Owlet: Enabling Spatial Information in Ubiquitous Acoustic Devices

Nakul Garg‡
nakul22@umd.edu
University of Maryland College Park

Yang Bai‡
yangbai8@umd.edu
University of Maryland College Park

Nirupam Roy
niruroy@umd.edu
University of Maryland College Park

(‡ Co-primary Student Authors)

ABSTRACT

This paper presents a low-power and miniaturized design for acoustic direction-of-arrival (DoA) estimation and source localization, called *Owlet*. The required aperture, power consumption, and hardware complexity of the traditional array-based spatial sensing techniques make them unsuitable for small and power-constrained IoT devices. Aiming to overcome these fundamental limitations, *Owlet* explores acoustic microstructures for extracting spatial information. It uses a carefully designed 3D-printed metamaterial structure that covers the microphone. The structure embeds a direction-specific signature in the recorded sounds. *Owlet* system learns the directional signatures through a one-time in-lab calibration. The system uses an additional microphone as a reference channel and develops techniques that eliminate environmental variation, making the design robust to noises and multipaths in arbitrary locations of operations. *Owlet* prototype shows 3.6° median error in DoA estimation and $10cm$ median error in source localization while using a $1.5cm \times 1.3cm$ acoustic structure for sensing. The prototype consumes less than 100^{th} of the energy required by a traditional microphone array to achieve similar DoA estimation accuracy. *Owlet* opens up possibilities of low-power sensing through 3D-printed passive structures.

CCS CONCEPTS

• **Computer systems organization** → **Embedded and cyber-physical systems**; **Sensors and actuators**.

KEYWORDS

Low-power sensing; IoT; Acoustic metamaterial; Spatial sensing

ACM Reference Format:
Nakul Garg, Yang Bai, and Nirupam Roy. 2021. Owlet: Enabling Spatial Information in Ubiquitous Acoustic Devices. In *The 19th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys '21), June 24–July 2, 2021, Virtual, WI, USA.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3458864.3467880

MobiSys '21, June 24–July 2, 2021, Virtual, WI, USA

ACM ISBN 978-1-4503-8443-8/21/06…$15.00
https://doi.org/10.1145/3458864.3467880

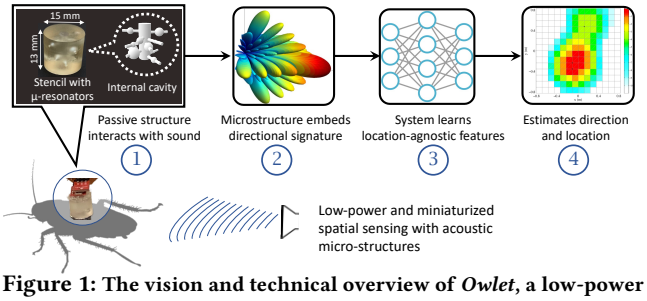

Figure 1: The vision and technical overview of *Owlet*, a low-power and miniaturized system for extracting spatial information from sound. *Owlet* uses acoustic microstructures to embed direction-specific signatures on the recorded sound and develops a learning-based approach for signature recovery and mapping in real-time.

1 INTRODUCTION

Acoustic devices in various forms are becoming ubiquitous in our environments. Besides voice interfaces, a wide range of applications are emerging that explore multiple dimensions of context-awareness and analytics. Applications include indoor activity monitoring using sound [52, 70, 73], health monitoring with acoustic cues [12, 71], speech development and acoustic environment tracking with on-body wearables [24, 66], and many outdoor applications with distributed sensor nodes [3, 8]. With emerging low-power and battery-free solutions [40, 59], it is even possible to continuously collect and process sound on standalone sensing modules scattered in the environment. Spatial analysis of sound and source localization can add new capabilities to such context-aware applications. On the other hand, spatial sensing of sound plays a key role in robotic navigation and situational awareness systems, both in air [2, 26, 27] and underwater [32, 43]. However, traditional techniques for obtaining spatial information of sound require multiple streams of simultaneously recorded sound using an array of microphones – an energy-hungry hardware requirement often difficult to meet on the stand-alone sensing modules. In this paper, we seek to develop an acoustic sensing system that can enable spatial information processing in power-constrained ubiquitous computing devices with small form factors.

Sensing spatial features of sound, such as direction-of-arrival (DoA) or source location, requires sampling the wave in space using an array of microphones. Given the conventional DoA estimation algorithms fundamentally depend on this spatial sampling model, the dimension of the array and number of microphones are crucial

for their performance. According to the sampling theorem [37], a separation equal to the signal's half-wavelength ($\lambda/2$) between microphones in a linear array is considered ideal for DoA estimation. Moreover, the angular resolution (in terms of the inverse of the Half Power Beam-width) of the DoA is proportional to the total length of the array. Therefore, the traditional algorithms require a large hardware setup to achieve fine-grained resolution for DoA estimation. Moreover, arrays often require simultaneous samples from the microphones, which increases power consumption and hardware complexity proportional to the number of microphones. Despite the tremendous proliferation of acoustic devices in ubiquitous computing, the hardware and power requirements, and limitation in form-factor hinder solutions requiring high-resolution spatial information. In this paper, we seek to develop an alternative method for spatial signal processing. We break away from the spatio-temporal sampling model and explore the interaction of the waves with structures for a low-power, low-complexity, and miniaturized solution.

The use of acoustic structures for directional hearing is common in nature. The symmetric left and right ears in most mammals, including humans, resemble a two-element array for the directional processing of sound. However, biophysical studies show that these species use cues from sound's interaction with the three-dimensional structure of their heads for fine-grained localization of the source [10]. In most owl species, the two ears are asymmetrical in their positioning in both horizontal and vertical planes [22]. This structural diversity helps them to precisely localize low-frequency noises, which is not possible through symmetric sensing given the separation between their ears. Surprisingly, some insects with body dimensions much smaller than a tenth of the wavelength of the relevant sound, achieve a localization performance similar to that of mammals [58]. For instance, a small grasshopper with a body width of $3mm$, a fraction of the target sound's wavelength, can sense precise location information. The secret lies in the asymmetric orientation and structural formations that lead to a different response based on the direction of the incoming sound. The sensory system and the neural network in these species have evolved to relate these responses to the direction of arrival of the sound. We take inspiration from such structure-aided hearing techniques to design a DoA estimation system for power-constrained miniaturized devices.

In this paper, we present a design and prototype of an acoustic localization system that introduces acoustic structures around a microphone to embed directional cues. Acoustic wave interacts with physical structures on its propagation path and as a result, the wave field is transformed. Such behavior of waves is clearly observable at a large scale in room acoustics, where the same sound appears different due to the shape, size, and object placement in the room. We show that it is possible to manipulate sound waves using a small 3D-printed acoustic structure, such that it leaves a unique signature to the passing sound. When we place a microphone inside that structure of few centimeters in dimension, it records sounds carrying that signature. If designed carefully, this structure can embed distinct signatures for sounds coming from different angles even at the resolution of a few degrees. Our system can detect these signatures to identify the DoA of the recorded signal. We call this system *Owlet*, named after the bird with marked auditory finesse.

We are not the first to observe the opportunity in environmental variations of sound fields. Past work has explored localization by fingerprinting multipath environments and analyzing nearby reflections [64]. Probably closest to our work is [18] that places objects in a 60×60 cm space with a microphone at the center. This work shows that the sound scattered by the nearby objects holds directional cues and can be processed to find its direction of arrival. The concept of *Owlet* fundamentally similar to these studies but differs in two important ways. We focus on developing a small centimeter-scale sensing system that can potentially be used on resource-constrained robots or as a ubiquitous sensing solution. *Owlet* prototype has shown angular resolution similar or better than the past work with a tiny $1.5cm \times 1.3cm$ sensor. Secondly, we address the issue of system's robustness to environmental changes. *Owlet* is designed to perform beyond an anechoic chamber or controlled lab environment and eliminate the requirement of location-specific training data.

One of the main challenges faced by *Owlet* is harnessing multipath diversity in a small form factor. Due to the large wavelength of low-frequency acoustic signal, it requires reflectors of comparable size to achieve such diversity which is directly related to the achievable spatial resolution of the system. We work around this limitation by developing a diffraction-based technique, as opposed to the reflection-based approach, to design miniature acoustic structures. The idea is based on the observation that when sound passes through a small aperture, it undergoes diffraction and appears as an independent sound source. We exploit this diffraction property to design a 3D-printed cylindrical cover, called stencil, for the microphone. These stencils carry optimally coded patterns of holes on the surface that create a complex but predictable multipath interference inside the structure. The interference pattern carries a signature of the direction of arrival of the recorded sound. We include principles of metamaterial designs in the stencil for improved angular diversity. *Owlet* system learns these signatures through a one-time calibration process and maps them to the DoA of sound at run-time.

The other major challenge is to make the design robust to environmental changes that can influence the incoming sound in arbitrary ways. In other words, for the system to be useful in practice, it should be able to function in arbitrary environments while calibrated only once during manufacturing. As mentioned before, room acoustics can influence a sound field and make the mapping of direction-specific signature fails. *Owlet* introduces a reference microphone to the design and takes a communication theoretic approach to eliminate the transient multipath effects during signature generation and mapping. This technique makes *Owlet* robust to environmental change and suitable for real-world applications.

This paper explores acoustic structures as passive components in new types of low-power and miniaturized solutions for ubiquitous sensing. Representative applications include wearable devices for acoustic environment sensing toward speech development

assessment in infants [29, 51] or personal analytics [21, 23] that require the direction of the sound. Navigation techniques in SWaP-constrained [14, 45] in-air and underwater mobile robots can be benefitted from spatial sensing with *Owlet*. Moreover, *Owlet* can enable direction estimation and localization in energy harvesting systems which is difficult to achieve with traditional microphone arrays. Figure 1 presents the broader vision and technical overview of our work. While several opportunities for applications open up, this paper focuses on developing the core capabilities and assessing the limits of the systems. We have made the following three specific contributions at the current stage of this project:

- A novel method of using passive elements for directional sensing that leads to low-power, low-complexity, and miniaturized system for acoustic localization. Developed sensing and signal processing ensure robust DoA estimation in a diverse environment with a one-time in-lab calibration. The system has achieved 3.6° of median error which is comparable to existing microphone array-based solutions with a fraction of its power and space requirements.
- A replicable method for designing and 3D-printing optimal acoustic structures to encode incoming sounds with directional cues. It presents a method for sound field shaping with controlled diffraction in small physical metamaterial structures.
- A hardware/software prototype of the entire system for the community to reproduce, evaluate, and build on the *Owlet* system.

Next, we elaborate on the core intuition, system design, and key findings of this project.

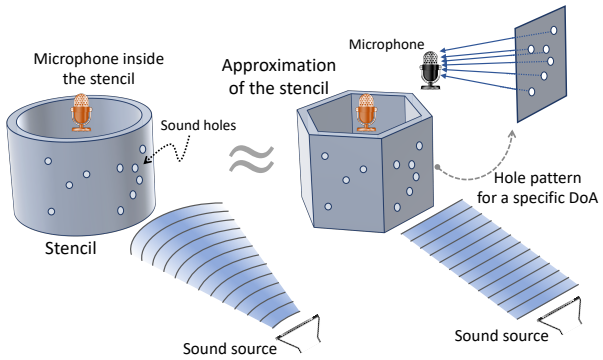

Figure 2: The concept of using a stencil with direction-specific hole patterns and microstructures for passive filtering of the incoming sound. The stencil embeds a directional response to the recorded signals.

2 CORE INTUITIONS AND PRIMERS

Fundamentally, we aim to design a controlled environment around the microphone such that the recorded signal contains a unique 'direction-specific' channel impulse response. This impulse response can be extracted from the microphone recording and will serve as a signature of the sound's angle of arrival. While a regular room environment or larger objects near a microphone are known to create a diverse multipath effect to add direction-specific response to the signal, we envision achieving much fine-grained diversity with a compact form factor by combining the concepts of diffraction, interference, and structural resonance. To this end, we design a porous cap for the microphone, called stencil. It has particular hole patterns at different sides as shown in Figure 2. Sound coming at a specific angle pass through the unique patterns of holes and combines at the microphone. The holes on the stencils are connected to microstructures of different parameters leading to a unique frequency response.

The stencil forms a metamaterial with internal microstructures that naturally modulates incoming sound to introduce a unique directional signature. As the impact of the microstructures depends on the frequencies of the sound, the signature is basically a vector of complex gains, G_θ, of the frequency response. The concept is explained in Figure 3.

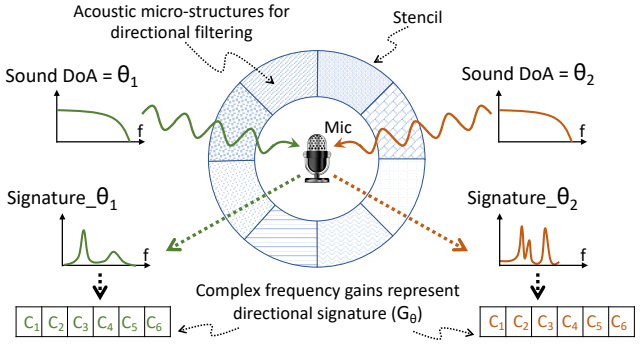

Figure 3: The concept of passive directional filtering using a stencil of acoustic microstructure. The stencil embeds a directional signature to the recorded sound unique to its direction of arrival (DoA). The spectrum of complex gains represents the signature for further computation.

2.1 Metamaterials for passive filtering

Sound frequencies get amplified or attenuated when it interacts with structures. At a large scale, multipath reflections show such variations in frequencies due to constructive and destructive interferences. While reflections can create directional signatures in sounds, it requires large form factors comparable to the wavelength of sound. Given *Owlet* focuses on the low audible frequencies, wavelengths are large, and it would require reflectors almost a half of a meter in size. To miniaturize the acoustic structure for passive filtering of the passing sound we use concepts of metamaterials. Metamaterials are specially designed structures with assemblies of substructures that give new property to the material. In designing our metamaterial stencil, we employed principles of (a) diffraction, (b) capillary effects, and (c) structural resonance.

(a) Diffraction: Waves, when encountering the edge of an obstacle in their path, tend to bend or deflect around it. This phenomenon is called diffraction. Diffraction leads to an interesting property of sound waves when it passes through a hole [56]. If the aperture of the hole is small compared to the wavelength of the sound, the wave diffracts at the edge of the hole and the hole behaves as a virtual point source of that sound. If the receiver is on the other side of a barrier having multiple such holes, it observes multipath-like environment with multiple virtual sources discussed earlier. Interaction between signals from these virtual sources creates a pattern of constructive and destructive interferences depending both on the location of the receiver and the frequency

of the signal. We use diverse patterns of small sound holes on the stencil to create a multipath effect at the inside microphone within a small form factor.

(b) Capillary effect: Acoustic impedance varies when sound propagates through small capillary tubes [54]. Moreover, the length and cross section of the tubes impacts the speed of passing sounds. We implement capillary tubes of various shapes in the stencil to emulate the effect of phase differences for sound paths. This leads to prominent diversity in frequency spectrum despite small separations between sound holes.

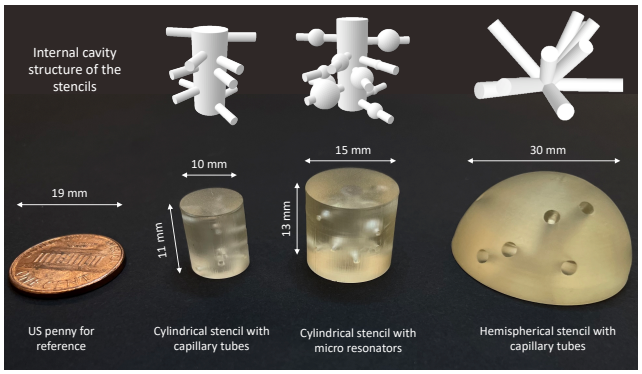

Figure 4: Different types of metamaterial stencils used in our experiments.

(c) Structural resonance: Certain sound frequencies get amplified when oscillating air pressures meet cavities on their way [69]. This property is called Helmholtz resonance and it is commonly observed in whistling bottles. We designed millimeter scale Helmholtz resonators embedded in the stencil and connected to the sound holes. We vary the shape of these tiny structural resonators to generate arbitrary resonance effects at different frequencies.

Figure 4 shows 3D printed stencils with embedded microstructures for directional filtering. In Figure 5 we show the effect of microstructure stencil in improving angular diversity of the sensor. It compares the angular variation of the amplitude of the $7kHz$ tone on the *Owlet* microphone setup with and without the stencil. Figure 6 shows diversity in corresponding direction-specific frequency responses of these stencils.

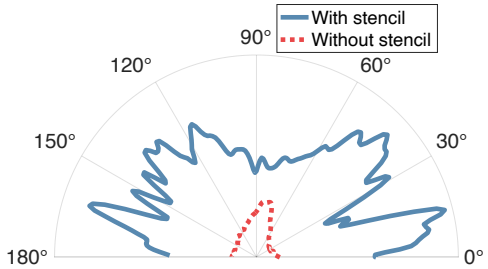

Figure 5: Angular diversity of the microphone with and without the microstructure stencil.

3 SYSTEM DESIGN

The system design focuses on two main tasks: (a) developing an optimal stencil structure that offers maximum angular diversity and (b) developing computing techniques to find DoAs from the recorded signal. Naturally, the accuracy of the system directly depends on the diversity introduced by the stencil. Our algorithms optimize this design by considering parameters of wave propagation around small structures in simulation and then 3D prints it for experiments. Before we go into the details of the stencil design, we explain our processing and DoA estimation techniques which also serve as an overview of the entire system.

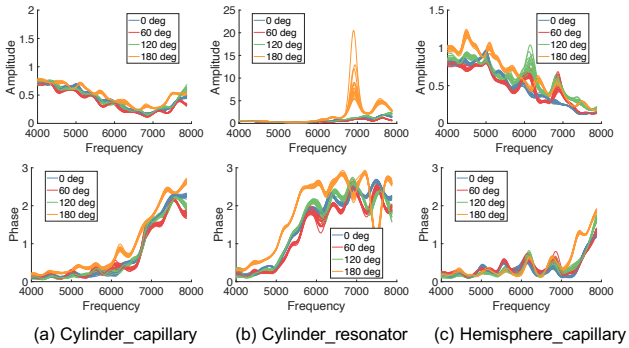

Figure 6: Comparison of the diversity in frequency responses (amplitude and phase) of the three types of metamaterial stencils.

3.1 Processing for DoA estimation

In rather simplistic terms, *Owlet*'s DoA estimation technique is a two-step process. First, we create a table of direction-specific signatures G_θ of the stencil by sending known signals from various directions. We perform this signature generation by playing a wideband sound signal from a specific direction and recording the signal with a microphone having the stencil cap over it. This is a one time in-lab calibration, similar to the calibration done for commercial-grade microphone arrays. The second step is performed at the run-time when the system is being used for DoA estimations. Here we process the incoming signal to extract the signature introduced by the stencil, $h_{stencil}$, and then look it up in the table of pre-collected signatures that maps it to a specific DoA. In practice, we train a deep learning model with variations of the signature table and use the pre-processed signal at run-time to get predicted DoA from the model. Note that the signature extraction from the real-world signal is a crucial part of the processing and it meets two challenges: (i) estimating $h_{stencil}$ by separating it from the frequency diversity of the source signal and (ii) eliminating additional distortions added to the signal by the environmental multipaths. We explain the technique for signature extraction for eliminating the source dependency, followed by a technique to deal with the environmental multipath in arbitrary locations.

3.2 Eliminating source signal dependency

The signal recorded by the microphone inside the stencil is basically the source signal distorted by the directional-specific response of the stencil. If we assume no environmental effect on the source signal $X(\omega)$, the signal received by the inside microphone $Y_{in}(\omega)$ can be expressed as a multiplication between this source signal and the stencil's response $H_{stencil}$ in frequency domain:

$$Y_{in}(\omega) = X(\omega)H_{stencil} \quad (1)$$

Therefore, when the source signal $X(\omega)$ is known we can obtain the stencil's response by simply calculating $\frac{Y_{in}(\omega)}{X(\omega)}$. The source signal can be user-defined and known for some applications, like in navigations where the robot localizes itself by finding DoA of a known control signal. However, DoA estimation is useful in many other applications including localizing an ambient noise source or finding the user's direction from spoken words. In such scenarios, the source signal is unknown to the system, and it is difficult to separate $H_{stencil}$ from the arbitrary source signal. We eliminate this problem by introducing a secondary microphone which is placed outside the stencil. The incoming sound is recorded simultaneously by these two microphones, but the outside microphone's recording is unaffected by $H_{stencil}$. Note that, unlike in microphone arrays, the secondary microphone can be placed arbitrarily close to the primary microphone. Figure 7 shows the physical design and the realistic signal model of the system.

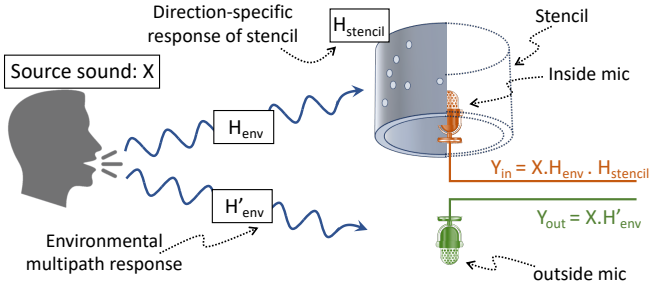

Figure 7: The two-microphone model for eliminating source and environmental dependency.

Consider the channel frequency responses to the inside and outside microphone are H_{env} and H'_{env} respectively. These channel responses manifest the effects of the multipath signal propagation from the source to the microphones and the signal's reflections from nearby objects. The presence of the stencil around the internal microphone introduces additional modulation to the recorded signal, represented by the frequency response $H_{stencil}$. Considering the linearity of the channels, the signal recorded by the inside microphone will experience both the impulse responses as shown in Figure 7. Therefore, the signals recorded simultaneously by these microphones, $Y_{in}(\omega)$ and $Y_{out}(\omega)$, can be formulated as the following equations. The source sound is $X(\omega)$ and independent noise at the two channels are $N(\omega)$ and $N'(\omega)$ at the frequency ω.

$$\begin{aligned} Y_{in}(\omega) &= X(\omega)H_{env}H_{stencil} + N(\omega) \\ Y_{out}(\omega) &= X(\omega)H'_{env} + N'(\omega) \end{aligned} \tag{2}$$

If we divide $Y_{in}(\omega)$ by $Y_{out}(\omega)$, it successfully eliminates the dependency on the source signal. However, the environmental dependency remains in the form of $\frac{H_{env}}{H'_{env}}$.

$$\frac{Y_{in}(\omega)}{Y_{out}(\omega)} = H_{stencil}\frac{H_{env}}{H'_{env}} + N''(\omega), \tag{3}$$

Here, $N''(\omega) \ll H_{stencil}\frac{H_{env}}{H'_{env}}$. This means the stencil calibration process, or the deep learning module training process has to be trained for all locations in the target environment to capture the environmental dependency to make the angle prediction effective. Such a system may be applicable for scenarios where the locations of the sound sources and the sensing modules are predefined. For instance, when acoustic localization is used to track objects on a conveyor belt or on a track. However, for most practical scenarios the location of the sound source is unknown, and it will require collecting data from virtually every point in the scene and train the prediction module – leading to an impractical solution. Next, we explain our technique to eliminate this location dependency. With this technique *Owlet* can function with one round of in-lab calibration of the stencil and does not require collecting any calibration data at the target environment.

3.3 Eliminating environmental dependency

This final stage of the technique is based on the observation that despite diverse and unpredictable nature of the environmental channel responses H_{env} and H'_{env}, the ratio of the channels, $H_{ratio} = \frac{H_{env}}{H'_{env}}$ is bounded when the microphones are closely placed. This idea can be intuitively understood by first analyzing the reason for diversity in environmental response H_{env}. The sound wave reflects off various objects in the environment after leaving from the source. These reflections follow paths of varying lengths to get superimposed at the recording microphone along with the direct line-of-sight path. The diversity in the path distances creates time delays in the reflected components leading to a unique response of the environment. Therefore, two microphones, even when recording the same signal, can observe different responses as the path lengths of the reflections are different. However, if the locations of the microphones are close to each other, these path differences of reflections are bounded and at one extreme when two microphones are exactly collocated, they will observe same environmental response. Therefore, $H_{ratio} = \frac{H_{env}}{H'_{env}}$ has a narrow distribution of values for each frequency in the response when two microphones are a few centimeters apart from each other. We obtained the probability distributions from simulated ray tracing and real-world experiments.

Once the distributions of H_{ratio} is known and $H_{stencil}$ is collected through the calibration stage, we generate a synthetic training data for $H_{ratio}H_{stencil}$ drawing from the distribution and use it for training the deep learning module. This process can train our angle prediction module robust to environmental variations at run-time without requiring real-world sound traces for training. Interestingly, if the dimension of the target environment and locations of the major reflectors are known, the synthetic training data can be customized to that environment. This customization reduces the time for convergence during training and improves prediction accuracy.

The run-time processing now requires to extract $H_{ratio}H_{stencil}$ from the two channels of sound $Y_{in}(\omega)$ and $Y_{out}(\omega)$. We improve this process by employing a recursive least square (RLS) adaptive filter [1] in system identification mode. The adaptive filter takes advantage of the uncorrelated Gaussian noise in the recorded signals to estimate $H_{ratio}H_{stencil}$ by minimizing the following error term with gradient descent.

$$e(\omega) = Y_{in}(\omega) - Y_{out}(\omega)\frac{H_{env}H_{stencil}}{H_{env'}} \tag{4}$$

3.4 Synthetic training for deep learning

We use the synthetic channel response mentioned in the previous section to introduce training diversity to the neural network model. To this end, we calculate $H_{ratio}H_{stencil}$ by simulating different room environments and diverse placements of the source and the microphones. We use the distribution of these channel values to generate additional $H_{ratio}H_{stencil}$ data and feed them to deep learning architecture for training. Vectors of 400 equally separated frequencies between $0 - 8kHz$ represent the discrete spectrum of $H_{ratio}H_{stencil}$ for each angle. Instead of using the complex vectors, we separately calculate the amplitude- and phase-spectrum from the channel responses for the training data.

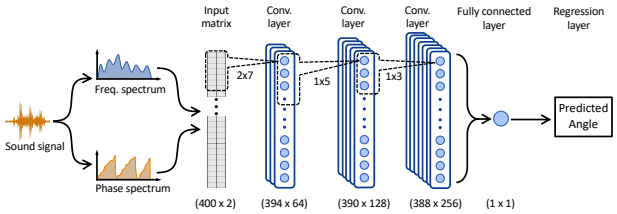

Figure 8: The architecture of the proposed CNN model.

We select a Convolutional Neural Network (CNN)-based regression model for the DoA estimation. CNN is known to have superior performance in environmental sound processing [55] and low-latency operation for its reduced set of parameters. We use a one-dimensional CNN model with three hidden convolutional layers followed by a fully connected layer and an output regression layer. The convolution layers have 64, 128, and 256 filter maps with 2×7, 1×5, and 1×3 filter sizes respectively. The regression layer computes the half-mean-squared-error loss for angle estimation. We customize the loss function according to the target range and resolution of the directional angles. In this model, we use ReLU (Rectified Linear Unit) activation function and add batch normalization layers between the convolution layers to speed up the training process. We apply stochastic gradient descent as the optimizer. The model is trained for 100 epochs at a learning rate of $1e^{-6}$. The block diagram of the CNN architecture is shown in Figure 8.

Besides the regression model, we develop a CNN classification model for evaluation and comparison, as detailed in section §5.8. For this model, we mostly follow the regression architecture but design a fully connected layer with length 360, one for each angle, followed by a Softmax layer and a classification layer.

3.5 Optimizing 3D stencil design

The accuracy of our proposed system depends on the diversity of the frequency gain pattern for different angles. Our feasibility study with random hole distribution on the stencil cap shows reasonable diversity of the gain pattern to distinguish sound direction with a median error of 7°. The resolution of detected DoA is not uniform across all angles, meaning the system's accuracy to detect signal from certain directions are poor compared to the other. We trace back this problem to the suboptimal distribution of holes and microstructures on the stencil that result in similar gain patterns for multiple directions. We address this problem through systematic development of the 3D stencil design which is optimized to guarantee a minimum DoA detection resolution in any direction.

An ideal stencil cap should provide maximum diversity of the frequency gain pattern for each possible DoA in the recorded signal. This problem of achieving maximum diversity is analogous to the information theoretic problem of designing maximally diverse code sequences. Consider a frequency gain pattern (G_θ), associated to a specific angle θ, to be a codeword. We want to design a set of N codewords which are maximally distant from each other (e.g., maximum possible Euclidean distance between all pair of codewords). Here the number N defines the DoA detection resolution, $\Delta\theta = \frac{2\pi}{N}$. In an initial attempt, we aim to design such codewords of lengths equal to a set of discrete frequencies and then use them as guidelines to generate a set of desired gain patterns (G_θ). The next step is to map G_θ to a pattern of pinholes on the surface of the stencil cap at angle θ. Given the number of holes in stencil, N, distance between microphone and holes, r_n, distance between microphone and stencil, D, and wavelength of the wave, λ; Equation 5 gives the resulting value, $u(\lambda)$, at the microphone. In other words, $u(\lambda)$ is the result of superposition of all the waves coming from the holes. Note that for different wavelengths the equation gives overdetermined systems of equations to solve for the hole patterns. We can derive an approximate solution to determine the optimal design of the hole arrangements. Unfortunately, this analytical approach quickly becomes intractable even for a moderate number of holes (> 10) on a three-dimensional stencil. The other limitation that renders this approach unsuccessful is the fundamental difference between our model and the actual wave field near small objects. We elaborate on this wave property before presenting our simulation-based approach for the optimal stencil design.

$$u(\lambda) = \sum_{n=1}^{N} \frac{D}{j\lambda r_n^2} e^{\frac{j2\pi r_n}{\lambda}} \tag{5}$$

Behavior of wave fields near the stencil: In our stencil model we implicitly assume that the incoming sound wave only passes through the holes on the side of the cylinder that directly faces the wavefront. Basically, we approximate the cylinder as a N-gonal prism, as shown in Figure 2, such that the unique hole locations on each face can generate a particular frequency gain pattern at the microphone. This approximation holds for a large object with a diameter more than 10 times of the signal's wavelength [72]. However, as we aim to design a miniaturized interference shaping structure, this wave propagation model differs significantly for the dimension of our stencil cap. Just like at the openings of the holes on the surface, wave fronts diffract by the outer surface of the small stencil and wrap around covering almost the entire cap as shown in Figure 9.

We verified this phenomenon using a cylindrical shape with a pinhole on one side (Figure 10(a)) and measured the sound pressure at the microphone kept at its center. Figure 10(b) shows significantly high sound pressure even when the hole is more than 90° off from the direction of sound source, indicating the bending of sound waves. Interestingly, it shows a high sound pressure level diametrically opposite to the incoming direction of sound, a result of

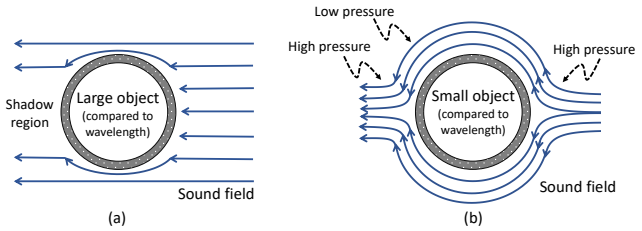

Figure 9: The behavior of sound field at the outer surface of an obstacle. (a) When the object's size is much larger compared to the wavelength of the sound, the obstacle creates a shadow region. (b) When the object's size is comparable to the wavelength of the sound, the wave diffracts around the object creating high-pressure at a larger region of the surface. It also creates a high-pressure region directly opposite to the sound's directions where sound fields from the top and bottom sides meet.

merging fields from two sides of the cylinder. This angular intensity variation at the surface of the stencil is dependent on the sound frequency influencing the received frequency gain pattern.

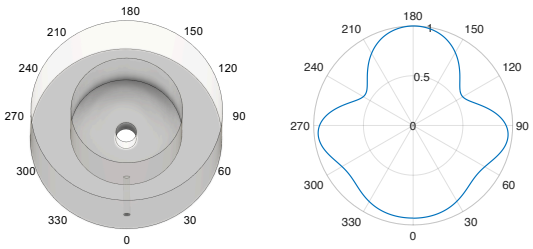

Figure 10: (a) A one-hole stencil to measure surface pressure levels. (b) Sound amplitude at different angles from the sound's direction of arrival.

We modify our model and take a forward simulation-based technique to find the optimal design for the stencil cap. Now, instead of tracing back a hole pattern from a desired frequency gain pattern, we select the best hole-pattern from a set of random sets. This Monte Carlo simulation-based approach takes repeated random sampling of the stencil pattern and then analytically simulates frequency-gain pattern for each of these variations for all directions of the DoA (360 source locations at 1° of separation). We then optimize for diversity of the gain-patterns for all directions as mentioned in the steps below. The following process ensures the selected hole-pattern is close to the global optimal solution given the cap size and other design parameters.

(1) Random stencil pattern generation:
The diversity of the directional gain-pattern is directly related to the multipath diversity created by the hole pattern on the stencil. We fixed the outer and inner diameters and the height of the cylinder for the simulation. The diameter of the pinholes is set to 2 mm. Then we generate random patterns of the pinhole locations on the side of the cylinder. However, uniform sampling of locations for the pinholes does not ensure minimum separation between the holes, where separation of half of the maximum wavelength is necessary to have an individual impact on the frequency gain-pattern. We introduce this constraint by modifying the *Fast Poisson Disc* sampling method [9]. At each iteration, the Poisson Disc method generates a 2-dimensional location for the hole, starting from a few existing seed locations, on the flattened surface of the cylinder. The location of the next hole is chosen randomly from the region within a circular annulus of radius 3 mm, ensuring the minimum separation. We randomly vary the width of the annulus at each iteration to introduce diversity in the hole-pattern.

(2) Estimating frequency gain-patterns:
The algorithm computes the frequency dependent gain-pattern for each stencil generated in the previous step using Equation 5. The path differences between the holes and the microphone are calculated considering sound's diffraction at the outer surface of the cylinder. We consider 400 equally separated frequencies between 0-8kHz to have a 400-point complex gain-pattern for each of the 360 angles of source location. We apply the amplitude and phase corrections due to the diffraction of sound waves around the surface of the cylinder as described earlier (Figure 10). After this step we have 360 400-point gain-patterns to be used in the next step.

(3) Assessing the diversity of gain-patterns:
Next, we measure the diversity of the gain-patterns using the all-pair Euclidean distance as the metric, called *chord-distance*. Two distinguishable gain-patterns will show higher values for chord-distance compared to the two similar patterns. We use this metric for maximin decision criteria in the final step.

(4) Stopping criteria and selecting the best stencil:
At each iteration with a new stencil pattern the algorithm records the minimum of the all-pair chord-distance derived in the previous step. The stopping criteria of the iterations is reached when the distribution of the metric fits a gaussian curve. We then pick the maximum value of the chord-distance and select the corresponding stencil design for fabrication. Figure 11 compares the diversity in the frequency gain-pattern of an optimal and a sub-optimal stencil.

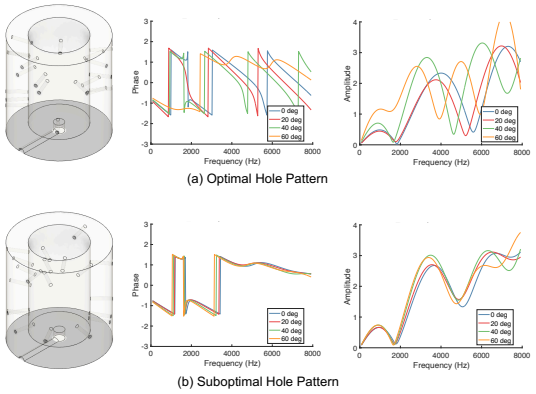

Figure 11: Comparison of diversity in phase and amplitude patterns for an optimal and a sub-optimal design of the stencil.

4 PROTOTYPE DEVELOPMENT

4.1 3D-printing stencil caps

We first run our optimization algorithm on Matlab to obtain a stencil design. Next, we use the Autodesk Fusion 360 Python API [4, 6] to generate the 3D model of the stencil. The script takes

the design parameters of the stencil as input, builds the structure including internal substructures and cavities, and adds the holes on the surface. Finally, we export the STL model of the stencil and slice it for 3D printing. We used the Elegoo Mars photocuring 3D printer[19] to print the stencils. We use an ultraviolet light-curable resin with 1.195 g/cm^3 density that solidifies when exposed to the light of 405nm wavelength. Compared with jetting-based printing, it provides a high resolution and smooth finish which is ideal for the tiny sub-structures on the stencil. More importantly, photocuring method leads to dense surfaces and makes the acoustic behavior of the stencil predictable [74].

4.2 Calibration and data collection

We first generate a wideband calibration signal on Matlab and export it to an '.arb' file which is loaded to the Keysight Waveform Generator [33]. We used two off-the-shelf speakers along with a 40W dual channel amplifier to transmit this signal. We used an external wired channel to trigger the waveform generator for precise time synchronization. The stencil was attached to a stepper motor and an arduino [5]. We rotated the stencil at 1° steps ranging from 0° to 360° and recorded the calibration signal. We used omnidirectional ADMP401 MEMS microphones [15] sampled at 16 kHz. The collected data was processed offline on a computer.

5 EVALUATION

We aim to assess the performance of our microstructure-guided spatial sensing technique. To this end, we implement a prototype of *Owlet* and perform experiments in several indoor and outdoor settings and under various acoustic environments. We use traditional uniform linear microphone arrays (ULA) of various sizes for the baseline performance comparison and benchmarking the energy consumption. Next, we elaborate on the experimental setup, followed by the evaluation results.

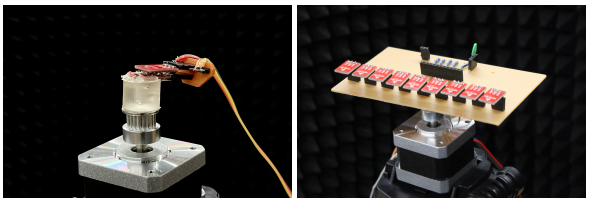

Figure 12: The *Owlet* prototype used in the evaluation experiment (left) and a 9-element uniform linear microphone array used as baseline for comparison(right). The array is 12cm wide, where *Owlet* is significantly smaller measuring less than 2cm in its largest dimension.

5.1 Evaluation setup and results summary

In the *Owlet* prototype, we use a 3D-printed stencil and two microphones placed on the top of each other facing in opposite directions. The separation between these microphones is 4mm and opening of one of these microphones is covered with the stencil. We also developed a 9-element linear microphone array with 1.3cm separation between the elements and each of the microphones are sampled simultaneously using a multi-channel DAQ system [34]. Figure 12 shows the sensor font-ends of the *Owlet* and the ULA. We used omnidirectional ADMP401 MEMS microphones [15] sampled at 16kHz for both *Owlet* and the ULAs. The collected data is processed offline using Matlab scripts on a computer. The transmitted sound sources include multi-frequency wideband sounds, white noise, drone sounds, and car engine noises. Otherwise mentioned, the sound source is a multi-frequency wideband signal, the default noise level is 40 dB SPL, the default distance between sound source and microphones is 3ft, elevation angle is 0°, and the size of stencil is 1.5cm × 1.3cm with internal capillary tubes and structural resonator cavities. We perform our experiments in several representative environments, such as indoor laboratory, lobby, and outdoor, as shown in Figure 13. Note that the recorded sound source is not influenced during our structure-guided DoA estimation techniques, since the secondary microphone is placed outside of the stencil and can perfectly record the sound source.

System	Prototype cost	Size	Error	Energy
Owlet	$15	1.9cm	3.6°	16.7mJ
9-element array	$70	11.4cm	4°	2078mJ

Table 1: Comparison of prototype cost, size, median error, and energy consumption of Owlet with a microphone array.

Figure 13: Various locations for system evaluations: (a) indoor laboratory, (b) indoor lobby, (c) outdoor.

Summary: Figure 14 summarizes the overall performance of *Owlet* in comparison to the traditional ULA based DoA estimation technique. As elaborated later in this section, *Owlet* outperforms even a 9-element ULA running the standard multiple signal classification (MUSIC) algorithm for direction estimation, while consuming a fraction of the energy required by the array. Table 1 presents a comparison of estimated manufacturing costs of these prototypes in the lab, their sizes, the median DoA estimation errors, and the energy requirements.

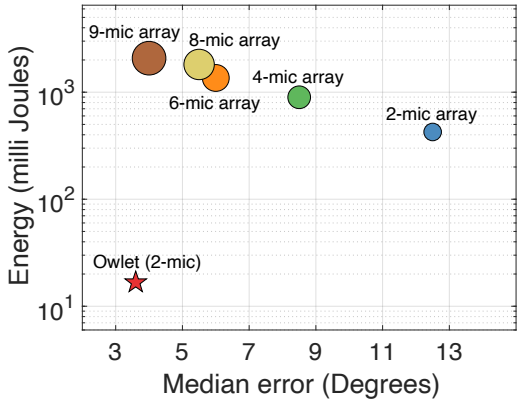

Figure 14: Overall performance of the *Owlet* system compared to the traditional microphone arrays of various sizes. *Owlet* requires 100× less energy than the state-of-the-art array systems while achieving better accuracy than a 9-element array.

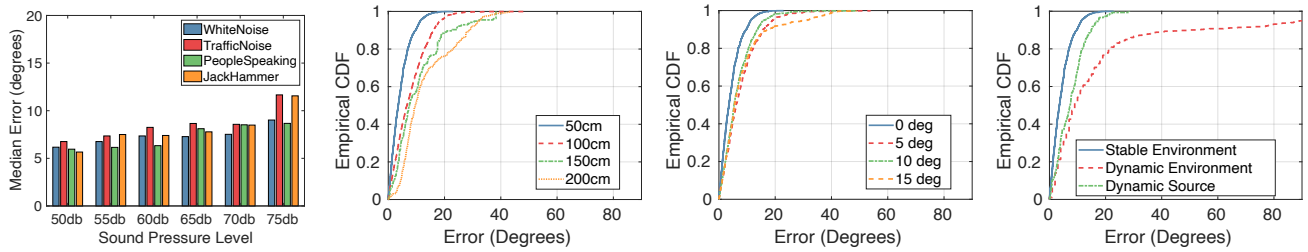

Figure 15: Performance under external conditions: (a) The impact of varying types and loudness levels of ambient noise on the median DoA estimation error. (b) The CDF of errors when the sound source is located at varying distances from the receiver. (c) The CDF plot of estimation error for different elevation angles or the vertical positions of the sound source. (d) The CDF plots of errors that show the impact of dynamic movements in the environment.

5.2 Impacts of external conditions

We evaluated the performance of our prototype under various adversarial conditions. We present the results below.

(a) Ambient noise: The regular noise level at the locations of experiments is around 40 dB of sound pressure level (SPL). We introduced four different kinds of noise with distinct spectral properties: (i) white noise, (ii) traffic sound, (iii) human conversation, and (iv) sound of machineries like jackhammer. We played these noise sounds from three speakers from different angles and at different levels of loudness near the receiver while performing the direction estimation. The loudness of the source sound used for DoA was 60 dB SPL, a loudness comparable to natural conversions. Figure 15(a) shows the median error of *Owlet*'s performance under these experiments remains stable for a wide range of noise loudness.

(b) Distance from the receiver: We measure the performance while the sound source is placed at various distances from the receiver. Figure 15(b) shows the median error for DoA estimation for these experiments. The error is mainly dominated by the change in signal-to-noise ratio at the receiver due to increasing distance. The intensity of the sound source was kept constant despite the location of the source. When we change the model to maintain a fixed signal loudness at the receiver, varying distance shows limited impact.

(c) Elevation angle: Current prototype of *Owlet* limits its direction estimation to the azimuth angles (directions on the horizontal plane) only. Ideally, azimuth-only DoA estimation system should not be affected by the elevation angle (i.e., vertical location) of the sound source. However, in practice microphones are not fully omni-directional and therefore regular microphone-array based DoA systems perform correctly for a certain limit on the elevation angle of the source. In addition to the microphone's limited response in the vertical plane, *Owlet* has pinhole patterns on the stencil that may be projected differently on the microphone. We evaluated the impacts of elevation angle by increasing the vertical distance of the sound source, while the horizontal distance is fixed at 150 cm. Results in Figure 15(c) show *Owlet*'s performance does not vary significantly when the vertical location of the source is up to 15 cm from the center.

(d) Dynamic multipath: *Owlet* is designed to mitigate the effects of the environmental multipath. We have evaluated this feature by changing the test location in previous experiments. We further evaluate it by moving the sound source during the test to add additional disturbance to the environment. The median error remains within 7° for this test. Next, we introduce moving subjects near the setup so that the multipath environment changes during the experiment. The median error is close to 9° when three people keep walking within 3 meters from the sensor. Figure 15(d) shows the CDF of the error for these experiments along with the CDF for stable multipath environments. Figure 16 shows *Owlet*'s performance for tracking the sound source while the source is moving near the sensor.

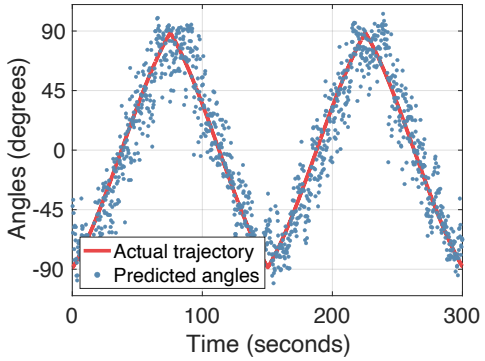

Figure 16: The performance of sound tracking while the source is constantly moving near the sensor. The movement of the source creates a dynamic multipaths scenario.

5.3 Performance in different environments

We evaluated *Owlet* in several represented indoor environments, such as indoor laboratory, lobby, and open-air outdoor places as shown in Figure 13. To make our model robust in diverse environments, we train the deep learning model using synthetic room impulse responses as mentioned in Section §3.4. Figure 17(a) shows the DoA estimation performance in multiple locations of these environments. The median error is within 4° and 90^{th} percentile error of less than 10°. This result shows *Owlet*'s ability to function in unknown environments with a one-time calibration during the development of the prototype.

5.4 Impact of different sound sources

In this experiment we evaluate the system's DoA estimation performance for parallel frequency signal and other types of signal

sources. We evaluate the system's DoA estimation performance for different types of sounds. These signals are different in their active bandwidths, frequency spectrums, and loudness. Figure 17(b) shows comparable performance across different sounds.

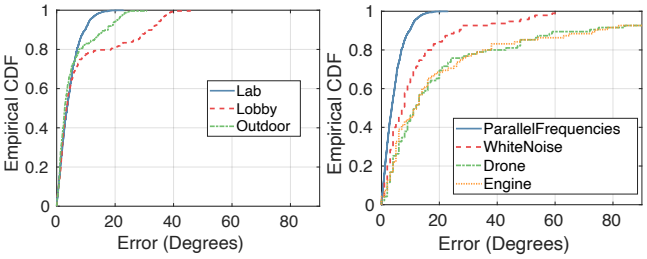

Figure 17: The CDF of median error for (a) different environments and (b) different types of sound sources.

5.5 Performance in known environment

Owlet's synthetic training data generation process can be customized if the geometry of the target location is known. We evaluate this feature by generating training data according to the test location. Figure 18 shows *Owlet*'s overall performance for estimating signals' direction of arrival. In this experiment we transmitted signals from a speaker placing it at various angles with respect to the *Owlet* system. The ground truth DoAs of the signal covered the 0−180° angles in front of the receiver with 1° separation between locations. Note that, unlike microphone arrays, *Owlet* does not have any 'mirror location' (or front-back) ambiguity in DoA estimation by design. The confusion matrix in Figure 18(a) visually presents the spread of error for every ground truth angle. Figure 18(b) shows the empirical cumulative distribution of the error. *Owlet* exhibits a median error less than 3.3° and 90^{th} percentile error of less than 10° in this scenario.

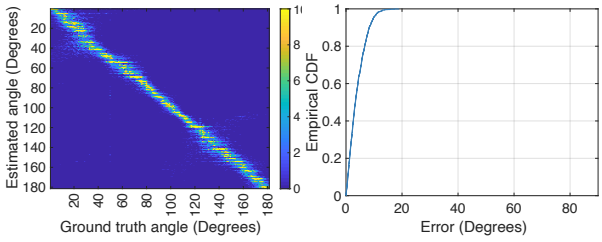

Figure 18: The performance for DoA estimation with known room size: (a) The confusion matrix and (b) the CDF of error in degrees of angle.

5.6 Localization Performance

Owlet primarily focuses on the DoA estimation of the sound. However, combining information from multiple such units can localize a sound source using triangulation. We created a setup for localization using two speakers continuously playing 50*ms* of parallel frequency pulses. We placed the *Owlet* receiver at various places within a grid in front of the speakers. *Owlet* system estimated DoA of both the speakers and estimated the location using triangulation method. Figure 19(a) shows the heatmap of the localization error and Figure 19(b) shows the corresponding CDF plot. The median localization error is 10 cm.

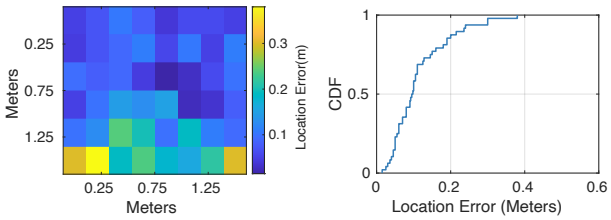

Figure 19: The localization error as (a) heatmap and (b) empirical CDF.

5.7 Comparison with traditional methods

Figure 20 compares performance of *Owlet* with traditional array-based DoA estimation techniques. We implement three popular and fundamentally different array-based methods – beamscan, minimum variance distortionless response (MVDR), and MUSIC algorithm. We apply these techniques on the microphone arrays having different number of elements. Results in Figure 20(a) show that *Owlet* significantly outperforms the other algorithms under similar conditions and signal SNR, while using only two microphones. *Owlet*'s median error is even slightly better than the MUSIC algorithm with 9-microphone array. For an estimate of the DoA resolution, we compare the spatial spectrum of each of the traditional algorithms with *Owlet*. Given *Owlet* uses a regression-based method, it does not directly produce a spatial spectrum, we rather plot the distribution of confidence score for all angles. Figure 20(b) shows the spatial spectrums for the signal coming at 20°. *Owlet* exhibits narrowest beamwidth comparable to MUSIC.

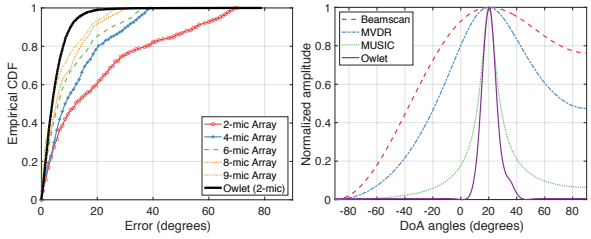

Figure 20: Performance comparison of *Owlet* with the implementation of beamscan, MVDR, and MUSIC algorithms: (a) The CDF of median errors, (b) The spatial spectrum for an incoming signal from 20° angle.

5.8 Comparison between learning models

In Figure 21, we compare the performance of different deep learning models and algorithms. The regression model performs slightly better than the classification algorithm in certain scenarios. We also compare different architectures of the regression model. Instead of using 64, 128, and 256 filters for three convolution layers, we first reduce the filter sizes to half and got a median error of 5.6°. We also apply only the first two convolution layers and reduced size filters, which leads to a median error of 5.8°. When we use two convolution layers with filters 64 and 128, the median error is 7.8°. These results show the opportunity to customize the

model for resource-constrained computational environment while maintaining the target DoA performance.

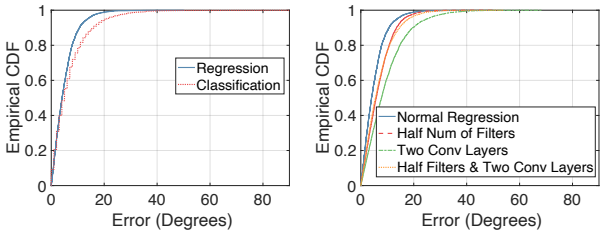

Figure 21: Performance comparison of *Owlet* with different deep learning models and architectures.

5.9 Energy consumption

In this section, we evaluate and compare the energy efficiency of *Owlet* with the traditional array-based systems. We measure the power consumption of each submodule, including hardware frontend, analog to digital conversion (ADC), and DoA computation. While we monitor the frontend and ADC directly, the setup for the computation part requires porting the runtime codes to a Raspberry Pi 4 module and monitoring the overall power variation of the module. We used a Keysight E6313A power supply and monitoring unit [35] for precise and high-resolution tracking. The setup is shown in Figure 22.

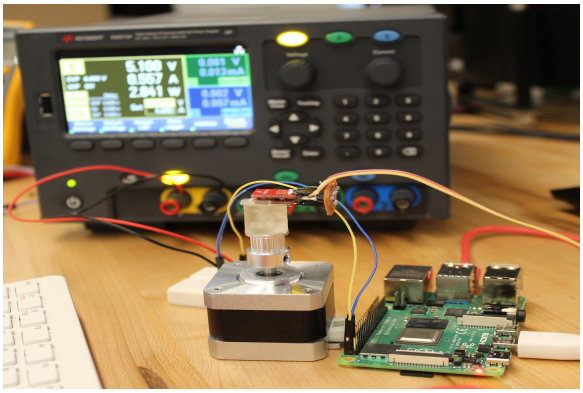

Figure 22: The setup for evaluating energy consumption. The setup tracks the energy requirements of *Owlet* and baseline microphone arrays under various conditions using a Keysight E6313A power supply and monitor.

Computation: We write the codes in Matlab and use Matlab Coder [44] to generate executable C files for the Raspberry Pi 4. We use Mathworks Raspbian image optimized for deep learning applications and cross compile the code for ARMv7 architecture with Neon Acceleration. This acceleration uses special registers for parallel operations which is an advantage for neural network systems over traditional techniques. We deploy the executable code and run it on offline data for 10,000 iterations. We collect the voltage and current readings from the power meter.

We also record the total time taken to complete the estimations. We learn that although the instantaneous power of *Owlet* is $1.92Watts$, which is approximately twice of traditional algorithms $1.05Watts$, but the time taken to complete the estimation is exponentially low, $8.3ms$ for *Owlet* compared to $2050ms$ for traditional algorithm. This contrast is because of the highly paralleled operations of the neural network, which is not possible in sequential traditional algorithms.

ADC: Figure 23 reports the energy consumed by the ADC of an MSP430 ($3 microcontroller) [31] and a Keysight Data Acquisition System ($2500 DAQ) [34]. We use the low-power 12-bit ADC of the MSP430FR5969 and vary its sampling rate to emulate the multiplexing of multiple microphones. For the Keysight DAQ, we use the 12-bit parallel channel ADC in single-shot data acquisition mode and vary the number of channels. We record the power consumption for both the devices from the power supply and we do not connect the microphones to remove the effect of microphones and their amplifiers.

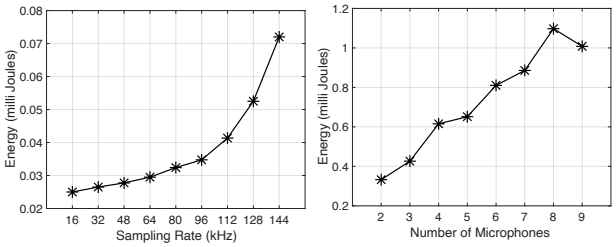

Figure 23: Energy consumption of (a) MSP430FR5969 low-power ADC [31] for different sampling rates and (b) Keysight Data Acquisition System [34] for different number of microphones.

Microphone frontend: We record the power consumption of the ADMP401 MEMS microphones [15]. To optimize the total energy consumption of the system, we only consider the time duration of $50ms$, a typical time duration to collect 800 samples at $16kHz$. We multiply this time with the average power consumed by the microphones to obtain the total energy consumed in Joules.

The evaluation summary in Section §5.1 presented a comparison of power consumption and accuracy for *Owlet* and traditional arrays. Figure 14 in Section §5.1 presents the energy consumption of *Owlet* and other state-of-the-art arrays and compares the median errors in DoA estimation. Here we show the energy consumption of each submodule, i.e., computation, ADC, and microphone frontend, separately in Figure 24. *Owlet* consumes less than a 100^{th} of the power required by the traditional arrays for similar angular resolution and accuracy.

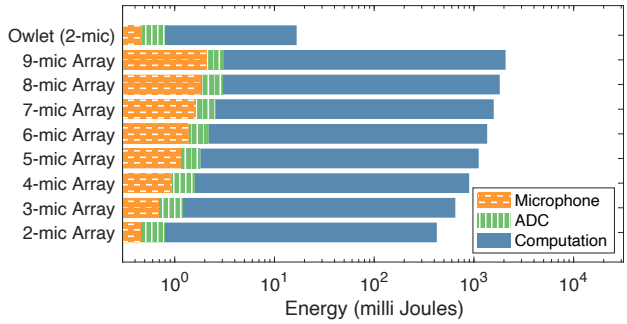

Figure 24: Overall energy consumption of array-based systems and *Owlet*.

6 LIMITATIONS AND DISCUSSION

Needless to say, current version of *Owlet* is an exploratory first realization of the concept and there is room for improvements and further work. We discuss a few points here.

• **Multiple sound sources:** We tested our prototype in various acoustic environments, with various noise sources and for different types of target sounds. At this stage, we assume one signal source for DoA estimation. When multiple sources overlap, the system attempts to analyze the strongest signal for direction estimation and considers other sources as noise. While considering one significant sound source is deemed practical for many applications, we believe multiple DoA estimation will be possible with *Owlet* system. Probably the most viable method would be adapting statistical separation source signals and then seek for an optimal match with the signatures for DoA detection. We leave this topic to future work.

• **Theoretical bounds on capacity:** Array signal processing has been studied deeply over the past three decades. As a result, it is possible to accurately estimate the theoretical bounds on the achievable spatial resolution under various limiting factors and array formation. Such information is crucial for array design and simulations. The concept of *Owlet* differs significantly from the array processing techniques for spatial information retrieval. However, we can analyze the bounds of its performance through information-theoretic treatment of the entropy of the available directional gain patterns. The shape and size of the stencil along with the frequency of the sound include additional constraints on *Owlet*'s capacity to produce diverse gain-patterns. Such a theoretical assessment is likely to benefit the understanding of the system and guide improvement efforts.

• **Mobility:** The Doppler frequency shift due to fast motion may affect the frequency gain-patterns that *Owlet* uses as directional cues. Our prototype operates at the low-frequency audible signal range, which is less affected by the mobility of the sound source or the receiver. Moreover, the DoA estimation with the parallel frequency signal provides a certain degree of robustness against Doppler frequency shift. Therefore, we sidestepped the analysis of the system for mobility. However, *Owlet* can potentially operate at higher frequency signals, and it will require design considerations to detect and compensate for the frequency shifts.

• **Inaudibility of sound signal:** In this paper, we have considered audible sound frequencies for system calibration and source signals. Long wavelength of the low-frequency signals likely to show less diversity in the frequency gain-pattern which affects the achievable angular resolution. We deliberately selected this operational frequency to show system performance at the lower end of the spectrum, where higher frequency can show better performance in terms of both the spatial resolution and reducing the stencil size. We plan to explore inaudible near-ultrasound ($17 - 24kHz$) and ultrasound ($> 24kHz$) ranges for subsequent versions of the system.

7 RELATED WORK

The literature is rich in techniques for spatial analysis of sound. Seminal work in direction of arrival estimation using microphone arrays [7, 47, 48, 50], array signal processing for beamforming [20, 36, 57], and subspace-based super-resolution algorithms [53, 65] have significantly advanced this field of study. In the recent past, new innovations in ubiquitous spatial acoustic sensing [13, 26, 41, 42, 49, 60–62, 67, 68] have opened up new opportunities. We sample below two topics closely related to *Owlet*.

• **Acoustic structures:** The study of the impact of structures on sound fields has a long lineage. The usage of large building structures to amplify sound or reduce noise is found in many ancient architectures. Architectural acoustics is also scientifically explored and implemented in modern houses and auditoriums for reverberation control and sound isolation. Research relevant to *Owlet* include designing 3D-printed acoustic metamaterials to absorb specific frequency bands [11] and developing meta-surfaces to generate diffraction-limited acoustic fields [46]. The study of acoustic structures in sensing applications is relatively new. Li et al. [39] construct acoustic filters using additive manufacturing that controls the impedance of discrete frequencies. In [30], authors create physical notches on a surface to make acoustic barcodes. Some recent works [38, 63] make tangible user interfaces by 3D printing tiny acoustic structures. These works vary the structure shapes to create distinguishable frequency responses and use smartphone microphones to collect the signal and perform classification.

• **Monaural DoA:** Existing studies [47, 48, 50, 67] explored the usage of microphone array for DoA estimation. Lately, several papers have emerged that focus on reducing resources in directional acoustic sensing. For instance, [16] keeps a single microphone in a known room and makes use of the reflections and scattering from the walls of the room to localize the sound source. In [64], a small vertical wall of varying shape is placed next to a microphone which changes the frequency response for different directions of sound. A few recent works [17, 18] place small structures like legos and cubes around a microphone to produce scattering. These works on monaural localization either keep a dictionary of possible source models or predict the source model before estimating the DoA. The structures used in these works are big which add diversity but are unsuitable mobile sensor systems. Similarly, [28] learns the spectral features of a sound signal to detect the distance of the speaker but only in controlled indoor environments.

8 CONCLUSION

This paper presents *Owlet*, a practical system for low-power acoustic DoA estimation and source localization using acoustic structures. The core idea is to shape the directional impulse response of a microphone by covering it with a carefully designed stencil. Sound diffracted by this stencil carries a direction-specific signature, an indicator of the directional information of the recorded sound. *Owlet* develops a robust DoA estimation technique using this concept and creates a prototype for demonstration [25] and evaluation. The prototype exhibits a similar angular resolution of a large aperture microphone array, while using only two microphones in a compact form-factor. The core idea opens up new applications in ubiquitous acoustic sensing and new possibilities of sensing architecture using 3D-printed passive structures. This paper is the first step towards this broader vision.

REFERENCES

[1] Ahmad, S., and Ahmad, T. Implementation of recursive least squares (rls) adaptive filter for noise cancellation. *International Journal of Scientific Engineering and Technology 1*, 4 (2012), 46–48.

[2] Andrade, J., Santana, P., and Almeida, A. Motion-induced acoustic noise awareness for socially-aware robot navigation. In *2018 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)* (2018), IEEE, pp. 24–29.

[3] Anuntachai, A., and Pavaranchanakul, K. Analyze traffic conditions and events with sound processing. In *2020 20th International Conference on Control, Automation and Systems (ICCAS)* (2020), IEEE, pp. 344–349.

[4] API, A. P. https://autodeskfusion360.github.io/, 2020.

[5] Arduino. https://www.arduino.cc/, 2020.

[6] Autodesk. https://www.autodesk.com/products/fusion-360/overview, 2020.

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

[8] Bernas, M., Płaczek, B., Korski, W., Loska, P., Smyła, J., and Szymała, P. A survey and comparison of low-cost sensing technologies for road traffic monitoring. *Sensors 18*, 10 (2018), 3243.

[9] Bridson, R. Fast poisson disk sampling in arbitrary dimensions. *SIGGRAPH sketches 10* (2007), 1.

[10] Brughera, A., Mikiel-Hunter, J., Dietz, M., and McAlpine, D. Brainstem biophysics contribute to sound-source localisation in reverberant scenes. *bioRxiv* (2019), 694356.

[11] Casarini, C., Tiller, B., Mineo, C., MacLeod, C. N., Windmill, J. F., and Jackson, J. C. Enhancing the sound absorption of small-scale 3-d printed acoustic metamaterials based on helmholtz resonators. *IEEE Sensors Journal 18*, 19 (2018), 7949–7955.

[12] Chan, J., Rea, T., Gollakota, S., and Sunshine, J. E. Contactless cardiac arrest detection using smart devices. *NPJ digital medicine 2*, 1 (2019), 1–8.

[13] Choudhury, R. R. Earable computing: A new area to think about. In *Proceedings of the 22nd International Workshop on Mobile Computing Systems and Applications* (2021), pp. 147–153.

[14] Coppola, M., McGuire, K. N., De Wagter, C., and de Croon, G. C. A survey on swarming with micro air vehicles: Fundamental challenges and constraints. *Frontiers in Robotics and AI 7* (2020), 18.

[15] Devices, A. Admp401: Omnidirectional microphone with bottom port and analog output, 2013.

[16] Dokmanić, I., and Vetterli, M. Room helps: Acoustic localization with finite elements. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2012), Ieee, pp. 2617–2620.

[17] El Badawy, D., and Dokmanić, I. Direction of arrival with one microphone, a few legos, and non-negative matrix factorization. *IEEE/ACM Transactions on Audio, Speech, and Language Processing 26*, 12 (2018), 2436–2446.

[18] El Badawy, D., Dokmanić, I., and Vetterli, M. Acoustic doa estimation by one unsophisticated sensor. In *International Conference on Latent Variable Analysis and Signal Separation* (2017), Springer, pp. 89–98.

[19] Elegoo. https://www.elegoo.com/product/elegoo-mars-uv-photocuring-lcd-3d-printer/, 2020.

[20] Elko, G. W. Microphone array systems for hands-free telecommunication. *Speech communication 20*, 3-4 (1996), 229–240.

[21] Feng, T., Nadarajan, A., Vaz, C., Booth, B., and Narayanan, S. Tiles audio recorder: an unobtrusive wearable solution to track audio activity. In *Proceedings of the 4th ACM Workshop on Wearable Systems and Applications* (2018), pp. 33–38.

[22] for Ornithology, B. T. Owl hearing. https://www.bto.org/our-science/projects/project-owl/learn-about-owls/owl-hearing. Last accessed 19 August 2020.

[23] Furui, S. Speech recognition technology in the ubiquitous/wearable computing environment. In *Acoustics, Speech, and Signal Processing, 2000. ICASSP'00. Proceedings. 2000 IEEE International Conference on* (2000), vol. 6, IEEE, pp. 3735–3738.

[24] Ganek, H., and Eriks-Brophy, A. Language environment analysis (lena) system investigation of day long recordings in children: A literature review. *Journal of Communication Disorders 72* (2018), 77–85.

[25] Garg, N., Bai, Y., and Roy, N. Demo: Microstructure-guided spatial sensing for low-power iot. In *The 19th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys '21)* (2021).

[26] Garg, N., and Roy, N. Acoustic sensing for detecting projectile attacks on small drones. In *Proceedings of the 21st International Workshop on Mobile Computing Systems and Applications* (2020), pp. 104–104.

[27] Garg, N., and Roy, N. Enabling self-defense in small drones. In *Proceedings of the 21st International Workshop on Mobile Computing Systems and Applications. ACM* (2020).

[28] Georganti, E., May, T., van de Par, S., Harma, A., and Mourjopoulos, J. Speaker distance detection using a single microphone. *IEEE transactions on audio, speech, and language processing 19*, 7 (2011), 1949–1961.

[29] Gilkerson, J., and Richards, J. A. The lena natural language study. *Boulder, CO: LENA Foundation. Retrieved March 3* (2008), 2009.

[30] Harrison, C., Xiao, R., and Hudson, S. Acoustic barcodes: passive, durable and inexpensive notched identification tags. In *Proceedings of the 25th annual ACM symposium on User interface software and technology* (2012), pp. 563–568.

[31] Instruments, T. Texas instruments msp430fr5969. https://www.ti.com/tool/MSP-EXP430FR5969, 2021.

[32] Jaffe, J. S., Franks, P. J., Roberts, P. L., Mirza, D., Schurgers, C., Kastner, R., and Boch, A. A swarm of autonomous miniature underwater robot drifters for exploring submesoscale ocean dynamics. *Nature communications 8*, 1 (2017), 1–8.

[33] Keysight. https://www.keysight.com/us/en/products/waveform-and-function-generators.html, 2020.

[34] Keysight. Keysight data acquisition system u2331a. https://www.keysight.com/us/en/assets/7018-03510/data-sheets/5991-0566.pdf, 2021.

[35] Keysight. Keysight e6313a power supply. https://www.keysight.com/us/en/assets/9018-04576/user-manuals/9018-04576.pdf, 2021.

[36] Krishnaveni, V., Kesavamurthy, T., and Aparna, B. Beamforming for direction-of-arrival (doa) estimation-a survey. *International Journal of Computer Applications 61*, 11 (2013).

[37] Landau, H. Sampling, data transmission, and the nyquist rate. *Proceedings of the IEEE 55*, 10 (1967), 1701–1706.

[38] Laput, G., Brockmeyer, E., Hudson, S. E., and Harrison, C. Acoustruments: Passive, acoustically-driven, interactive controls for handheld devices. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (2015), pp. 2161–2170.

[39] Li, D., Levin, D. I., Matusik, W., and Zheng, C. Acoustic voxels: computational optimization of modular acoustic filters. *ACM Transactions on Graphics (TOG) 35*, 4 (2016), 1–12.

[40] Liu, V., Parks, A., Talla, V., Gollakota, S., Wetherall, D., and Smith, J. R. Ambient backscatter: Wireless communication out of thin air. *ACM SIGCOMM Computer Communication Review 43*, 4 (2013), 39–50.

[41] Mao, W., and Qiu, L. Dronetrack: An indoor follow-me system using acoustic signals. *GetMobile: Mobile Computing and Communications 21*, 4 (2018), 22–24.

[42] Mao, W., Wang, M., and Qiu, L. Aim: Acoustic imaging on a mobile. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services* (2018), pp. 468–481.

[43] Mason, F., Chiariotri, F., Campagnaro, F., Zanella, A., and Zorzi, M. Low-cost auv swarm localization through multimodal underwater acoustic networks. In *Global Oceans 2020: Singapore–US Gulf Coast* (2020), IEEE, pp. 1–7.

[44] Mathworks. Mathworks matlab coder. https://www.mathworks.com/products/matlab-coder.html, 2021.

[45] McGuire, K., De Wagter, C., Tuyls, K., Kappen, H., and de Croon, G. C. Minimal navigation solution for a swarm of tiny flying robots to explore an unknown environment. *Science Robotics 4*, 35 (2019).

[46] Memoli, G., Caleap, M., Asakawa, M., Sahoo, D. R., Drinkwater, B. W., and Subramanian, S. Metamaterial bricks and quantization of meta-surfaces. *Nature communications 8*, 1 (2017), 1–8.

[47] Nakadai, K., Nakajima, H., Yamada, K., Hasegawa, Y., Nakamura, T., and Tsujino, H. Sound source tracking with directivity pattern estimation using a 64 ch microphone array. *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems* (2005), 1690–1696.

[48] Nakajima, H., Kikuchi, K., Daigo, T., Kaneda, Y., Nakadai, K., and Hasegawa, Y. Real-time sound source orientation estimation using a 96 channel microphone array. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems* (2009), IEEE, pp. 676–683.

[49] Nandakumar, R., Iyer, V., Tan, D., and Gollakota, S. Fingerio: Using active sonar for fine-grained finger tracking. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (2016), pp. 1515–1525.

[50] Niwa, K., Hioka, Y., Sakauchi, S., Furuya, K., and Haneda, Y. Estimation of sound source orientation using eigenspace of spatial correlation matrix. In *2010 IEEE International Conference on Acoustics, Speech and Signal Processing* (2010), IEEE, pp. 129–132.

[51] Orena, A. J., Byers-Heinlein, K., and Polka, L. Reliability of the language environment analysis recording system in analyzing french–english bilingual speech. *Journal of Speech, Language, and Hearing Research 62*, 7 (2019), 2491–2500.

[52] Pan, Z., Ge, Y., Zhou, Y. C., Huang, J. C., Zheng, Y. L., Zhang, N., Liang, X. X., Gao, P., Zhang, G. Q., Wang, Q., et al. Cognitive acoustic analytics service for internet of things. In *2017 IEEE International Conference on Cognitive Computing (ICCC)* (2017), IEEE, pp. 96–103.

[53] Pavlidi, D., Griffin, A., Puigt, M., and Mouchtaris, A. Real-time multiple sound source localization and counting using a circular microphone array. *IEEE Transactions on Audio, Speech, and Language Processing 21*, 10 (2013), 2193–2206.

[54] Peat, K. A first aproximation to the effects of mean flow on sound propagation through cylindrical capillary tubes. *Journal of sound and vibration 175*, 4 (1994), 475–489.

[55] Piczak, K. J. Environmental sound classification with convolutional neural networks. In *2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP)* (2015), IEEE, pp. 1–6.

[56] Pierce, A. D. Diffraction of sound around corners and over wide barriers. *The Journal of the Acoustical Society of America 55*, 5 (1974), 941–955.

[57] Ramos, A. L., Holm, S., Gudvangen, S., and Otterlei, R. Delay-and-sum beamforming for direction of arrival estimation applied to gunshot acoustics. In *Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense X* (2011), vol. 8019, International Society for Optics and Photonics, p. 80190U.

[58] Römer, H. Directional hearing in insects: biophysical, physiological and ecological challenges. *Journal of Experimental Biology 223*, 14 (2020).

[59] Rostami, M., Sundaresan, K., Chai, E., Rangarajan, S., and Ganesan, D. Redefining passive in backscattering with commodity devices. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking* (2020), pp. 1–13.

[60] Roy, N. *Inaudible acoustics: techniques and applications.* PhD thesis, University of Illinois at Urbana-Champaign, 2018.

[61] Roy, N., Hassanieh, H., and Choudhury, R. R. Backdoor: Sounds that a microphone can record, but that humans can't hear. *GetMobile: Mobile Computing and Communications 21*, 4 (2018), 25–29.

[62] Roy, N., Shen, S., Hassanieh, H., and Choudhury, R. R. Inaudible voice commands: The long-range attack and defense. In *15th USENIX Symposium on Networked Systems Design and Implementation (NSDI)* (2018), USENIX Association, pp. 547–560.

[63] Savage, V., Head, A., Hartmann, B., Goldman, D. B., Mysore, G., and Li, W. Lamello: Passive acoustic sensing for tangible input components. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (2015), pp. 1277–1280.

[64] Saxena, A., and Ng, A. Y. Learning sound location from a single microphone. In *2009 IEEE International Conference on Robotics and Automation* (2009), IEEE, pp. 1737–1742.

[65] Schmidt, R. Multiple emitter location and signal parameter estimation. *IEEE transactions on antennas and propagation 34*, 3 (1986), 276–280.

[66] Semenzin, C., Hamrick, L., Seidl, A., Kelleher, B., and Cristia, A. Towards large-scale data annotation of audio from wearables: validating zooniverse annotations of infant vocalization types. In *2021 IEEE Spoken Language Technology Workshop (SLT)* (2021), IEEE, pp. 1079–1085.

[67] Shen, S., Chen, D., Wei, Y.-L., Yang, Z., and Choudhury, R. R. Voice localization using nearby wall reflections.

[68] Shen, S., Roy, N., Guan, J., Hassanieh, H., and Choudhury, R. R. Mute: bringing iot to noise cancellation. In *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication* (2018), ACM, pp. 282–296.

[69] Tang, P., and Sirignano, W. Theory of a generalized helmholtz resonator. *Journal of Sound and Vibration 26*, 2 (1973), 247–262.

[70] Telembici, T., and Grama, L. Detecting indoor sound events. *Acta Technica Napocensis 59*, 2 (2018), 13–17.

[71] Wang, A., Sunshine, J. E., and Gollakota, S. Contactless infant monitoring using white noise. In *The 25th Annual International Conference on Mobile Computing and Networking* (2019), pp. 1–16.

[72] Wiener, F. M. Sound diffraction by rigid spheres and circular cylinders. *The Journal of the Acoustical Society of America 19*, 3 (1947), 444–451.

[73] Zhou, B., Elbadry, M., Gao, R., and Ye, F. Batmapper: Acoustic sensing based indoor floor plan construction using smartphones. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services* (2017), pp. 42–55.

[74] Zieliński, T. G., Opiela, K. C., Pawłowski, P., Dauchez, N., Boutin, T., Kennedy, J., Trimble, D., Rice, H., Van Damme, B., Hannema, G., et al. Reproducibility of sound-absorbing periodic porous materials using additive manufacturing technologies: Round robin study. *Additive Manufacturing 36* (2020), 101564.