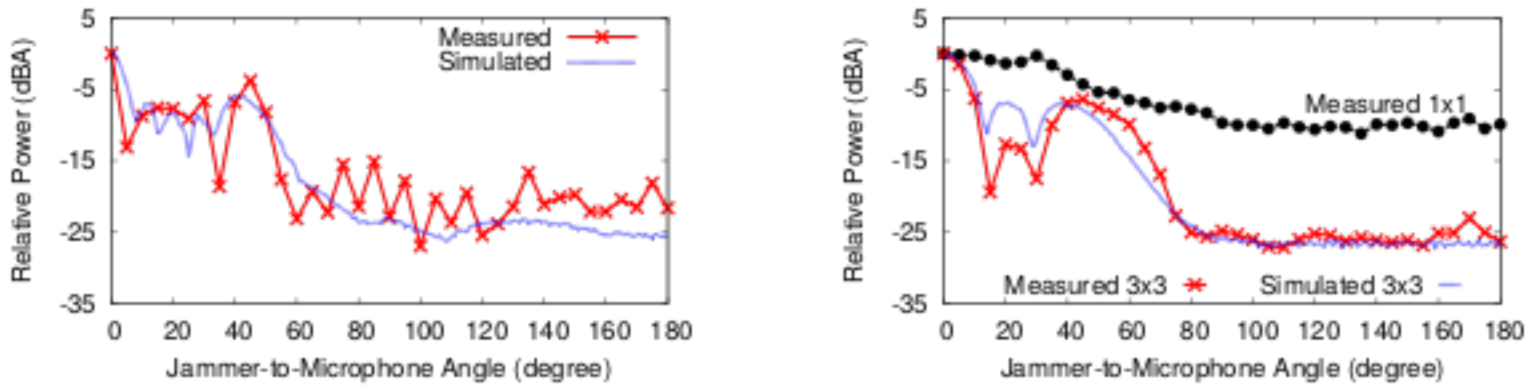


Fig. 3. Our evaluation scenario where we vary α, the angular separation between the jammer and the microphone device.



(a) i4 Tech jammer

(b) *Backdoor 3x3*

Fig. 4. Real-world measurements of the jammer's angular coverage, in terms of the signal power as the jammer-to-microphone angle α increases from 0° to 180°, normalized by that of ($\alpha = 0°$). The distance between the jammer and the microphone is kept at 1m.

Ultrasonic Signals for Device Interactions. Researchers in the HCI community have used signals in ultrasonic bands [20, 37] and near-ultrasonic bands (*e.g.* 18.8kHz) [25, 31] to enable interaction with/among devices. As an example, Gupta et al., utilize Doppler shifts in emitted ultrasound to enable a laptop to perform simple gesture tracking [31]. A variety of smartphone apps use ultrasonic signals as beacons to perform device localization and tracking [8, 19, 29], again based on their leakage to the audible band.

3 EVALUATION OF EXISTING ULTRASONIC JAMMERS

We begin our work by evaluating current designs for ultrasonic jamming devices: (1) a commercial jammer purchased from *Amazon.com* (i4 Technology, $750), and (2) the *Backdoor 3x3* jammer that we built using off-the-shelf ultrasonic transducers following[2] recent work by Nirupam et al. [42, 43]. For both devices, we evaluate the jammer's signal coverage (§3.1) and its effectiveness in disrupting microphone recordings (§3.2).

Jammers. (1) The i4 jammer[3] is shown in Figure 1(b), and consists of a row of five ultrasonic transducers on the side and two more on the top. These transducers operate at the very low end of ultrasonic frequency (24KHz), and unfortunately even produce disturbing audible sounds due to signal leakage in the transducer. This device weighs 380 grams and consumes 4.2W of power. (2) The *Backdoor 3x3* jammer, which is depicted in Figure 1(c), is an array of nine ultrasonic transducers. These transducers operate at 25kHz (±1Hz) and the sound output is completely inaudible. This is not a stand-alone device and its power supply and circuitry are not integrated.

Experimental Setup. Our evaluation considers a typical scenario in which the ultrasonic jammer is used to jam microphones in the room. As shown in Figure 3, we placed the jammer on the table and distributed smartphones (serving as microphones) some distance away. We performed experiments in four rooms of varying sizes and furniture arrangements. We found our measurements to be consistent across all rooms, thus we present aggregated results.

3.1 Jamming Coverage and Blind Spots

Instead of just evaluating the known-effect that distance has on jamming[4], we focus on the angular coverage of the ultrasonic jammer. Since jammers seek to disrupt microphones in the surrounding area, angular coverage is a key performance metric; decreasing blind spots is crucial for effective jamming. Our evaluation used both

[2] We implemented amplitude modulation based jamming using a band limited white noise as the ultrasonic source signal. The signal bandwidth is 1kHz because it is the operating limit of our ultrasonic transducers.

[3] The i4 jammer includes a traditional audio jammer (in the audible band) and a ultrasonic jammer. We only activate the ultrasonic module.

[4] We note that the physical distance covered by a jammer depends on the ultrasonic transducer' power level, the ultrasonic signal frequency, and the volume of the human speaker. All of these can vary across scenarios.

real-world signal measurements using a sound level meter at coarse-grained locations, and simulated signal emission maps at fine-grained locations.

Real-world Measurement. We placed a HT-80A sound level meter (which includes a well-calibrated microphone) 1m away from the jammer. We moved the sound level meter around the jammer (a 1m radius) to vary the angular separation between them. Figure 4 shows the measurement results of the i4 and *Backdoor 3x3* jammers in the absence of any human speech. We present the measured meter power at different degrees of angular separation ($\alpha = 0^\circ$ to 180°), normalized by that of ($\alpha = 0^\circ$). For the *Backdoor 3x3* jammer, we also show result for a single ultrasonic transducer.

We made two key observations.

First, *both jammers have very limited angular coverage.* As shown, moving away from a perfect alignment ($\alpha = 0^\circ$) results in a drop of the jamming signal strength by 25 to 30dB. This implies that in order to disrupt microphones in the surrounding area a user either points directly at the microphones (which is not possible if these are hidden) or each jamming device must raise its transducer's power level by at least 25 to 30dB. Otherwise several jammers are necessary to fully cover potential microphones at different angular positions relative to the user. This lack of angular coverage is caused by the inherent directionality of commodity ultrasonic transducer, as shown in Figure 4(b).

Second, jamming signal power for both jammers shows heavy local fluctuations at different jamming angles. Even within the angular sector of [0°,40°], a subtle angle change of 2° leads to 5-10dB change in jamming power level. This uneven distribution is a fundamental problem facing transducer arrays, often referred to as the blind spot problem [36]. Mutual coupling of signals emitted by different transducers creates unevenness in the jammer's emission pattern, leading to undesired blind spots at certain angular directions.

Mapping Jamming Power using Propagation Models (Simulation). To further illustrate the above two artifacts, we followed the ultrasonic signal propagation model [6] to generate an ultrasonic signal map for both jammers. Our simulation used the single transducer's emission pattern, provided by the manufacturer of the ultrasonic transducer, which we used to replicate the *Backdoor* jammer. We utilize the same emission pattern to simulate the behavior of the i4 jammer (since the manufacturer does not provide any information regarding their transducers). We marked the jammer location as (0m, 0m) pointing at (0m, 1m). While the jammer emits white noise signals of 1KHz on the 25KHz band, we computed the signal power received at each location on the 1m×1m area, normalized by received power at (0m, 0m).

The results plotted in Figure 5 show the relative jamming power for the 3×3 jammer. For visual clarity, we omitted the i4 jammer results as they are very similar. As shown, our simulation confirmed the limited angular coverage (shown as a blue triangle in the bottom right) and the directions of the blind spots (shown as two blue stripes in the top left). These simulations are in line with our earlier observations. Note that we also compared our model-generated signal power values to measured power values in Figure 4, which we found to be consistent.

3.2 Speech Recognition under Jamming

For an end-to-end evaluation of jamming effectiveness, we measured the ability of jammed microphones to record human speech for recognition and content extraction. We tested the two jammers using built-in microphones of three different smartphones: iPhone X (2018), Xiaomi Mi 6 (2017), and iPhone SE (2016). We used the same experimental setup as shown in Figure 3.

In each experiment, we used a high-quality bluetooth speaker as a proxy[5] of a human speaker by playing pre-recorded human speech at a standard sound level of human conversation (55-60dBA measured at 1m away according to [39]). This setup avoided inconsistency caused by potential participants and ensured a fair evaluation

[5] We did experiments to study the potential difference between human speakers and bluetooth speakers by doing extensive recordings of both. For both audio spectrogram and speech recognition, the two are quite similar.

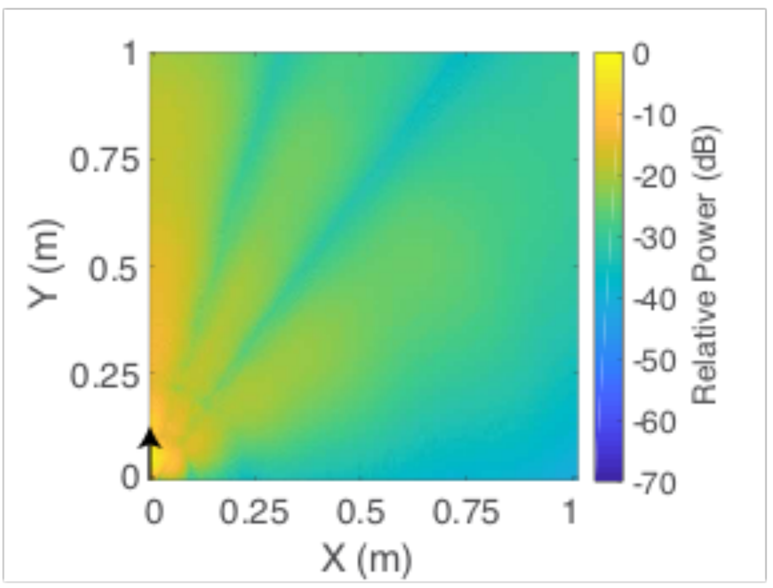
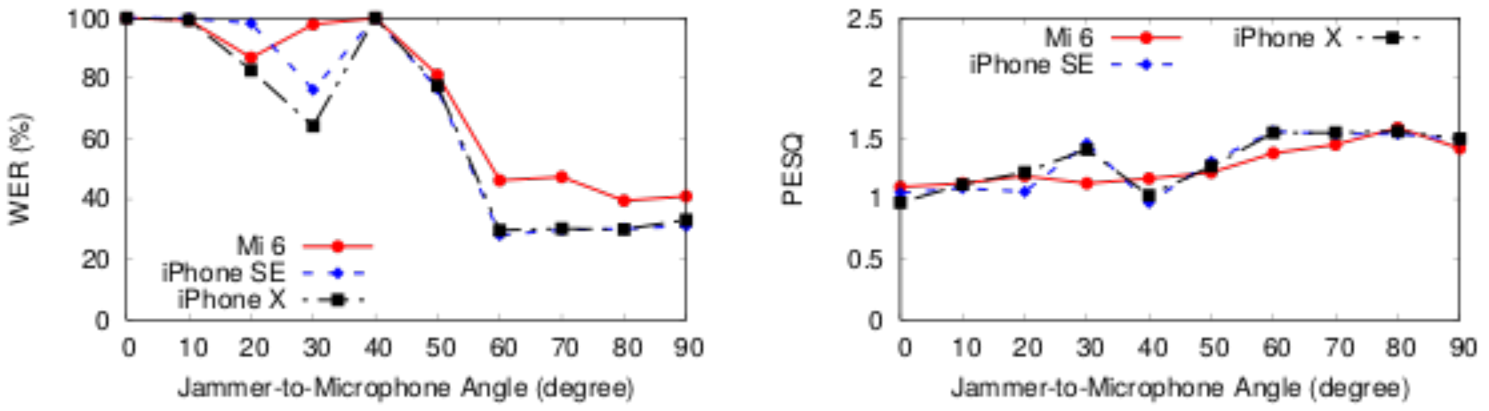


Fig. 5. Simulated jamming power of the *Backdoor* 3×3 jammer an 1m×1m area. The jammer is placed at (0,0) and points to (0,1) (marked by an arrow). The i4 jammer shows a similar pattern.



(a) WER

(b) PESQ

Fig. 6. WER and PESQ when the i4 jammer was 1m away from the target microphone at varying angular separation α, using the same setup of Figure 3. The *Backdoor* 3×3 jammer showed a similar trend (results omitted).

of the jammers. The pre-recorded speech used in our experiments was taken from the LibriSpeech dataset [2], which is commonly used by speech recognition researchers, and includes randomly selected 1000 sentences of clean human speech.

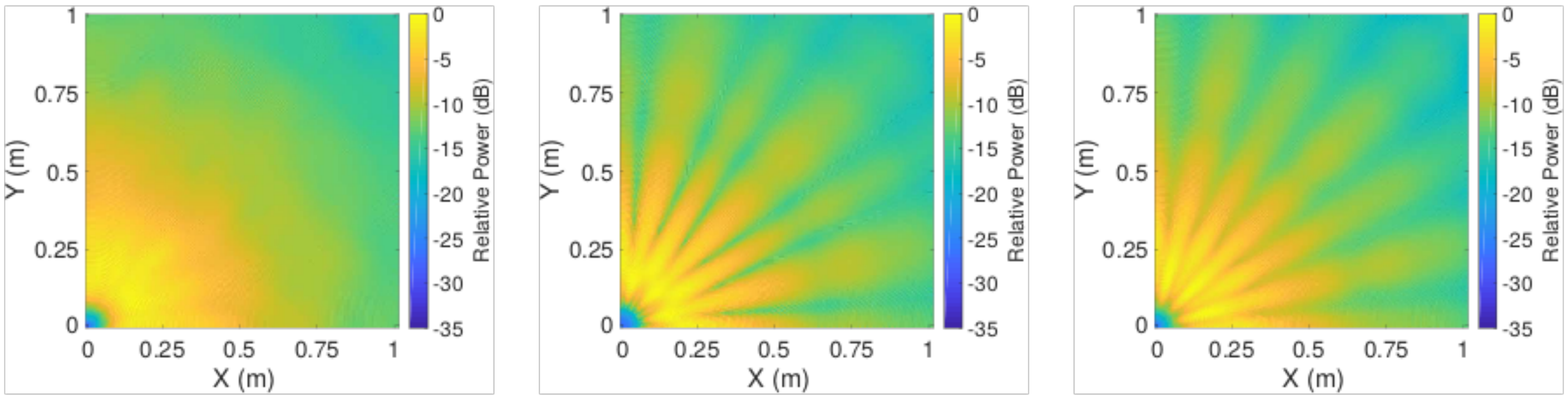
We used two metrics to evaluate the jamming effectiveness. *First*, for each test, we recorded audio on the target smartphone under active jamming, and used the recordings to compute the Perceptual Evaluation of Speech Quality (PESQ) [22], which is an objective voice quality metric. The PESQ ranges between -0.5 and 4.5, where lower scores mean lower voice quality. *Second*, we fed our recordings into five popular speech recognition systems, CMUSphinx [13], Google Speech Recognition [14], Microsoft Bing Voice Recognition [11], IBM Speech to Text [12], and Kaldi toolkit with ASpIRE model [41]. The last two systems are particularly known for their robustness against noisy speech signals. We picked the best speech recognition results of these systems, and recorded its Word Error Rate (WER), the common performance metric on speech recognition. As a baseline, we note that WER without jamming is around 30% for the smartphones.

Effectiveness of Omni-directional Jamming. For the i4 jammer, Figure 6 depicts the WER and PESQ results as a function of the angular separation α, for each of the phones tested. When pointing the microphone device (smartphone) directly ($\alpha = 0^\circ$) at the jammer we observed a WER of almost 100%. But when angular separation α exceeded 50°, we observed a significant drop in WER from 100% down to 30-40%, indicating that the jammer was no longer effective. PESQ also increased from 1 to beyond 1.5, following the same trend (again, a larger PESQ means better voice quality).

Furthermore, the two iPhone models also showed large local fluctuations between 20° and 30°, indicating the existence of blind spots. The effect was observed to a lesser extent on the Mi phone because its microphone is more sensitive to jamming (*i.e.* by having a higher degree of non-linearity). In the interest of visual clarity and brevity, since the results of the *backdoor* 3×3 jammer led to similar observations, we omitted these.

4 A WEARABLE JAMMER BRACELET

Our evaluation results show that while ultrasonic signals can be used to effectively disrupt microphone recordings, existing jammer designs offer very limited angular coverage, and can only target microphones in a few specific directions. This limits the privacy protection they might offer to users in practice, since the user must know the location of nearby microphones and aim their jammer accurately at the microphone.



(a) w/ 24 independent sources

(b) w/ 1 source

(c) w/ 2 independent sources

Fig. 7. Simulated power map of a static wearable jammer in the form of a circular array of 24 ultrasonic transducers, with 24, 2 and 1 input sources. The jammer is placed at (0,0) in the 2D map but 10cm taller (*i.e.* on a user's wrist).

In this paper, we consider a broader and more practical scenario where the goal is to simultaneously jam and disrupt all microphones in an area around the user. For this purpose, we explore the design of a wearable ultrasonic jammer, which not only allows effective omni-directional jamming, but also enables our solution to be highly portable, *i.e.* the jammer follows the human speaker it is designed to protect.

In the following, we describe key design elements of our wearable jammer, and describe how they overcome limited angular coverage and blind spot issues faced by existing jammers (§4.1). We perform detailed benchmarks (§4.2), and describe our current prototype as a self-contained bracelet (§4.3).

4.1 Key Design Elements

Omni-directional Jamming via Circular Array. In theory, omni-directional signal emission can be achieved using a circular array of ultrasonic transducers, which emit signals simultaneously in many directions. In practice, the (angular) coverage of the jammer depends heavily on the number of independent signal sources used to drive these ultrasonic transducers.

If the circular array can provide an independent input source for each ultrasonic transducer, the simulated power map of the jammer will display a uniform angular coverage, as in Figure 7(a). But this is impractical for wearable devices, since each input source is a high-resolution digital audio player, and no more than 2 can fit on a form factor consistent with a single wearable device.

When the array of transducers is driven by a smaller number of input sources, interactions between transducers will again produce blind spots (or blind angular directions), just like those produced by rectangular arrays in the existing jammers. Figures 7(b) and 7(c) plot the simulated power map where our circular array jammer has one or two input sources. While the jammer radiates signals from all directions, we can observe multiple strips of locations where the jamming signal is 10dB lower than nearby locations.

We note that for the above simulations, the jammer is placed at (0,0) (as in Figure 3) but 10cm taller since it is now on the user's wrist (rather than the table). As such, the signal is weak at locations within 5cm to the jammer due to the lack of vertical coverage. This can be addressed in practice by adding more transducers along the vertical direction.

Removing Blind Spots via User Movement and Gestures. A significant benefit to embedding the jammer as a wearable device is that we can mitigate the blind spot problem by leveraging natural user movement, *e.g.* making a gesture or walking back and forth. As the wearable jammer moves with the user (her wrist), its instantaneous signal emission map also changes. Such natural signal fluctuations mimic the fading effect in radio transmissions, creating instantaneous signal peaks and valleys[6]. These frequent signal peaks, although short in time, can effectively disrupt the recording of individual words by a nearby microphone.

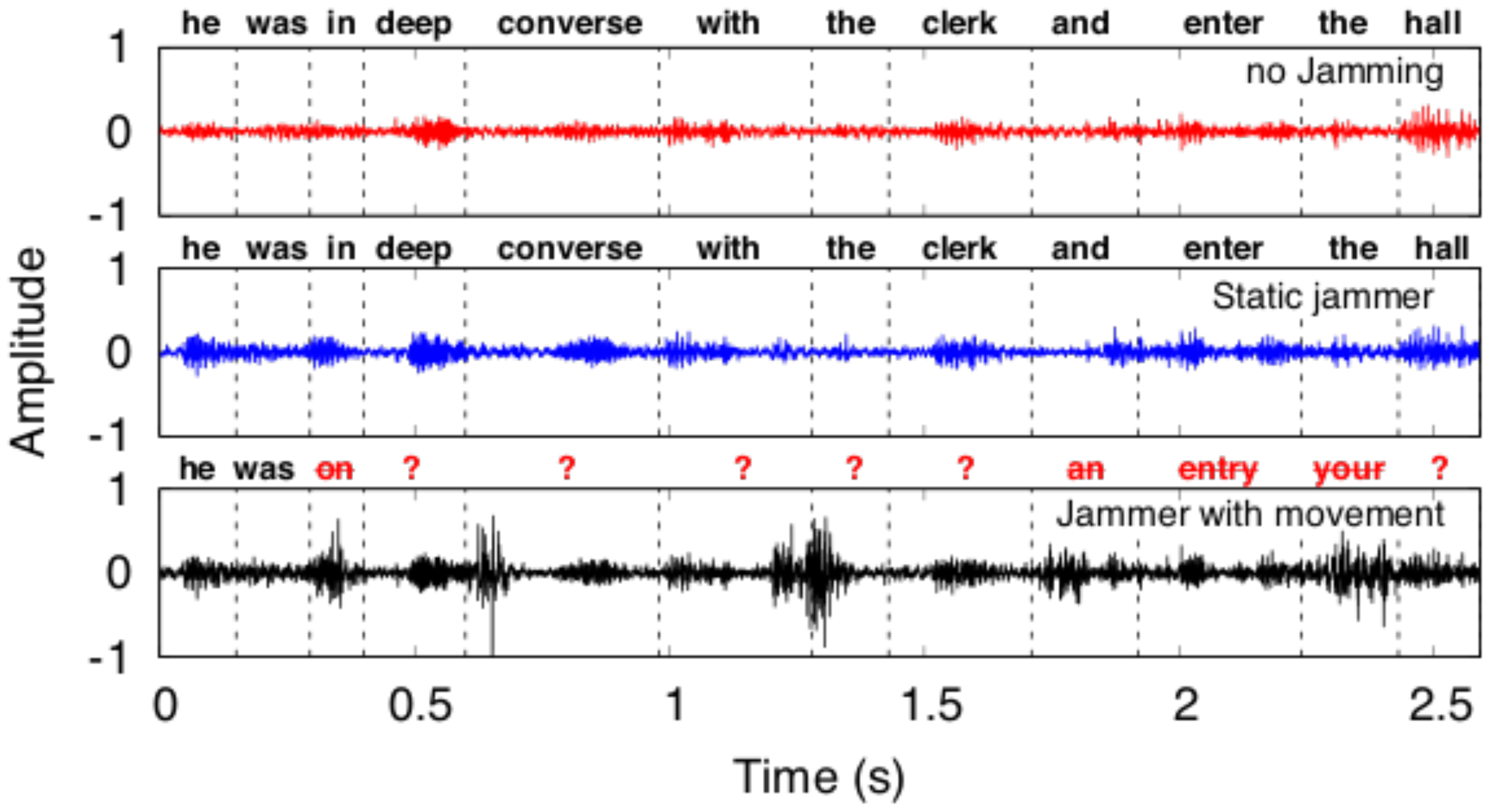


Fig. 8. A microphone is placed in the blind spot of a wearable jammer. If the jammer is static, its jamming has minimal impact on decoding human speech. The same jammer with small user gestures produces artifacts that make many words unrecognizable or introduce errors in word recognition.

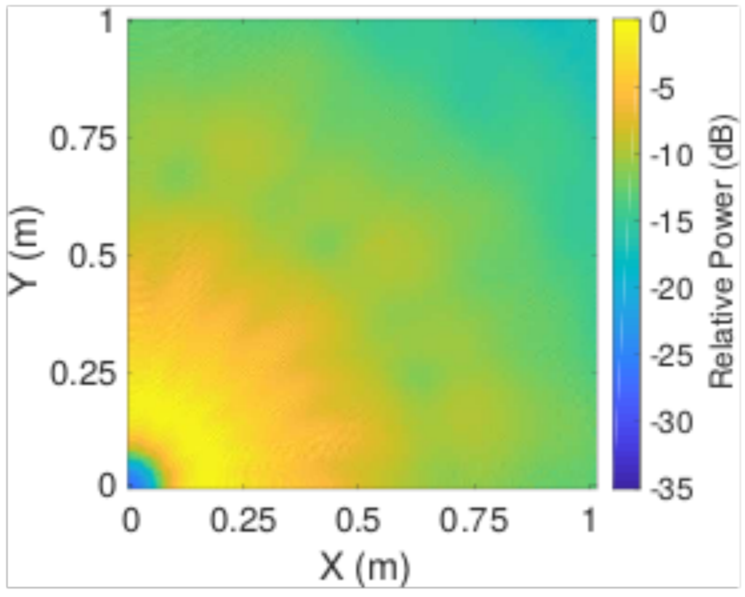


Fig. 9. Simulated power map of a wearable jammer with 1 source under small user gestures. The jammer is placed at (0,0) in the 2D map but 10cm taller (*i.e.* on a userâĂŹs wrist).

We test the impact of movement on jamming efficacy through an experiment. Using a jammer with 24 transducers, we carefully place a microphone in its blind spot. We then test the ability of the microphone to pick up and decode pre-recorded human speech from the LibriSpeech dataset [2] in 3 scenarios: normal recording (no jamming), static jammer (user staying completely still), and jammer with movement (user with small gestures). Decoded results in Figure 8 show that a microphone in a jammer's blind spot can decode human speech near perfectly, but even small gestures are enough to make the majority (10 out of 12) of words in the test unrecognizable. The two unaffected words were short, monosyllabic words that recorded before jammer gestures started disrupting the signal.

With this in mind, Figure 9 plots the computed signal power map of the wearable jammer with natural human movements (random rotation by up to 45°), but with only a single input source. Since the instantaneous jamming signal fluctuates significantly, we show the map averaged over a window of 0.4s (average duration of a human spoken word [1]). The resulting map closely approximates the signal map of the (oracle) case where each transducer has its own input source (shown in Figure 7).

Colocation with the Human Speaker. A wearable jammer is always co-located with the human speaker it seeks to protect. This not only increases coverage (since the jammer moves with the user), but the short distance between the jammer and the speaker's vocal cords also prevents the use of beamforming microphone arrays to separate the signals of the human speaker and the jammer [18].

4.2 Validation via Benchmarking Experiments

Next, we use detailed benchmark experiments to evaluate our wearable jammer design. Again we consider the scenario in Figure 3, but replace the jammer with our wearable prototype (placed on the human speaker's wrist, 10cm above the table). We examine both signal power distribution and speech recognition accuracy.

[6] Since the ultrasonic carrier frequency (25KHz) is much higher than that of human voice (85-180Hz), its signal fluctuation will be significantly larger and more frequent than that of human voice.

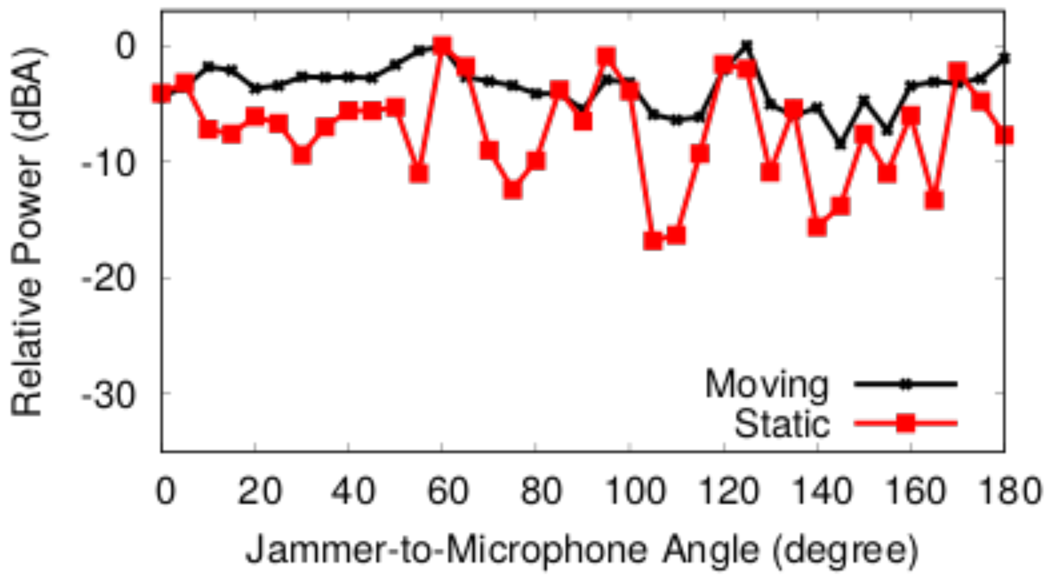


Fig. 10. Angular coverage of the wearable jammer when completely static and under natural human movement. Jammer is 1m away from microphone.

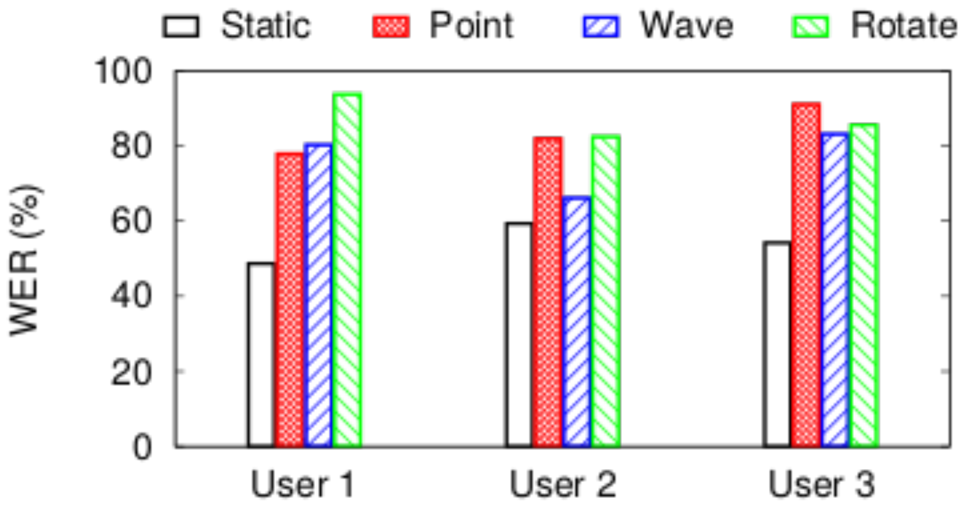


Fig. 11. WER of when each of the three human volunteers wears the jammer, and applies three gestures (point, wave, rotate) in their own styles. The microphone is placed at the blind spot of the jammer (when it is static). The naturally-occurring human gestures largely increase jamming effectiveness.

Angular Coverage of Wearable Jammer. We first look at the angular coverage of our wearable jammer. As in Figure 3, we measure signal power when the sound meter is 1m away from the jammer, but with different angular separation α to the jammer/human speaker. Figure 10 shows the measured signal power (normalized by the highest power seen across $\alpha = 0^\circ$ and 180°), for cases where the jammer is completely static and moving with the wearer. Clearly, the jammer's movement helps to smooth the signal power across α, effectively removing blind spots to offer omni-directional jamming.

Speech Recognition Accuracy. We also validate the benefits of natural jammer movement. In each experiment, while keeping the jammer static, we identify a blind spot (1m away from the jammer) and place a microphone device there. We ask our volunteer to make one of the three popular gestures (point, wave, rotate) as suggested by [10], and run the experiment for 30 sentences per gesture per round. To minimize inconsistency in human speech, we again use the bluetooth speaker to replay the same human speech audio clip.

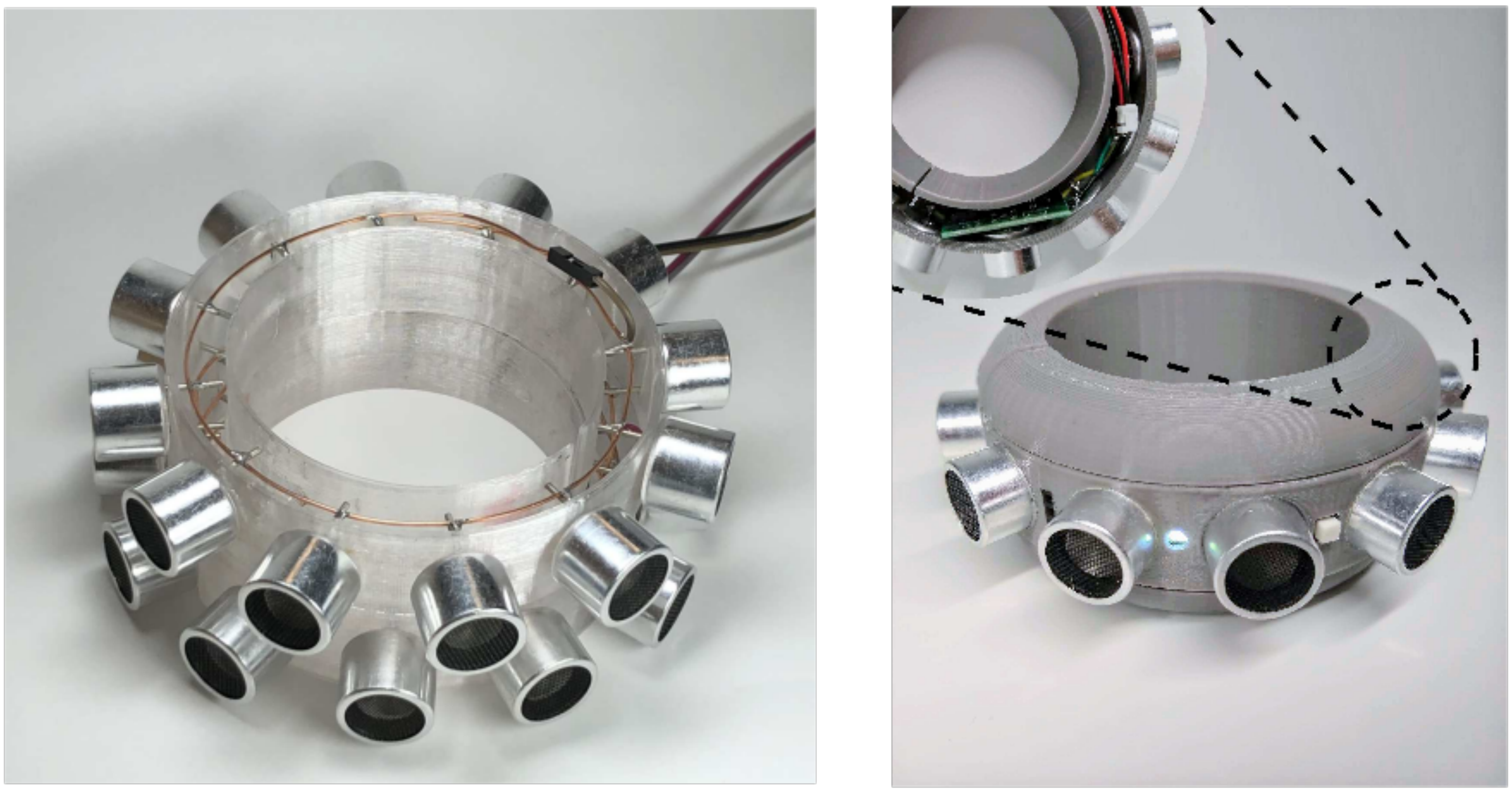
Figure 11 shows the WER for three volunteers. We can see that naturally occurring hand gestures greatly increase jamming effectiveness and remove blind spots. The resulting WER increases to 70-92%.

4.3 Prototype

We choose to design the wearable jammer as a bracelet that can be easily activated [17, 28, 40, 48, 53] whenever the user decides to engage in a private conversation. We now provide all the necessary technical details to implement our prototypes. To assist readers in replicating our prototypes, we provide all microcontroller code, signal parameters, circuits, and 3D files[7].

Our initial version, which is depicted in 12(a), of our prototype bracelet is made from a simple 3D printed ring that holds an array of 12 ultrasonic transducers (NU25C16T-1, 25kHz) evenly spread in its perimeter. These transducers all connect to a single ultrasonic signal generator and an audio amplifier (PAM8403), which both sat outside the bracelet. For simplicity in our initial version, we used a Galaxy S7 edge smartphone as the signal generator, which is capable of playing up to 192Khz through the line-out port. We configure the signal generator to produce amplitude modulated white noise centered on 25kHz (±1 Khz). We configured the audio amplifier such that all the ultrasonic transducers operate at their maximum power level. In all our experiments, we used two of these bracelets stacked together (totalling 24 transducers), allowing us to get a sense for the upper bound of the design. As our experiments confirmed that making a jammer wearable does improve its effectiveness (*e.g.*, it reduces blind spots), we engineered an improved and stand-alone version.

[7] Anonymized for review.



(a) Our initial prototype of the jamming bracelet; in this prototype the signal generator, amplifier, and power supply sat outside the device (not shown).

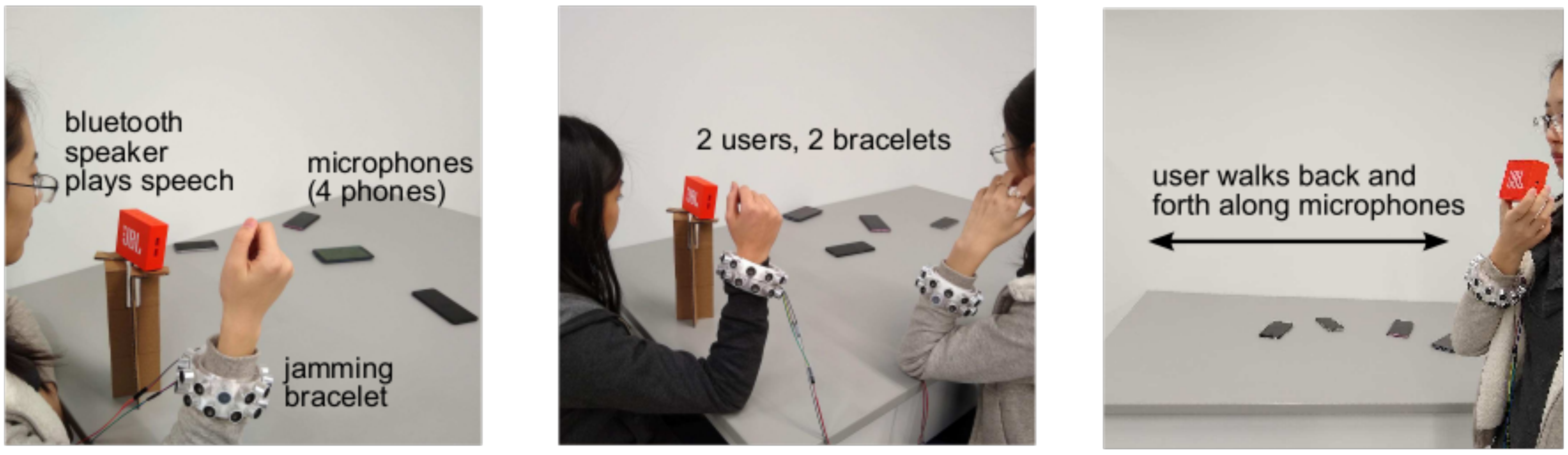
(b) Our final prototype is a self contained wearable device (battery, signal generator, microcontroller, touch button, LED status, and amplifier are all integrated).

Fig. 12. Our prototypes: (a) initial version used for experiments and (b) an improved and stand-alone wearable jammer.

Our improved prototype, which is depicted in Figure 12(b), is a self contained wearable device comprised of the following components: a 3D printed shell, 12 ultrasound transducers (same as the above), a small low-powered signal generator (AD9833, up to 12.5MHz with 0.004Hz programmable steps), a ATMEGA32U4 microprocessor, an LED status, a touch button, and a small rechargeable LiPo battery (105mA, which is 26 times smaller than an iPhoneX's battery). The microprocessor controls the signal generator via Serial Peripheral Interface (SPI).

We measured energy consumption of our prototype bracelet. It consumes about 148mW when jamming (which is 28 times less energy than the i4 jammer). To put this into perspective, our jammer uses roughly *15%* the energy consumed by the internal WiFi module of a typical smartphone [24, 47]. Our resulting device (including its battery) weighs 91 grams.

Current Limitations. While we believe our prototypes are a step towards wearable jammers, they have limitations. First, like all current ultrasonic jamming techniques, the user cannot selectively jam devices: i.e., a user cannot choose to avoid jamming their own smartphone while the signal is on. On this limitation, our approach does provide much more control than existing stationary jammers, because the user does not have to walk all the way to the jammer to disable it and can do so by simply touching the bracelet. Furthermore, in Section 6 we present some initial steps to further include selective jamming. Second, due to hardware limitations, the current prototype is larger than a typical bracelet and has limited vertical coverage. However, we believe that switching to newly developed ultrasonic transducers – like [7] (which are 1.4 mm in diameter) – enables the construction of a slim, stylish version of our wearable jammer. Despite these shortcomings, we believe this prototype offers a great blueprint towards a low-cost and ubiquitous microphone jammer.


bluetooth speaker plays speech
microphones (4 phones)
jamming bracelet
2 users, 2 bracelets
user walks back and forth along microphones

(a) One participant setup.

(b) Two participants around the microphones.

(c) Participant walks back and forth in front of microphones.

Fig. 13. User study setup we designed to investigate our jammer in more complex situations that a user might face at home, these include: walking around the room, interacting with other users, etc.

5 VALIDATION IN REALISTIC SCENARIOS

In this section, we test the wearable jammer bracelet using four scenarios designed to capture realistic situations that one might face at home or at work, these are depicted in Figure 13.

Experimental Procedure. We used a within-subjects design with 2 interface conditions (with our jamming bracelet or without) and 5 tasks.

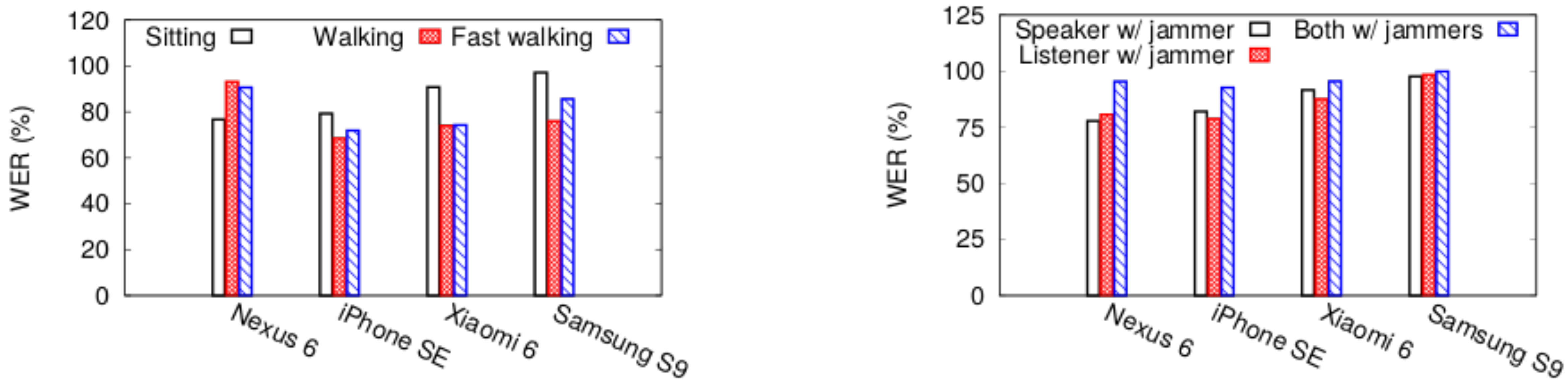
Setup. Participants were asked to wear our jamming bracelet (our initial laser-cut prototype, featuring 2x12 transducers) on their dominant arm. We left the participants in the experimental room for 80 minutes (as in Figure 13(a)), and they could do whatever normal daily activities they wanted to, either sitting at or walking around the table, as long as they completed the "tasks" we planned for them (see below). Participants later reported activities such as looking at their phones, reading books, etc. We also asked them to not speak, so we could again use our bluetooth speaker playing pre-recorded speech as our consistent proxy for a human voice. Around the conference table, at a distance between 0.8m and 1m, we placed 4 different smartphones (Samsung Galaxy S9, Xiaomi Mi 6, iPhone SE, and Nexus 6), which we used to record the resulting jammed speech that was later used to perform speech recognition.

Tasks. We asked participants to perform 5 tasks during their experimental time: (1) sit on a chair at the table with a bluetooth speaker positioned in front of them (see Figure 13(a)); (2) same activity, but with the bluetooth speaker positioned away from the participant to mimic another voice in the room; (3) work with another participant, where one was tasked as the "speaker" while the other was a "listener" (see Figure 13(b)); (4) same as before but with the participants flipping their roles as speaker and listener; and finally, (5) walk back and forth in front of the table. Each task took 10 minutes.

Participants We recruited six participants from our local institution for these experiments (aged 20-30 years old).

5.1 Jamming Visible Microphones

We now detail the results for all tasks in which the microphones were placed visibly on the table (as depicted in Figure 13). These placements are resemblant of how home-assistant devices are placed in a user's living room.



(a) Single Speaker

(b) Two-user Conversation

Fig. 14. Speech recognition results for the tasks in which the microphones were placed visibily on the table. (a) A single participant, either sitting or walking. (b) Two participants, either one or both wearing our jammer.

Participant sitting down. We compare the speech recognition results when the participant is with and without the jammer bracelet. Results are quite consistent across participants, so we aggregate speech recognition results for all participants, shown in Figure 14(a). Clearly, jamming was effective for different smartphones positioned at different locations. Baseline WER was 30% without our bracelets, and ranged between 75% and 100% when the participant wore the bracelet. The PESQ results are consistent and thus omitted for brevity.

To see whether 75% WER is sufficient to jam voice recordings, we look at the recognized words. Table 1 shows four examples of the recognized sentence in different WER cases. At a WER of 30%, there was some loss and mis-recognition. But at a WER of 75%, the recognition results had almost no overlap with the original sentence. Further, we found that in higher WER cases, what few words were actually recognized were not at all useful for understanding. For example, in WER of 99% cases, we can only recognize words of "and," "do," "sure," "his," "show," "this," "make," "to," "of," "for," "the," "think," "is," "he"].

Participant Walking Around. We asked participants to randomly walk near the table (distance within 0.8 m), while holding the bluetooth speaker near their mouth in one hand, as shown in Figure 13(c). Participants were asked to walk in a mix of styles, such as fast, slow, and at their own pace. Each round takes 1 minute of walking. Figure 14(a) shows the speech recognition results. In all cases, WER with jamming is high (>70%). We observe no significant difference in speech recognition performance when walking at different speeds.

Multi-participant Scenario. We consider a two-person scenario, to see if another jammer-equipped person in the same room can help improve the jamming effect. We asked volunteers to work in pairs. For each round of the experiment, a pair of volunteers will sit in the room and do whatever they want. One will be the "speaker," *i.e.* we place the bluetooth speaker in front of them, and then we evaluate jamming effects where: 1) only the speaker wears the bracelet; 2) only the listener wears the bracelet; 3) both participants wear the bracelet. Figure 14(b) shows the results. We see that if both of them wear the bracelets, the jamming performance is the best - the WER can reach more than 90% for all 4 microphones (smartphones). With an extra jammer, received jamming power increases and boosts WER. When only one bracelet is worn, there is little difference based on whether the jammer is worn by the speaker or the other participant; both are able to disrupt the voice recording/recognition.

5.2 Jamming Hidden Microphones

Next, we consider the task of jamming microphones hidden nearby, *e.g.* a smartphone hidden inside a pocket or an attacker trying to stealthily record a conversation by covering up a microphone.

Sentence Transcript	**Clean case of WER 30%**	**Jamming case of WER 75%**
Now to bed boy.	No too bed boy.	and they weren't too bad boy.
Gamewell to the rescue.	<noise> to the rescue.	You should be able to get rid.
Most of all, Robin thought of his father what would he counsel.	Most of all, Robin thought of his father what would he counsel.	List of a father we see counsel.
He began a confused complaint against the wizard, who had vanished behind the curtain on the left.	You begin to confused complaint against the wizard would vanished behind the curtain on the left.	Get confused complete to get to the wizards.

Table 1. Examples of recognized sentence in clean speech case (WER 30%) and jamming case (WER 75%.)

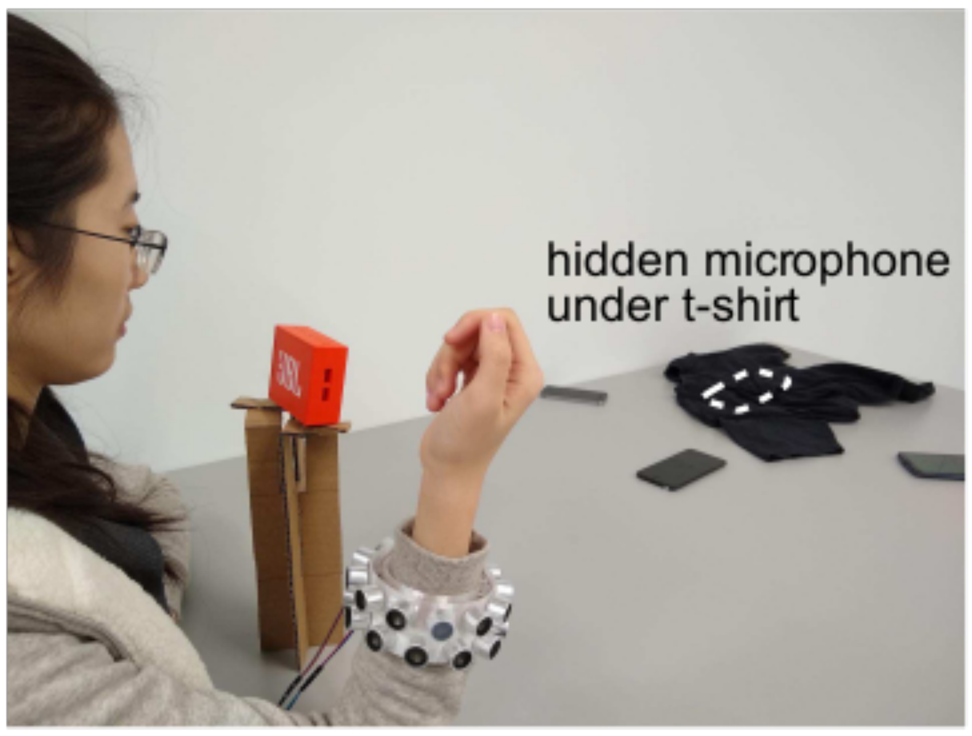


Fig. 15. Experimental setup we designed to investigate how our jammer disrupts hidden microphones. Here, we depict one of the cases we explore, to hide a microphone under a T-shirt (but we also explored covering them up with boxes, etc).

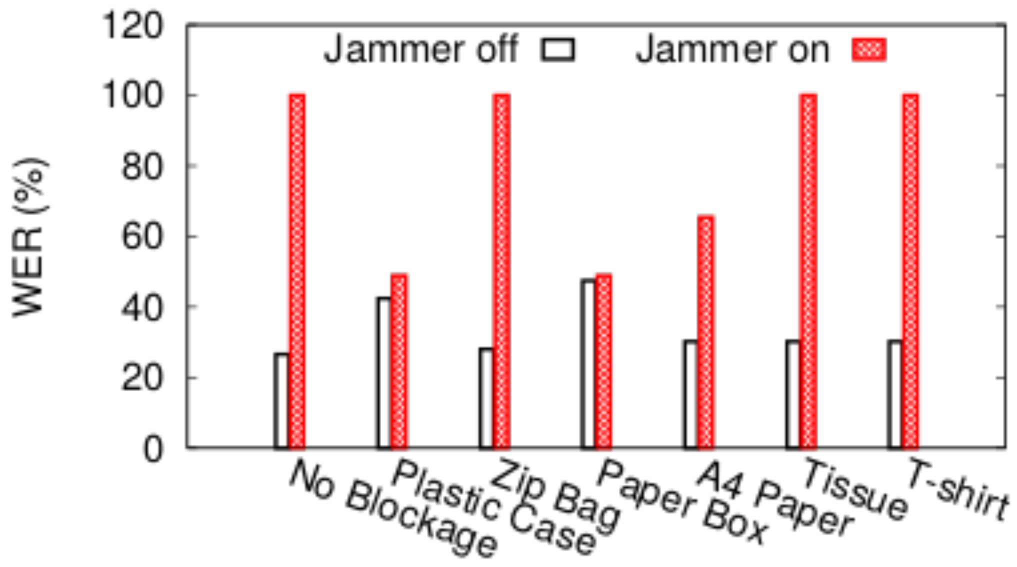


Fig. 16. Speech recognition results when the microphone is covered up with various objects.

To understand how different kinds of blockages can affect jamming performance, we put different blockage materials on the microphone (shown in Figure 15). We then record the human voice audios with and without jamming turned on. From our results in Figure 16, we see that zip bags, tissues, and T-shirts have little impact on jamming performance. Although the jamming effect drops when blocked by A4 paper, WER is still over 60%. When the microphone is covered by a plastic case or paper box, jamming performance drops considerably, but the WER for clean audio also increases. In other words, blocking our jammer signal also blocks normal audio. In general, most thin blockage materials have little impact on jamming performance while thick blockage materials will decrease the jamming performance. However, thick blockage materials will also decrease the quality of the audio recording.

6 DISCUSSION

6.1 Non-linearities of Microphone Hardware: Transient or Permanent?

One may question whether the non-linearity of today's MEMS microphone hardware is just a transient artifact of today's devices, and whether it will disappear with improvement to microphone hardware. We believe non-linearity is likely permanent for the foreseeable future, because MEMS microphones for smartphones and voice-interface IoT devices are designed for low-cost and small form-factors [23, 34, 45]. Device manufacturers have little incentive to use materials with stronger linearity properties, given the associated increases in cost and device size. Our experiments on multiple smartphone devices released over the last 5 years show that non-linearity has not decreased.

6.2 Ethics, Safety and Unintentional Jamming

Building mobile systems with ultrasonic signals requires an implicit assumption that ultrasonic signals are harmless to surrounding users and their devices. This assumption does not always hold. There was even speculation that high powered ultrasonic subharmonics have been weaponized to produce undetectable discomfort to human targets [30], though the effects were later identified as being caused by the Indies short-tailed cricket [32]. Here we discuss some considerations for ethics and safety and possible risks of unintentional jamming.

Risks and Experimental Precautions. Our proposed system uses ultrasonic frequencies in the 25kHz range, while the upper limit frequency that the human ear can hear is between 15 and 20kHz. The U.S. Occupational Safety and Health Administration (OSHA) warns that audible subharmonics can be harmful at intense sound pressures of 105 decibels or above [44].

To examine the safety of our jammer device, we also measured the sound pressure level (SPL) of our prototype bracelets using decibel meters. Our prototype bracelets showed a maximum sound pressure of <100dB when measured directly at the transducer, which quickly attenuates down to 73dB when 25cm away.

During our experiments, coauthors conducted tests with multiple bracelets (24 transmitting transducers), for hours at a time, and reported no pain or discomfort. All experiments were designed to keep jammers at least 30cm away from human users' ears.

Unintentional Jamming. It is possible that our jammer could accidentally jam legitimate microphones in nearby IoT devices, including hearing aids or personal emergency response devices. This is of course non-ideal. But given that our blocking range is limited, we believe it would be easy for a user to detect an unintentionally jammed person or device and turn off jamming as appropriate. More work is necessary to understand the impact of ultrasonic signals on these devices and to design workarounds.

Finally, we consider a scenario in which a user is trying to have a private phone conversation with the jammer turned on. Users would like to avoid inadvertently jamming their own phone. We have conducted some initial tests, which show that, when a user speaks into her phone (held up to her face), she can effectively block the ultrasonic jamming signal by shielding her phone microphone and mouth with a hand. Here the user's hand selectively blocks the bracelet, but not her own mouth. As ongoing work, we are investigating device designs that would make it easier for the jammer to speak into her own phone.

7 CONCLUSIONS

This paper describes our efforts in designing and validating a wearable device that restores a user's sense of agency and control over their personal voice privacy. Users can tap the device on, and be confident that their conversations remain private from nearby microphones (both visible and hidden). As a disconnected device with no built-in microphone, our jammer provides a tamper-resistant tool users can trust. Our prototype is lightweight (91g), low-cost ($36 for a complete 24-transducer bracelet, including circuitry and battery), power-efficient (148mW), uses only commodity components, and can effectively jam a range of listening devices from digital assistants to smartphones.

There is certainly room for improvement in our current prototype. For instance, recent developments in transducer design might result in commercially available 1.4 mm diameter [7] transducers; these can dramatically reduce the bracelet size. This will open up new wearable form factors that go beyond our bracelet, such as badges, rings and so forth. In addition, further tuning of the ultrasonic signal will enable jamming over longer distances, while enabling selective exceptions to more easily allow the user to carry on a phone conversation.

REFERENCES

[1] [n. d.]. Average Speaking Rate and Words per Minute. https://virtualspeech.com/blog/average-speaking-rate-words-per-minute.

[2] [n. d.]. LibriSpeech Dataset. http://www.openslr.org/12.

[3] [n. d.]. New Generation of High Grade Smartphone Scrambler. https://www.globaltscmgroup-usa.com/.

[4] [n. d.]. SILENT ULTRASONIC MICROPHONE DEFEATER. https://www.uspystore.com/silent-ultrasonic-microphone-defeater.

[5] [n. d.]. Speech jammer TOWER-A for blocking proffessional microphones / counter-surveillance. https://www.detective-store.com/speech-jammer-tower-a-for-blocking-proffessional-microphones-counter-surveillance-1516.html.

[6] [n. d.]. Ultrasonics. https://courses.washington.edu/bioen508/Lecture6-US.pdf.

[7] 2014. Tiny Forward-Looking Ultrasound Transducer for 3D View Inside Coronary Vessels. https://www.medgadget.com/2014/02/tiny-forward-looking-ultrasound-transducer-for-3d-forward-view-inside-heart-blood-vessels.html.

[8] 2015. https://www.theatlantic.com/technology/archive/2015/11/your-phone-is-literally-listening-to-your-tv/416712/. Your Phone Is Listening - Literally Listening - to Your TV.

[9] 2017. Google is permanently nerfing all Home Minis because mine spied on everything I said 24/7 [Update x2]. https://www.androidpolice.com/2017/10/10/google-nerfing-home-minis-mine-spied-everything-said-247/.

[10] 2018. 20 Hand Gestures You Should Be Using. https://www.scienceofpeople.com/hand-gestures/.

[11] 2018. Bing Speech. https://azure.microsoft.com/en-us/services/cognitive-services/speech/?v=18.05.

[12] 2018. IBM Speech to Text. https://www.ibm.com/watson/services/speech-to-text/.

[13] 2018. OPEN SOURCE SPEECH RECOGNITION TOOLKIT. https://cmusphinx.github.io/.

[14] 2018. SpeechRecognition 3.8.1. https://pypi.org/project/SpeechRecognition/.

[15] 2019. Hidden Microphone dictaphone Bug Recording supressor ultrasonic + Noise Generator by i4 Technology. https://www.amazon.com/Microphone-dictaphone-Recording-supressor-ultrasonic/dp/B01MG4WACJ/.

[16] Muhammad Taher Abuelma'atti. 2003. Analysis of the effect of radio frequency interference on the DC performance of bipolar operational amplifiers. *IEEE Transactions on Electromagnetic compatibility* 45, 2 (2003), 453–458.

[17] Leonardo Angelini, Maurizio Caon, Stefano Carrino, Luc Bergeron, Nathalie Nyffeler, Mélanie Jean-Mairet, and Elena Mugellini. 2013. Designing a Desirable Smart Bracelet for Older Adults. In *Proceedings of ACM Conference on Pervasive and Ubiquitous Computing Adjunct*. 425–434.

[18] Xavier Anguera, Chuck Wooters, and Javier Hernando. 2007. Acoustic beamforming for speaker diarization of meetings. *IEEE Transactions on Audio, Speech, and Language Processing* 15, 7 (2007), 2011–2022.

[19] Daniel Arp, Erwin Quiring, Christian Wressnegger, and Konrad Rieck. 2017. Privacy threats through ultrasonic side channels on mobile devices. In *Proc. of EuroS&P*.

[20] Md Tanvir Islam Aumi, Sidhant Gupta, Mayank Goel, Eric Larson, and Shwetak Patel. 2013. DopLink: Using the Doppler Effect for Multi-device Interaction. In *Proceedings of ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*.

[21] Kazuki Awaki, Chun-Hao Liao, Makoto Suzuki, and Hiroyuki Morikawa. 2016. Speaker-less Sound-based 3D Localization with Centimeter-level Accuracy. In *Proceedings of ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct (UbiComp)*.

[22] John G Beerends, Andries P Hekstra, Antony W Rix, and Michael P Hollier. 2002. Perceptual evaluation of speech quality (pesq) the new itu standard for end-to-end speech quality assessment part ii: psychoacoustic model. *Journal of the Audio Engineering Society* 50, 10 (2002), 765–778.

[23] Joseph A. Boales, Farrukh Mateen, and Pritiraj Mohanty. 2017. Micromechanical microphone using sideband modulation of nonlinear resonators. *Applied Physics Letters* 111, 9 (2017), 093504.

[24] Aaron Carroll and Gernot Heiser. 2010. An Analysis of Power Consumption in a Smartphone. In *Proceedings of USENIX Annual Technical Conference*.

[25] Ke-Yu Chen, Daniel Ashbrook, Mayank Goel, Sung-Hyuck Lee, and Shwetak Patel. 2014. AirLink: Sharing Files Between Multiple Devices Using In-air Gestures. In *Proceedings of ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*.

[26] H. Chung, M. Iorga, J. Voas, and S. Lee. 2017. Alexa, Can I Trust You? *Computer* 50, 9 (2017), 100–104.

[27] Hyunji Chung and Sangjin Lee. 2018. Intelligent Virtual Assistant knows Your Life. *CoRR* abs/1803.00466 (2018).

[28] Luigi De Russis, Dario Bonino, and Fulvio Corno. 2013. The Smart Home Controller on Your Wrist. In *Proceedings of ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication (UbiComp)*. 8.

[29] Carl Fischer, Kavitha Muthukrishnan, Mike Hazas, and Hans Gellersen. 2008. Ultrasound-aided Pedestrian Dead Reckoning for Indoor Navigation. In *Proceedings of the First ACM International Workshop on Mobile Entity Localization and Tracking in GPS-less Environments (MELT '08)*.

[30] Kevin Fu, Wenyuan Xu, and Chen Yan. 2018. How We Reverse Engineered the Cuban Sonic Weapon Attack. *IEEE Spectrum* (March 2018).

[31] Sidhant Gupta, Daniel Morris, Shwetak Patel, and Desney Tan. 2012. SoundWave: Using the Doppler Effect to Sense Gestures. In *Proceedings of SIGCHI Conference on Human Factors in Computing Systems (CHI).*

[32] Tara John. 2019. Crickets could be behind the Cuba 'sonic attack' mystery, scientists say. CNN. https://www.cnn.com/2019/01/07/health/cuba-sonic-attack-crickets-scli-intl/index.html.

[33] Gary S. Kendall, Christopher Haworth, and Rodrigo F. CÃądiz. 2014. Sound Synthesis with Auditory Distortion Products. *Computer Music Journal* 38 (2014), 5–23. Issue 4.

[34] Junhong Li, Chenghao Wang, Wei Ren, and Jun Ma. 2017. ZnO thin film piezoelectric micromachined microphone with symmetric composite vibrating diaphragm. *Smart Materials and Structures* 26, 5 (2017), 055033.

[35] Sapna Maheshwari. 2018. Hey, Alexa, What Can You Hear? And What Will You Do With It? New York Times.

[36] Robert J Mailloux. 1982. Phased array theory and technology. *Proc. IEEE* 70, 3 (1982), 246–291.

[37] R. Mayrhofer and H. Gellersen. 2007. On the Security of Ultrasound as Out-of-band Channel. In *Proceedings of the 2007 IEEE International Parallel and Distributed Processing Symposium.*

[38] Tim Moynihan. 2016. ALEXA AND GOOGLE HOME RECORD WHAT YOU SAY. BUT WHAT HAPPENS TO THAT DATA? Wired.

[39] Wayne O Olsen. 1998. Average speech levels and spectra in various speaking/listening conditions: A summary of the Pearson, Bennett, & Fidell (1977) report. *American Journal of Audiology* 7, 2 (1998).

[40] Minna Pakanen, Ashley Colley, Jonna Häkkilä, Johan Kildal, and Vuokko Lantz. 2014. Squeezy Bracelet: Designing a Wearable Communication Device for Tactile Interaction. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational (NordiCHI '14).* 305–314.

[41] Daniel Povey, Arnab Ghoshal, Gilles Boulianne, Lukas Burget, Ondrej Glembek, Nagendra Goel, Mirko Hannemann, Petr Motlicek, Yanmin Qian, Petr Schwarz, et al. 2011. The Kaldi speech recognition toolkit. In *IEEE 2011 workshop on automatic speech recognition and understanding.*

[42] Nirupam Roy, Haitham Hassanieh, and Romit Roy Choudhury. 2017. Backdoor: Making microphones hear inaudible sounds. In *Proceedings of ACM MobiSys.*

[43] Nirupam Roy, Sheng Shen, Haitham Hassanieh, and Romit Roy Choudhury. 2018. Inaudible Voice Commands: The Long-Range Attack and Defense. In *Proceedings of the 15th USENIX Symposium on Networked Systems Design and Implementation (NSDI).*

[44] U.S. Occupational Safety and Health Administration (OSHA). 2013. Occupational Safety and Health Administration Technical Manual. https://www.osha.gov/dts/osta/otm/new_noise/#appendixc.

[45] Woon Seob Lee and Seung S. Lee. 2008. Piezoelectric microphone built on circular diaphragm. 144 (06 2008), 367–373.

[46] Liwei Song and Prateek Mittal. 2017. Inaudible Voice Commands. *CoRR* abs/1708.07238 (2017).

[47] Selvaganesh Dharmeswaran Swetank Kumar Saha, Pratham Malik and Dimitrios Koutsonikolas. 2016. Revisiting 802.11 power consumption modeling in smartphones. In *Proc. of WoWMoM.*

[48] Edward J. Wang, Tien-Jui Lee, Alex Mariakakis, Mayank Goel, Sidhant Gupta, and Shwetak N. Patel. 2015. MagnifiSense: Inferring Device Interaction Using Wrist-worn Passive Magneto-inductive Sensors. In *Proceedings of ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp).*

[49] Charlie Wood. 2017. Devices sprout ears: What do Alexa and Siri mean for privacy? Christian Science Monitor.

[50] Candid Wueest. 2017. Everything You Need to Know About the Security of Voice-Activated Smart Speakers. Symantec.

[51] Chenren Xu, Sugang Li, Gang Liu, Yanyong Zhang, Emiliano Miluzzo, Yih-Farn Chen, Jun Li, and Bernhard Firner. 2013. Crowd++: Unsupervised Speaker Count with Smartphones. In *Proceedings of ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp).*

[52] Tuo Yu, Haiming Jin, and Klara Nahrstedt. 2016. WritingHacker: Audio Based Eavesdropping of Handwriting via Mobile Devices. In *Proceedings of ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp).*

[53] Clint Zeagler. 2017. Where to Wear It: Functional, Technical, and Social Considerations in On-body Location for Wearable Technology 20 Years of Designing for Wearability. In *Proceedings of the 2017 ACM International Symposium on Wearable Computers (ISWC '17).*

[54] Guoming Zhang, Chen Yan, Xiaoyu Ji, Tianchen Zhang, Taimin Zhang, and Wenyuan Xu. 2017. DolphinAttack: Inaudible voice commands. In *Proceedings of ACM Conference on Computer and Communications Security (CCS).*