

SPiDR: Ultra-low-power Acoustic Spatial Sensing for Micro-robot Navigation

Yang Bai[‡]
yangbai8@umd.edu

University of Maryland College Park

Nakul Garg[‡]
nakul22@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 the design and implementation of *SPiDR*, an ultra-low-power spatial sensing system for miniature mobile robots. This acoustic sensor produces a cross-sectional map of the field-of-view using only one speaker/microphone pair. While it is challenging to have enough spatial diversity of signal with a single omnidirectional source, we leverage sound's interaction with small structures to create a 3D-printed passive filter, called a stencil, that can project spatially coded signals on a region at a fine granularity. The system receives a linear combination of the reflections from nearby objects and applies a novel power-aware depth-map reconstruction algorithm. The algorithm first estimates the approximate locations of the objects in the scene and then iteratively applies fractional multi-resolution inversion. *SPiDR* consumes only 10mW of power to generate a depth-map in real-world scenario with over 80% structural similarity score with the scene.

CCS CONCEPTS

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

KEYWORDS

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

ACM Reference Format:

Yang Bai, Nakul Garg, and Nirupam Roy. 2022. SPiDR: Ultra-low-power Acoustic Spatial Sensing for Micro-robot Navigation. In *The 20th Annual International Conference on Mobile Systems, Applications and Services (MobiSys '22)*, June 25–July 1, 2022, Portland, OR, USA. ACM, New York, NY, USA, 15 pages. <https://doi.org/10.1145/3498361.3539775>

1 INTRODUCTION

Miniature mobile robots are emerging with new capabilities and skills. Insect-sized robots, a few inches in size, can work as first responders to search for survivors in disaster debris [63], perform

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

MobiSys '22, June 25–July 1, 2022, Portland, OR, USA

© 2022 Association for Computing Machinery.

ACM ISBN 978-1-4503-9185-6/22/06...\$15.00

<https://doi.org/10.1145/3498361.3539775>

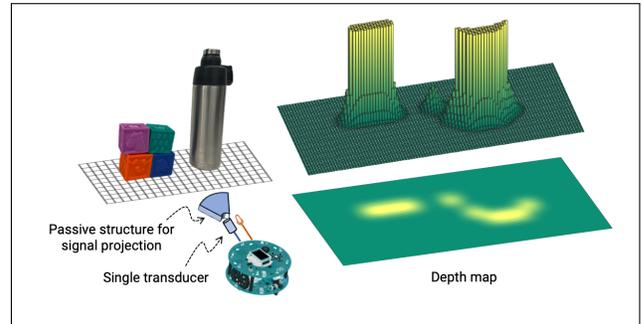


Figure 1: *SPiDR*, an ultra-low-power acoustic spatial sensing system for mobile robots. The system uses a carefully designed 3D-printed micro-structure for projecting spatially coded signals for imaging.

mining or agricultural foraging [34, 125], and can even move heavy objects to organize [22]. These small robotic systems, when paired with autonomous navigation, will create new possibilities in a wide range of applications, including precision farming, disaster management, and surveillance and monitoring. However, it will require overcoming a set of challenges to realize this vision and probably understanding the environment for navigation is the most crucial of them. Existing techniques are not directly applicable in small robotic systems for their unique constraints of limited energy source, small size, limited computational power, and the requirement of low-cost manufacturing (SWaP-C constraints [74]). In this paper, we present *SPiDR*¹, an ultra-low-power acoustic spatial sensing system capable of generating an accurate depth-map of nearby objects. Figure 1 gives an overview of the system.

Widely adopted methods for scene perception basically scan the surroundings and generate a depth-map of the scene [76]. Lidar [20], for example, uses a moving laser source to scan each point in the surrounding 3D space and finds the distance to that point by analyzing the reflected signal. Radar- or sonar-based systems [11, 70], on the other hand, eliminate mechanical movements using a phased array of antenna/transducers that is capable of electronically focusing the radio frequency or sound signals to a specific direction, called beamforming. Given mechanical maneuvers are power inefficient, beamforming would be preferable for our application. However, in beamforming, a fine angular resolution of the beam requires a large number of transducers to transmit the signal simultaneously. This architecture needs parallel transmit/receive hardware chains and processing that leads to higher computational complexity and power budget. Note that,

¹SPiDR stands for Structure-assisted Perception, Detection, and Ranging

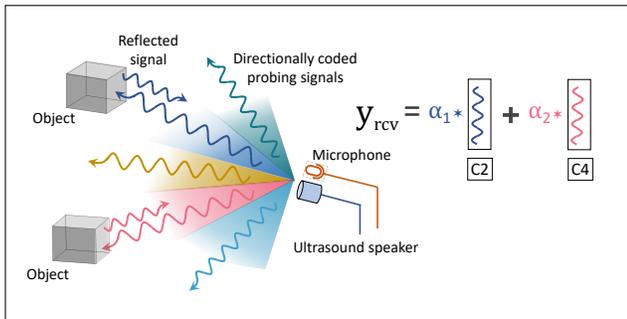


Figure 2: The concept of the spatially coded channel sounding method. The received signal is the weighted linear combination of the reflections that bear the direction-specific signature.

in micro-robot systems, only 10-30mW of power remains available for sensing after allocating power to the actuators for locomotion [18, 49, 51, 61].

In our quest to develop the ultra-low-power sensing system, we start by exploring the possibility of spending minimum power on the sensing hardware. We use acoustic signals for their small wavelength at a low frequency (8.5mm of wavelength at 40kHz). Small wavelength offers finer resolution for spatial sensing, while simple low-cost oscillators can produce the frequencies. For active sensing, where the sensor requires to send a signal and analyze the received reflections, it boils down to using at least one channel for signal generation and one for reception (a time-interleaved transmit/receive channel leads to similar performance). The single transmitter/receiver model rules out the possibility of electronic beamforming for steering the signal. Also, we do not consider mechanical steering for scanning the scene with one sensor – even an advanced low-power mechanical steering system consumes 33mW of power [61]. We overcome this issue by using a spatially coded signal that sends unique patterns of signal in each direction, as shown in Figure 2. The reflection of this signal carries the information of the direction of the reflecting object and its distance is known from the signal’s time-of-flight – the two parameters needed to know the locations of the obstacles in the scene. Sounding the channel with spatially diverse signals is still challenging given we have only one omnidirectional transmitter. Such transmitter spreads the same signal in all directions failing the purpose of spatial codes. We solve this problem by controlling the properties of the multipath channels, a well-studied concept in wireless communication, using a 3D-printed micro-structure.

Multipath is a natural phenomenon where a signal, after leaving the transmitter, reflects off objects in the environment to create replicas and the replicas propagate through paths of different delays before combining at the receiver. The lengths of these individual paths decide the phase delays of the replicas and therefore their superimposition leads to a specific amplitude and phase of the received signal. A unique combination of the path lengths can produce a unique signal similar to a code pattern. Of course, path lengths in environmental multipath are not defined and therefore do not help in our solution, but we use the concept of multipath superposition to create signals with direction-specific codes.

We design a 3D-printed cover for the speaker, called a stencil, that divides the speaker’s output into multiple replicas by passing it through small internal tubes. The lengths of these internal tubes are carefully calculated to channelize the signal replicas through different time-delayed paths before releasing them through separate output sound holes pointed in different spatial directions. These delayed replicas of the sound signals again interfere with each other and create complex but predictable patterns at different points of the scene. It is like sending a specific code to a particular 3D point in the space. Imagine the target space is divided into uniformly spaced 3D points or voxels. If a voxel is occupied by an object (or a part of the object), the coded signal will reflect back to the receiver with its unique signature. In fact, all the coded reflections from the target scene combine at the receiver and can be separated through processing to convert them to a 3D point cloud of the objects in the scene. The navigation application, however, requires only a 2D cross-sectional view of the scene, which reduces the required processing power. Given the stencil is simply a passive structure, only the transmitter (speaker) and the receiver (microphone) in this sensor frontend consume energy, which is less than half a milli-joule.

New architecture for the sensing frontend alone does not provide sufficient advantage in overall low-power sensing. The computation of the scene from the received combination of reflected codes requires a channel inversion approach, which can be a computationally expensive operation. As detailed in Section 3.1, *SPiDR* adopts the approach of fractional scene reconstruction and stitching algorithm that significantly reduces the energy cost for computation. The key idea is to have a coarse-grained probabilistic estimate of the objects’ locations in the scene and then apply iterative refinements using partial inversion of the channel matrix leading to exponentially less computation than the traditional approach. *SPiDR* can also take advantage of the movement of the robots to stack multiple maps of the same objects in the view to create a higher resolution representation of the scene. With optimized computing, the entire scene reconstruction process takes 0.83mJ per depth-map for the same 20 cm by 10 cm scene.

Along with past works in mobile sensing and localization, our work builds on the well-established theories of wireless channel estimation, multipath propagation of signal, and computational sensing. The idea of spatial sensing through a single measurement has been explored in computational photography [33] that applied “coded aperture” to pass spatial image information through incoherent functions and combine on a sensor. Coded aperture for sensing can be similar (in fact, the inverse) idea of using a 3D-printed stencil for projecting probing signals. Some of the recent works in medical ultrasound and radio-frequency (RF) imaging have also shown the use of coded aperture to improve image clarity. Probably the work closest to *SPiDR* is the compressed ultrasound imaging technique [75] that mechanically moved a coded aperture mask to image objects placed in water. However, *SPiDR*’s technique for low-power spatial sensing differs from the past works in four fundamental ways to meet the practical constraints of a micro-robot platform.

(a) No scanning required: Past works on radar or spatial sensing with a single sensor relied on movement of the sensor for creating a synthetic aperture [24] or added mechanical maneuver to the sensor [91] to introduce diversity in measurements. *SPiDR* does not require any mechanical motion and therefore does not need any power consuming actuator in the design.

(b) No location specific information needed: Unlike the past approaches with echo analysis [29, 64] and location fingerprinting [128], *SPiDR* does not attempt to learn the signal pattern in a specific location. The purpose of our calibration is to learn the ideal channel in an empty space with an ideal point reflector. Therefore, a one-time in-lab calibration is sufficient and the sensor comes with this channel calibration pre-loaded. Moreover, the design of the stencil does not use any location specific information and therefore one design works for all locations.

(c) Power-aware hardware/signal: The depth resolution is inversely proportional to the frequency. However, we restrict ourselves to relatively lower frequency of ultrasound and use commercially available 40kHz piezo transducer. These transducers are inexpensive and low-power oscillators can produce this signal.

(d) Power-aware computing: We exploit the sparsity in target scene to reduce the number of operations in computing the depth-map. The computational cost of our iterative scene reconstruction algorithm is proportional to the fraction of the scene occupied by objects.

This paper explores a broader vision of leveraging passive structure toward extremely low-power sensing and perception. As this research evolves, we envision that the core techniques will serve to enable applications beyond robotic navigation. Possible applications include ubiquitous gesture recognition, long-term underwater monitoring, and soft-tissue imaging with handheld devices. While several opportunities exist in research and application, this paper focuses on developing the core capabilities and assessing the limits of the systems. To this end, we have made the following three specific contributions at the current stage of this project:

- A novel channel sounding method for spatial sensing with a zero-power passive extension to a single sound source. We leverage sound’s interaction with small structures to create a 3D-printed passive filter, called stencil, that can project spatially coded signals on a region at a fine granularity.
- An ultra-low-power processing algorithm for depth-map reconstruction algorithm. The algorithm iteratively refines the initial coarse estimation of the map using our fractional multi-resolution reconstruction.
- A functional hardware-software prototype of the spatial perception system. We use this demonstrable implementation of *SPiDR* for benchmarking the power profile and evaluating the performance under real-life conditions.

2 CORE INTUITIONS AND PRIMERS

Traditional methods for scene perception rely on scanning the region with directional beams (either transmit or receive, or both) at one angle at a time. *SPiDR* takes a fundamentally different approach and used the ‘spatially coded channel sounding’ method. The key

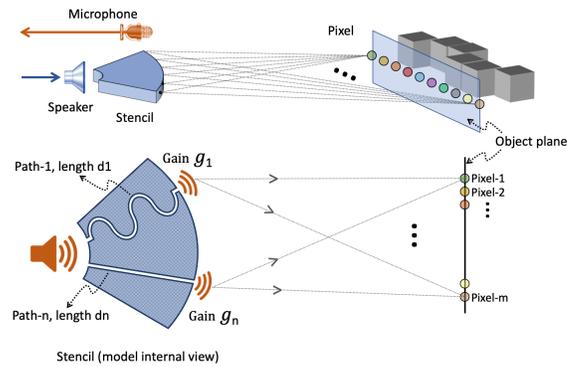


Figure 3: Diversity projection with the stencil with the internal channels to encode unique gains to signals to probe each pixel on the object plane with a unique signature.

intuition is to send the probing signal in a way such that it bears unique signatures or codes for every spatial angle. When reflected by an object in a certain direction the reflected signals bear the code of that specific direction. The scene reconstruction method essentially identifies the unique codes present in the reflected signal and maps them to the direction and distance of the objects. While the concept is clean in theory, there are several challenges in leveraging its advantage in a low-power and low-complexity system. In order to minimize the power consumption and the complexity of the hardware frontend, we use only one speaker and one microphone to send and receive the probing signal. We have shown that a carefully designed and 3D-printed passive structure can be used to embed directional signatures in the transmit probing signal without requiring any mechanical or computational resources. If this spatially coded channel sounding is done well, a single receiver can use a linear superimposition model to recover the direction and depth information of the nearby objects from the reflected signals. We elaborate on these two key intuitions next.

2.1 Coded signal projection with structures

The superimposition of reflected components of a signal, known as the *multipath effect*, alters the waveform of the signal and its impact can be diverse based on the number of paths and their complex gains. While the multipath effect is considered unwanted in communication and other acoustic applications, we leverage this concept to project spatially diverse probing signals on the scene. However, instead of relying on the ambient reflections, we use a carefully designed acoustic structure, we call stencil, to produce distinct channels for sound propagation with desired delays and attenuations. Figure 3 shows the design principle of the stencil.

The stencil is a 3D-printed porous cap that covers the speaker and channelizes the output signal through a number of internal tubes are connected to the openings pointed at different angular directions, as shown in Figure 4. The size and length of the tubular paths vary to control the amplitude and relative phase of the signals at the opening. Note that the openings of the stencil behave as individual signal sources that transmit the same signal with a unique delay and attenuation, i.e., multiplied with a unique complex gain.

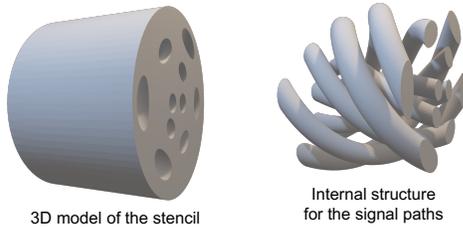


Figure 4: (left) The 3D design of the stencil and (right) the internal structure showing the tubular helical paths of different lengths.

Figure 5 shows that the stencil spreads the signal energy in a wide region. On the other hand, Figure 6 shows how the stencil diversifies the amplitude of the signal in different directions. These signals again combine with each other before reaching the object plane and the final signal at the object plane is reflected back to the receiver. Therefore, the stencil should be designed in a way such that it creates a specific signal pattern at the object plane, representing unique and diverse codes.

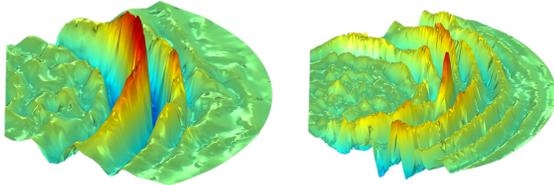


Figure 5: Ultrasound emitted from the speaker without (left) and with the stencil (right). The stencil spreads the signal energy over the region of interest.

2.2 Single receiver depth mapping

We discretize the scene to a collection of pixels on a vertical plane at a certain distance from the sensor. Multiple such planes placed at different depths cover the region of interest. A horizontal cross-section of the 3D scene, however, provides sufficient information about the location of the objects for navigation purposes. Therefore, we consider one layer of pixels per plane representing the 2D cross-section of the scene we are interested in reconstructing. The probing signal only reflects off the pixels that represent an object and combines linearly at the only microphone used for sensing. The received signal y_{rcv} can be formulated as $y_{rcv} = Hx$. Here H is the collection of ideal reflected signals from each individual pixel organized in columns. The vector x represents the reflectivity status of the pixels indicating the fraction of the ideal signal reflected from each pixel. The values of x are higher when the object at the corresponding location is a good reflector of the signal and a zero value indicates the absence of object at that pixel. In other words, the vector x selects columns of H to map a scene to the weighted sum of reflections y_{rcv} . Our scene reconstruction algorithm recovers x from the received signal y_{rcv} .

3 SYSTEM DESIGN

3.1 Low-power scene reconstruction

The cross-sectional scene reconstruction or depth-map estimation relies on the inversion of the channel matrix H to recover the pixel occupancy vector x , i.e., $x = H^{-1}y$. Here, y is the received time-domain signal with length M and x is the 2D pixel matrix

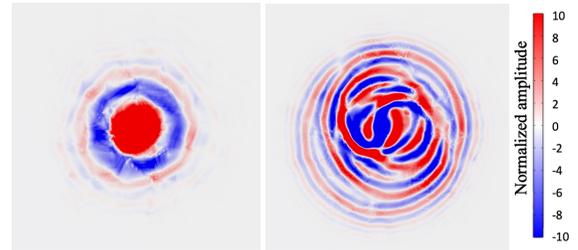


Figure 6: The amplitude of acoustic signal at the cross section of the image plane when (left) speaker does not have any stencil, and (right) when it has a stencil. The internal micro-structure of the stencil diversifies the signal amplitude as direction codes.

reformatted as a vector of length N . This makes the dimension of the channel matrix H to be $M \times N$. Note that, the value of M can be largely dependent on the time of sampling per measurement (t_{meas}) and the sample rate (f_s). For instance, 5 milliseconds of measurement at 200kHz sample rate, which is the Nyquist rate for low-frequency ultrasounds, makes $M (= t_{meas} \times f_s)$ to be 1000 samples. On the other hand, N is the total number of pixels covering the region of interest. Typically navigation of micro-robot requires scanning a small region due to their size. We scan a $20cm \times 10cm$ region at $1cm$ resolution, leading to $N = 200$ pixels per map. Manipulation of a 1000×200 H matrix is trivial on a standard computing platform, but this can consume a significant fraction of power and computational latency on a single board computer in a tiny robot. The power consumption and latency of a matrix inversion and multiplication operations increase exponentially with the dimension of the matrix. Figure 7(a) shows the increase in power for different column sizes of our channel matrix H for the reconstruction of each map. A 1000×200 channel matrix will consume around $400mJ$ of energy. We address this challenge through our multi-scale scene computation strategy and reduce the power consumption by $50\times$, without compromising the quality of the reconstruction. This technique uniquely leverages the linearity of the signal model and the ‘locality of pixel values’ to reduce computational latency and power, as elaborated next.

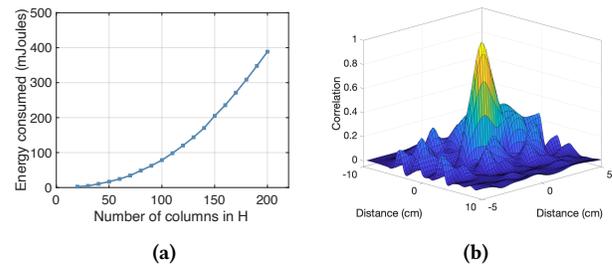


Figure 7: (a) Energy consumption for different number of columns in the channel matrix H . (b) The correlation between the signals at nearby locations. The pixels within $3cm$ have correlation higher than 0.5.

Fractional scene computing: If we divide our region of interest into a $M \times N$ grid, each block will correspond to a pixel of the reconstructed scene. Also, each column in the channel matrix represents the coded channel between a block in the grid and the sensor. Ideally, the channel for each block should be uncorrelated with any other blocks in the grid. However, despite our coded

channel sounding technique, the fundamental diffraction limit of the low-frequency signals used by our system makes it different in practice. As shown in Figure 7(b), the channel of a block shows a certain degree of correlation with nearby blocks with a gradual reduction over distance. This limitation in channel independence, which we overcome at a later stage in the processing, offers an opportunity to have a probabilistic measure of the occupancy status of a grid without directly measuring the channel. In other words, we can identify a certain block and use the knowledge of the correlation of channels in nearby grids to assume the presence of an object at that location with a certain confidence. With this intuition, we define *meta-pixels* that can cover a cluster of pixels based on correlated channels. Of course, the locations of the meta-pixels depend on the distribution of correlated channels or rather the projected signal codes. Given the projected codes are constant, we can estimate the location and cluster size of the meta-pixels beforehand. We create a new channel matrix H_{meta} by selecting rows corresponding to the meta-pixels which are small in number and therefore drastically reduce the size of the channel matrix. We use H_{meta} for computing an approximate scene x_{meta} at a fraction of power of the full reconstruction with H .

The approximate scene x_{meta} lacks usable resolution for perception but provides the possible locations of the objects along with a probabilistic score as the confidence of estimate. We use this cue to iteratively zoom in to those probable locations. As we select a smaller region to further reconstruct the scene to improve resolution, we again work with a smaller subset of the columns of H . Every such operation reveals a few more pixels in its vicinity that requires processing for higher resolution mapping. Considering the size and continuity of rigid objects, we ‘search’ these neighborhoods of the occupied pixels with this fractional computing method. Once the iterations do not generate new regions to compute, the algorithm converges and stitches all the fractional maps to a complete reconstructed scene. We have evaluated this technique extensively in section 4.6 & 4.7 that shows the process can cost $100\times$ less power than a single shot reconstruction to produce a map with similar accuracy and resolution.

Note that theoretically, the fractional scene reconstruction process can take more power in cases where the scene is extremely dense. In that case, x_{meta} will indicate all pixels to be potentially occupied and the iterative zooming will end up processing all the pixels in turns. Of course, in practice robots, movement pattern ensures that the field of view is not always obstructed by a dense scene. However, to sidestep such issues we develop a heuristic method based on the power of the total reflected signals to identify an overly dense or rare scene.

Predictive pre-fetching: To further improve the energy efficiency of the scene reconstruction algorithm, we pre-compute and store the inverse of all the coarse-grained and fine-grained measurement matrices. During real-time scene reconstruction, we pre-fetch these inverse matrices based on the coarse location of detected objects. We also minimize the number of pre-fetching operations by predicting the translation of objects in the scene based on the movement of the robot.

Tensor parallelization: Recent deep learning innovations have led to remarkable improvements in hardware accelerations. Several single board computers now have dedicated GPUs to perform tensor operations in real-time [2, 4, 43]. These hardware units can perform multiple floating-point operations in parallel with a single instruction which exponentially increases their energy efficiency. Unfortunately, traditional techniques have been lacking these speedups because of the sequential operations in CPUs. We change our matrix inversion and matrix multiplication operations from the ground up to take advantage of these intrinsic properties of the tensor parallelization. In our *SPiDR* prototype, we leverage the NEON SIMD unit [27] of the Raspberry Pi 4 and convert our operations to vector instructions using the MATLAB coder [86].

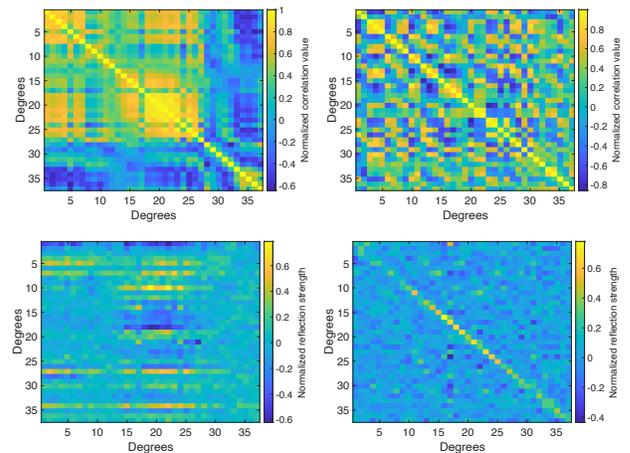


Figure 8: Row 1: Comparison of the correlation of the time domain signals project at different angles with an arbitrary stencil (left) and an optimized stencil (right). Row 2: Comparison of the location detection performance for a small object (1cm wide) placed at different angles from the sensor with an arbitrary stencil (left) and an optimized stencil (right).

3.2 Directional code projection

The accuracy of our map reconstruction or spatial sensing depends on the sensor’s ability to generate a diverse probing signal, such that reflections from each location of the scene carry a unique signature to the receiver. While mechanically rotated directional speakers or beamforming arrays can send different directional signals, they do not meet the latency and power requirements. We aim to use only one common $40kHz$ ultrasound speaker [55] that consumes around 90 micro-Joule of energy for sending a 5 milli-second of pulse for our system. To produce a directionally diverse signal, we use a 3D-printed passive microstructure cap for the speaker, called stencil. We have shown that it is possible to modulate spatially diverse sounds using a carefully designed zero-power, low-complexity, low-cost, and miniature acoustic structure. However, an arbitrary design for the stencil does not produce enough diversity for each pixel location at the target scene for successful mapping. Figure 8 shows the correlation of the signals projected at different directions for an arbitrary stencil and an optimized stencil, along with their scene reconstruction performance. The design of a stencil that can generate diverse signals for each target pixel starts with an optimization process, called *multipath synthesis*.

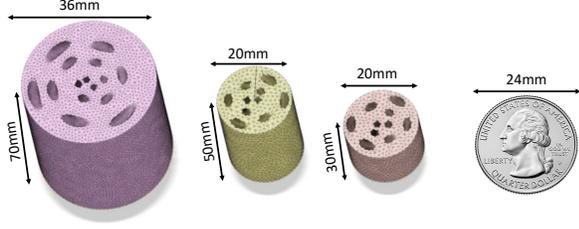


Figure 9: Different sizes of stencils used in our experiments.

Multipath synthesis: We consider the openings on the 3D structure as secondary sound sources and assume a specific combination of the amplitude and phase distributions of these signal sources can result in the desired signal pattern at the object plane. We then estimate the signal phases along the paths between all pair of pixel and the secondary signal sources. Next, we compute the channels due to superimposition of these estimated signals. For N secondary sources, the sound at the m -th location (pixel) is calculated by super imposing the propagations as follows.

$$p_m = i \frac{\rho c k S}{2\pi} \sum_{n=1}^N a_n e^{i2\pi f t} \frac{e^{-i d_{mn}}}{d_{mn}} \quad (1)$$

Here a_n and t are the amplitude and time delay for the n -th signal source, f is the frequency of transmitted signal, d_{mn} is the distance between the n -th source and the m -th location, ρ and c are the density and sound speed in the air respectively. S is the surface area of the reflecting object and we consider this to be a constant unit surface. This process is similar to the well-established Weighted Gerchberg-Saxton (GS) algorithm [28] used in holographic projection. Figure 9 shows the comparative sizes of the metamaterial stencils used in our experiments.

We estimate the channel matrix H with the following equation.

$$H(m, n) = iK \frac{e^{-i d_{mn}}}{d_{mn}} \quad (2)$$

Here K is a constant value equal to $\frac{\rho c k S}{2\pi}$. In theory, this assumption that any arbitrary signal pattern can be projected on the object plane will require a large projection filter with infinitely many openings on the structure. Limiting the size and openings on the 3D structure for practical purposes will restrict the achievable diversity on the object-plane and/or the size of the target pixel. Naturally, the signal wavelength also impacts the achievable diversity for a given pixel size. We optimize these parameters with an algorithm that varies each design parameters in iterations and picks a set of parameters that provide best overall diversity. We attempt to minimize the ‘F-measures’ with different support vectors that indicates the quality of the channel matrix H for map reconstruction [44]. A smaller F value indicates better recovery performance and a value smaller than 1 ensures the matrix H is robust to small environmental noises. We vary the maximum lengths of tubes based on the size of 3D structures to learn the trade-off between maximum tube length and imaging performance. We transmitted 40kHz 10-cycle tone-burst signal. The stencils have 6 tubes with 1.5mm diameter. As shown in Figure 10(a), the F value becomes smaller with the increasing tube length. In practice, we find a stencil of $7 \times 3 \times 3$ cm is required to support a 24cm tube length. The size of 16cm, 8cm, and 4cm tubes are $5 \times 3 \times 3$ cm, $3 \times 3 \times 3$ cm, $1 \times 3 \times 3$ cm, respectively.

The diffraction can also adversely affect the diversity of the received signal. When acoustic signal encounters the edge of the tube with a diameter smaller than the wavelength of signal, it tends to bend or deflect around it limiting overall resolution. In a COMSOL simulation, we vary the diameter of the tubes with the longest tube of length 16cm. As shown in Figure 10(b), the F value remains stable for small variation of the tube diameters.

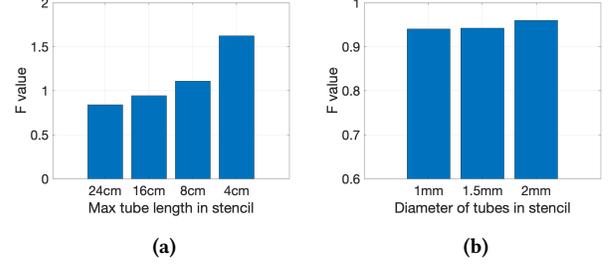


Figure 10: F values with (a) different lengths of tubes inside stencil, (b) different diameters of tubes.

Leveraging frequency diversity: In our measurement model, $y = Hx$, the number of columns in the matrix H indicates the number of unknowns, and the rows are the measurements. More independent measurements compared to the number of unknowns, i.e. a tall and skinny H matrix, leads to better estimation of the unknowns or better reconstruction of the map. Past works in computational photography with similar formulation attempted to increase the number of rows by taking multiple measurements while introducing controlled variations to the scene or mechanically moving the sensor for diversity. None of these approaches are feasible for applications with small robots due to power constraints. However, we overcome this limitation of the passive structure by leveraging multipath diversity over frequency.

Note that the spatial multipath patterns per pixel created by the stencil depend on the length of its internal tubes. However, the phase introduced by these paths depends on the frequency of the signal. This means the stencil that creates a specific spatial pattern for a certain frequency, can create components of completely different phases for another frequency leading to a different spatial code. We build on this intuition to optimize the stencil design jointly for a set of discrete frequencies, such that the projected signals are maximally different at any two locations and also for any two frequencies. We leverage this frequency diversity during scene reconstruction by filtering individual frequencies from the received signal and considering them as individual measurements. We stack these measurements on top of each other to get a new measurement signal y and similarly we stack the channel matrix for different frequencies before proceeding to our scene reconstruction process. Figure 11 shows the reconstruction performance for different individual frequencies and for combined reconstruction.

3.3 Optimal microstructure design

Once the required phase delays are determined through the optimization algorithm, we make a 3D model for the stencil. To achieve mentioned phase delays, we use an automated CAD script to design non-overlapping tubular paths inside the 3D structure.

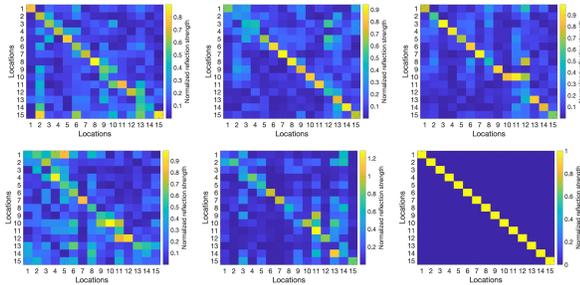


Figure 11: Confusion matrix of imaging accuracy for different frequencies and after combining all frequencies together.

The path lengths of these tubes determine the phase difference of secondary sources, i.e., the sound signals at the output holes of the stencil. The shape of the tubes impacts the propagation of the sound waves through them due to the capillary effects and resonance of the sound column. We experimented with a number of shapes for the tubes and converge on concentric helical paths for the microstructure. Figure 6 shows the improvement of signal diversity distribution on a cross-sectional plane in front of a transducer with the stencil.

Next, we export the design as an STL file and slice it for 3D printing. We used the Elegoo Mars photocuring 3D printer [38] to print the stencils. We use an ultraviolet light-curable resin with $1.195g/cm^3$ 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 [139].

3.4 Motion stacking

So far in our processing, we did not consider the movement of the robot that provides an opportunity for further improving the quality of the generated map. Perception in mobile robots requires continuous mapping of the scene during navigation. Therefore, it provides multiple snapshots where certain parts of the scene are common. It is possible to reorient the snapshots and superimpose them to generate a robust and high-resolution reconstruction of the scene. The concept of superimposition or stacking maps is a popular approach in robotic navigation as well as in photography and imaging. Some common techniques include linear and non-linear filtering [50, 101], optical flow [82], and deep fusion [127]. However, to remain within the power budget, we borrow the technique of mean and median stacking, which runs at $O(n)$ and $O(\log(n))$ complexity without significantly compromising accuracy. *SPiDR* is capable of producing snapshots of the scene at a maximum speed of 170 frames per second, but depending on the speed of movement of the robot even a slow frame rate (e.g., 30 frames per second) provides an opportunity for stacking. We implemented and tested a motion-based stacking algorithm on the cross-sectional depth-maps generated by *SPiDR*. We use the knowledge about the robot’s movement to compensate for the translational changes between consecutive maps before superimposition. Empirical results show that median stacking is capable of fine-grained map generated

along with improvements in outlier removal. Figure 12 shows the performance of stacking for five frames.

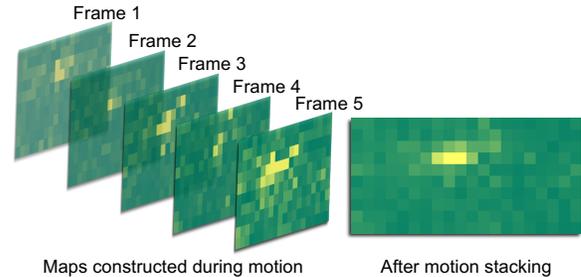


Figure 12: Stacking multiple frames suppresses spurious objects in the scene. Above we show the result of motion stacking of 5 frames taken as the robot moves.

4 EVALUATION

To this end, we implement a prototype of *SPiDR* and perform experiments with different scenes. We use lidar [59] and ultrasound distance sensor [56] for the baseline performance comparison and benchmarking the energy consumption. Next, we elaborate on the experimental setup, followed by the evaluation benchmarks.

4.1 Metrics

We evaluate the performance of *SPiDR* using the following metrics:

- Structural similarity (SSIM) index [131] $SSIM = l(x, y)c(x, y)s(x, y)$, where $l(x, y)$, $c(x, y)$, and $s(x, y)$ are respectively the luminance term, the contrast term, and the structural term calculated based on the local means, standard deviations, and cross-covariance for estimated image x and ground truth image y .
- RMSE error $r = \sqrt{\sum_{i=1}^N \frac{(p_i - \hat{p}_i)^2}{N}}$, where N is the total number of pixels in the ground truth, \hat{p}_i is the binary score (0 or 1) of the pixel, and p_i is the binary score of the ground truth. A score 0 means no object at the pixel, and vice versa.

4.2 Overall performance

In the *SPiDR* prototype, we embed an ultrasound speaker in a 3D-printed stencil and a microphone placed on the top of the stencil. We also compare the performance and the power consumption with lidar and ultrasound distance sensor. The collected data is processed offline using Matlab scripts on a computer. The sound sources are ultrasound 10-cycle tone burst signals with frequencies $38 - 42kHz$ with $1kHz$ apart, with signal strength 40 dB SPL. The size of the stencil is $5 \times 3 \times 2cm$ with internal tubes. We show the images of 3 representative scenes in Figure 13. The real-world scenes are shown in the first row, and the estimated scenes are shown in the second row. We observe that the scenes can be detected with *SPiDR* when up to 70% of the scene is occupied, where the width of the scene is $20cm$.

Figure 14 summarizes the overall accuracy (SSIM) and energy consumption compared with lidar and ultrasound distance sensor. *SPiDR* can produce similar accuracy as lidar but with $400\times$ less power consumption. To compare the power consumption with Radar, we refer to an existing work [30] that achieves similar



Figure 13: Depth-map reconstruction using *SPiDR* for various real-world scenes.

resolution for imaging using a 64-microphone array and 40kHz ultrasound speaker. With 64 microphones, the power consumption is up to 15.7mJ, 19× higher than our work. As shown in existing work [40], a depth camera for short-range scanning, named Primesense Carmine, has 7× less power consumption than Kinect V2, and it still consumes 75mJ energy for one image, 90× higher than *SPiDR*.

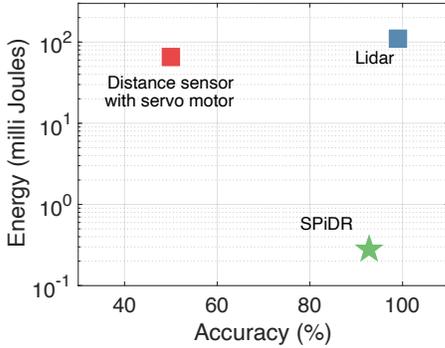


Figure 14: Overall performance of the *SPiDR* compared to Intel Realsense lidar and ultrasound distance sensor mounted on a servo motor. *SPiDR* consumes a fraction of power compared to Lidar and motor based systems while delivering high accuracy in depth-map reconstruction.

4.3 Impact of the environment

Environmental noise: In this experiment, we evaluate the system’s performance under different ambient noise. We played sound clips of noises from various indoor (regular household, mall, and library) and outdoor scenarios (traffic noise). As shown in Figure 15(a), natural ambient noise does not have impact on the performances, with almost constant RMSE error less than 0.2 and SSIM value more than 0.6. It is obvious as audible sounds do not impact the 40kHz ultrasound band *SPiDR* operates in. Next, we test the system with increasingly higher ultrasound noise at the 40kHz band so that Signal-to-Noise ratio (SNR) of the received signal degrades. As shown in Figure 15(b), the system is robust when the SNR is at 60dB. Note that naturally ultrasound noise level in the environment is extremely low and even with an inexpensive speaker with internal noise the SNR is higher than 82dB.

Different locations: We evaluated *SPiDR* in three different locations – indoor laboratory, lobby, and open-air outdoor places. In these experiments, we calibrated only once in the lab environment. Figure 16 shows that the performances in these environments are similar, with an RMS error below 0.2 and SSIM higher than 0.6. This result indicates that the system requires only a one-time calibration to function in arbitrary locations.

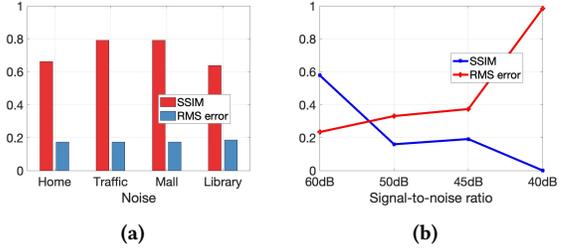


Figure 15: Impact of varying (a) types and (b) levels of noises on depth-map reconstruction.

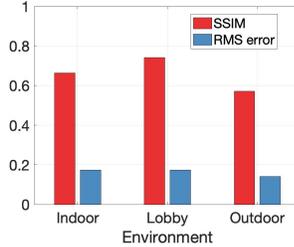


Figure 16: *SPiDR*’s performance in different environments.

4.4 Impact of system parameters

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

Cross-sectional depth-map: The width of the scene is 20cm and the depth is 6 – 15cm. We place the different sizes of blocks and their combinations in different locations of the scene. 5 scenes are collected for each of the sparsity levels. As shown in Figure 17, the RMS error becomes lower and the IoU is higher when the sparsity is larger. The estimated scene is calculated by checking if the coefficient is larger than a threshold 0.3. The pixel with a smaller confidence score than the threshold indicates no object from the corresponding position. The coefficient equals to 1 indicates perfect matching. This means the sparser the scene, the better the performance. The performance of *SPiDR* does not vary significantly when the sparsity of the scene is within 60%. The RMS error is within 0.5cm.

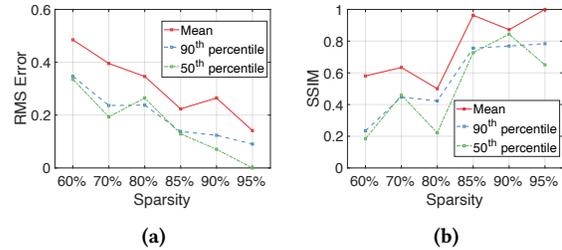


Figure 17: Cross-sectional depth-map reconstruction performance in terms of (a) RMS error and (b) structural similarity, as a function of sparsity of the scene.

Horizontal location: Instead of imaging the 2D scene (width and depth) as 2D imaging, we treat the depth as a constant value and only image the scene in the dimension of width. We show the result of 1D imaging when the distance between the stencil and the object

layer is 10cm. 10 scenes are collected for 1D imaging with vary number of objects and object sizes. The object sizes are within 1-6cm and there are up to three objects. As shown in Figure 18, all the RMS errors for sparsity 70 – 95% are smaller than 0.6cm, and the SSIM is larger than 0.6 when the sparsity level is above 84%.

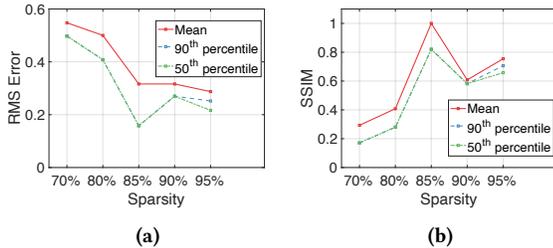


Figure 18: Horizontal localization performance in terms of (a) RMS error and (b) structural similarity, as a function of sparsity of the scene.

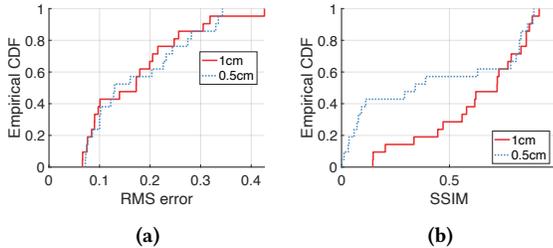


Figure 19: CDF plot of (a) RMS error and (b) structural similarity for depth-map reconstruction at varying resolutions.

Resolution: To test the resolution of *SPiDR*, we calibrate our system in 1cm and 5mm resolutions, respectively, and compare the scene reconstruction accuracy. We imaged 25 scenes with sparsity from 60 – 95% for both resolutions, and show the CDF of the RMS error and SSIM in Figure 19. For both 1cm and 5mm resolutions, all the RMS errors are within 0.5cm and 50% of the scenes have SSIM larger than 0.75. As a visual example shown in Figure 20, the performances of these two resolutions are similar, meaning our system is able to detect 5mm-wide tiny objects.



Figure 20: Scene reconstruction results at 1cm and 0.5cm resolutions. We modify the number of columns in the channel matrix to have a higher resolution.

4.5 Impact of scene parameters

Materials/Reflectivity: The performance may vary with different types of materials since the reflection rates of sound are different. We calibrate our system with 3D-printed resin blocks and test the performance with objects made of cardboard, wood, and resin. We image 5 scenes with a sparsity 70% – 95% for each kind of material. As shown in Figure 21(a), the RMS errors for all the materials are

below 0.3cm, which indicates *SPiDR* is robust to common types of materials.

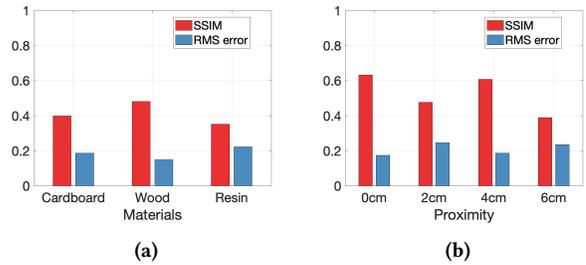


Figure 21: The depth-map reconstruction performance for (a) different materials and (b) different proximity between two objects.

Proximity: In this experiment, we evaluate the system’s performance when the distance between two objects varies. From the experiment with an ultrasound distance sensor, we find the distance sensor cannot detect the disconnection of two objects when the distance between them is within 3cm. We test the system when the distances between two objects are 0,2,4,6cm, respectively. We show the quantitative scene reconstruction accuracy in Figure 21(b). Figure 22 shows the output depth-map for varying gaps. Results indicate that *SPiDR* can detect open spaces as narrow as 2cm.

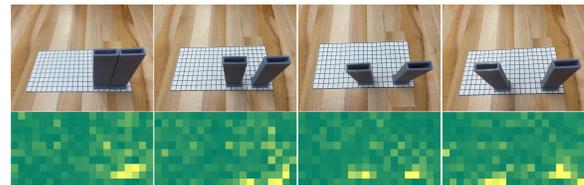


Figure 22: Depth-map reconstruction outputs for varying proximity between objects.

Depth: To evaluate the performance when the object is at different depths, we put a 3cm width object in depths 6,9,11,13cm, respectively. The first line of Figure 23 includes the ground truth of the scenes, and the second line shows the estimated images. The image becomes blurred when the distance is farther. But the location of the object is still accurate. As the system moves closer to the object, the image becomes clearer. The distance of the object from the sensor decreases the reflected signal strength which impacts the overall accuracy. The quantitative result of scene reconstruction accuracy with different depths is shown in Figure 24. The RMS error for all the depths is within 0.6cm and the SSIM is larger than 0.6 when the distance is within 11cm.

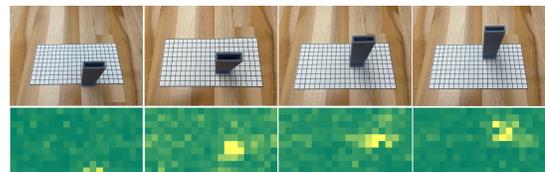


Figure 23: Depth-map reconstruction outputs of different depths of the objects.

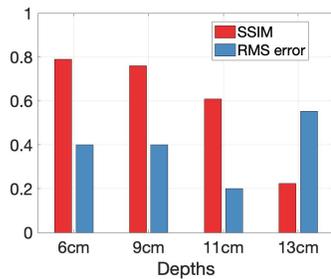


Figure 24: Performance of depth-map reconstruction with different depths of the objects.

4.6 Computation techniques

Frequency and Motion stacking: As introduced in Section 3, frequency stacking and motion stacking can increase the performance accuracy of scene reconstruction. Figure 25 shows an example of the stacking operations. The ground truth is a 3cm wide object located in the front middle of the scene. Before stacking, the object is recognized as two 1cm objects, and the figure has some other pixels mistakenly detected as 1cm-wide objects. After frequency stacking or motion stacking, only one object at the same location is detected, but the size is 1cm smaller or larger. If with both stacking approaches, the size and location of the object are exactly the same as the ground truth. Moreover, no points are mistakenly imaged.

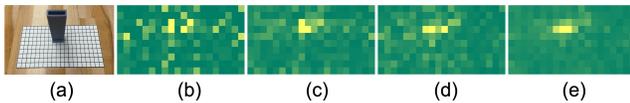


Figure 25: Scene reconstruction of a horizontal bar with and without frequency or motion stacking. (a) Ground truth, (b) Raw output, (c) Only motion stacked, (d) Only frequency stacked, (e) Both motion and frequency stacked.

Fractional scene observation: We propose fractional scene observation to achieve low-power scene reconstruction, as well as achieving comparable scene reconstruction accuracy. In Figure 26, we show two examples of scene reconstruction with one or multiple projects. We observe that the results of with and without fractional computing are comparable to each other. At the same time, we can reduce energy by 85%, as elaborated in the evaluation of the power consumption. We also evaluate the performance with different fraction rates, as shown in Figure 27. Naturally, the accuracy decreases with the increase of the fraction rate. In our experiments, usually a 3cm region shows a higher similarity between the signals. When the fractions become smaller, it fails to properly represent the whole scene with the H matrix after integration. This can causes error in the detection of the “region of interest”.

4.7 Power consumption

In this section, we develop the ultra-low-power prototype of *SPiDR* and benchmark its power consumption. Figure 28 shows the prototype of the setup used for evaluation. We compare its performance and energy efficiency with Intel Realsense lidar [60]

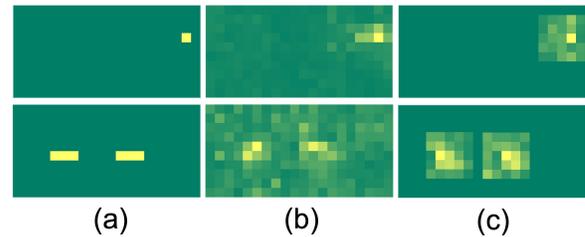


Figure 26: The scene reconstruction with and without the fractional computing method. (a) Ground truth, and the scene reconstruction (a) without and (b) with fractional computing.

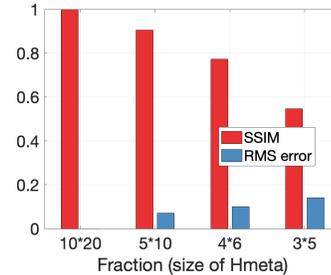


Figure 27: Performance with different sizes of H_{meta} .

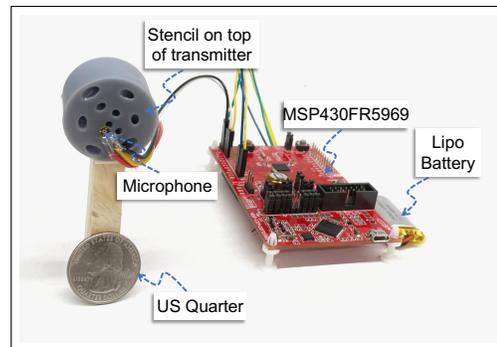


Figure 28: *SPiDR* prototype for power evaluation.

and an ultrasound distance sensor [56]. We measure the power consumption of each sub-module in our system, including hardware frontend, analog-to-digital conversion, and image computation. Table 1 shows the breakdown of power consumption for different hardware and software sub-modules. *SPiDR* consumes a total 9.9mW of power and runs at a rate of 12 depth-maps per second. This results in 0.83mJ of energy per depth-map, which means our system can continuously run for 3 days on a common 3volts CR2032 coin cell battery [65].

Submodule	Energy consumed
Transducer (Hardware)	0.035mJ
ADC (Hardware)	0.024mJ
Microphone (Hardware)	0.245mJ
Computation (Software)	0.526mJ
Total	0.83mJ

Table 1: Breakdown of energy consumption for the hardware and software submodules.

Ultra-low-power implementation: Hardware is a bottleneck for energy consumption in most of the depth imaging techniques, like lidar and distance sensors mounted on motors. Our focus is to reduce the power consumption of the hardware front-end without losing out the accuracy. We develop *SPiDR* using TI-MSP430FR5969 MCU [58], an off-the-shelf 40KHz ultrasound transducer [55], and the ADMP401 microphone [1]. We use a piezo-crystal based ultrasound transducer which has a narrow resonating bandwidth at 38 – 42kHz. We eliminate the need for a Digital-to-analog converter (DAC) by transmitting a square wave to the transducer. The impedance matching at the resonance frequency acts like a bandpass filter which converts the signal to a narrowband sinusoid. The transducer has extremely high efficiency when transmitting and receiving at the resonance frequency. We use ultra-low-power sigma-delta ADC of the MSP430 MCU to sample the received signal. The power consumption during sampling and computation is shown in Figure 29. We noted that the energy consumed by the ADMP401 microphone was 49mW so we power up the microphone only for the duration of sampling of signal, which reduces the average power per frame to 0.245mJ. We believe that better engineered and low-power pre-amplifiers [57] can further reduce the power consumption of the system. Here we have reported the prototype power consumption from commercially available components.

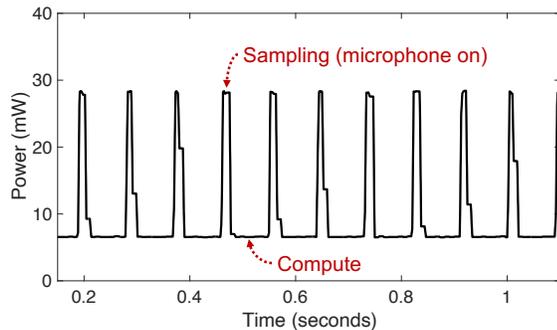


Figure 29: Power consumption during sampling and computation.

Computational optimization: To measure the power consumption of the computation module with different fractional computing optimizations we implement the system on the MSP430 with varying sizes of H_{meta} and record the power consumption using the Texas Instruments EnergyTrace++ tool [126]. We run the system to compute scenes using different prefetched matrices of different sizes and record the energy consumed using the Energytrace tool. Figure 30 shows the average energy consumed per scene reconstruction for four different fractional computing ratios. We see that the energy drops exponentially with the decrease in size of H_{meta} . We find that 4×6 is the best size based on the trade-off between SSIM score (shown in Fig 27) and energy consumed. To compare the power consumption of the prefetching and fractional optimization we port the codes to a Raspberry Pi 4 [43] and record the voltage and current traces using a Keysight E6313A programmable DC power supply [71]. We write the software for *SPiDR* in Matlab and cross-compile it

Computation optimization	Energy overhead	Gain
None	388.65mJ	1×
Only Fractional	9mJ	40×
Only Prefetching	1.42mJ	270×
Both	0.28mJ	1380×

Table 2: Comparison of different computation optimizations showing the total energy consumed per scene reconstruction. Prototype on Raspberry Pi 4.

for a Raspberry pi 4 using Matlab Coder [86]. We use the NEON SIMD unit [27] of the Raspberry Pi 4 and export an executable C file. We run the file for 10,000 iterations and record the voltage and current traces using the power supply. We calculate the overhead of the computation module by subtracting the idle CPU power consumption. Table 2 shows the energy overhead with different computation optimizations.

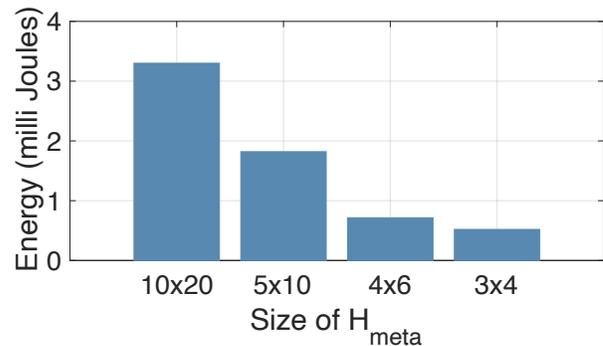


Figure 30: Comparison of power consumption with different sizes of H_{meta} . Prototype on MSP430FR5969.

5 RELATED WORK

Computational sensing has been an active field of research for over a decade now and is rich in foundational works [32] that paved the ways for applications like autonomous navigation [120], contactless perception and monitoring [80], human-machine interaction [84], and precision measurements [5]. Recent works explore acoustic sensing in the context of ubiquitous computing [21, 93, 106–110, 119, 137]. *SPiDR*, while building on the advancements in this field, takes a systems approach to develop a method for low-power acoustic perception leveraging sound’s interaction with passive microstructures. We discuss the ideas closely related to our work.

• **Scene perception and object avoidance:** Seminal work in object detection using image processing [83, 132, 142], Radar [3, 6, 14, 140], lidar [12, 113, 129], and ultrasound [8, 42, 47, 67, 81, 136] have significantly advanced the field of study. Since the dominant physical effects change as scale reduces, conventional object avoidance techniques such as electromagnetic motors and robot navigation algorithms cannot be applied on microrobots. Recent works [10, 133] use magnetic particle imaging to detect objects in front of a microrobot in phantom. Another underwater snake robot detects obstacles using a camera and computer vision algorithm [69].

- **Low-power sensing:** Low-power sensing has been applied on many emerging applications, such as monitoring human body motions [23, 130], chemical sensing [39, 96, 122], and health monitoring [17, 77, 98, 100]. Many signal processing techniques are designed for low-power sensing [62, 92]. Low-power long-range RFID technologies [48, 104] are developed to improve pervasive sensing. Low-power RF energy harvesting transceiver [72] is also designed for various sensing applications.

- **Acoustic structures for sensing:** The art of using structure to manipulate sound waves is known for a very long time and its application is seen on many ancient architectures [138]. However, the study of 3D-printed acoustic microstructures in ubiquitous sensing applications is relatively new. Li et al. [79] build acoustic filters to control the impedance of discrete frequencies. Some other works [114, 135] develop physical user interfaces by 3D printing acoustic structures. These methods change the structure shapes to create unique frequency responses and use smartphone microphones for data collection and classification.

Our previous work, called Owlet [45, 46], shows the possibility of using acoustic microstructure to embed directional clues to the signal recorded by a microphone, targeting applications like sound's direction-of-arrival estimation and localization. Owlet and *SPiDR* share the same high-level idea that a carefully designed passive structure can induce spatial diversity to sound. However, these two systems fundamentally develop different principles. First, Owlet introduces structural-assisted diversity to the received signal, where *SPiDR* uses it for projecting a modulated output signal. Owlet detects the angle of arrival of incoming sounds, while *SPiDR* produce a depth image of the scene.

- **Sensing with limited sensors:** Several works attempt to find alternative ways for spatial sensing without using a sensor array. A few recent works [31, 36, 37, 115] use structures like walls, legos, and cubes around a microphone to produce scattering. Similarly, [117] leverages the reflections from nearby walls to find the direction-of-arrivals of sounds and localize them. *SPiDR*, on the other hand, explores the transmitter diversity and spatially coded probing signals and uses statistical algorithms to find the directionality of the reflected signals. Coded apertures are designed by [19, 85, 99] to entail different light directions or different frequency bands for different locations in the scene. Several studies [15, 75, 90, 134] generated coded aperture to generate predefined acoustic fields that provide unique ultrasound measurements for each location in the scene. However, these works require MHz level frequency for imaging. To achieve diversity in the microphone responses, past works aggregated microphone techniques [87–89] use a directional microphone and mode beamforming techniques [112, 141] use vector sensors embedded in a larger object with known modes of vibration. Micromachined Ultrasound Transducers (MUT) provide ways to develop compact spatial sensors for high-frequency (several MHz) sounds [25, 35, 102]. Some other works leverage ambient channel responses [105, 118], frequency patterns of the sound [116], array interpolation [54, 66, 68], compressive sensing [41, 53, 95], and repurposing of acoustic vector sensor [78, 94, 124].

6 DISCUSSION

We mention below a few points of discussion regarding the technique and scopes for improvements.

- **Speed of the robot.** Speed of the movement can have a trade-off with the imaging latency. While motion stacking gives us a chance to produce more accurate depth-map, it also adds a limitation on the robot's speed. With 1cm resolution, our motion stacking algorithm expects the robot to move less than 1cm per frame for best performance. Since our current prototype can produce 170 depth-maps per second, it allows the robot to move at a maximum speed of 170cm/s (6.12km/h) without any significant impact on depth-maps recovery.

- **Applicable micro robots.** Existing micro-robots [49, 51, 61, 111] have power budget of 10–30mW for sensing and computation. Due the limited options of sensors available in this power budget, most micro-robots use proximity sensors (IR-based and ultrasonic-based) for their navigation. Proximity sensors can only detect the objects in front without motor for rotation. Vision based sensors [7, 26] have low power consumption for hardware, but they are computationally expensive, requiring many orders of power than the available budget. These micro-robots can utilize *SPiDR* for low-power and smarter navigation. Physical form factor is another consideration. Since the size of stencil is within 5cm and can be as small as 2cm, all the robots in cm-level [9, 16, 52, 121] and larger than that [13, 73, 97, 103, 123] can apply *SPiDR* on them. Since *SPiDR* is ultrasound-based, it can be applied in harsh environments, including low illumination, smoke, and opaque obstructions.

- **Interference from out of the region.** We defined the region of interest by carefully considering the properties of the signal. The frequency of the sound signal and its attenuation in the medium works as a natural filter for interference. We find that the signal strength of the reflected signal from out of the region is negligible. However, in theory, since we use at least 3.5ms of time series data for image recovery, unwanted reflections within 60cm of distance can interfere with the system. Time gating can be a potential solution. However, since our system depends on superimposition of signals for performance, it is difficult to separate the signal reflected from in and out of the region simply by time gating. We keep this as a future work.

7 CONCLUSION

This paper presents *SPiDR*, a low-power spatial sensing system for micro-robot navigation. We use only one ultrasound speaker/microphone pair and a 3D-printed passive structure to project diversity in the scene. *SPiDR* consumes only 10mW of power to produce precise depth-map at a rate of 12 frames per second. The core idea of structure-assisted acoustic sensing opens up new possibilities of low-power perception, obstacle detection, and navigation.

8 ACKNOWLEDGMENTS

We sincerely thank our shepherd, Prof. Inseok Hwang, and the anonymous reviewers for their insightful comments and suggestions.

REFERENCES

- [1] Mems microphone breakout. <http://www.analog.com/media/en/technical-documentation/obsolete-data-sheets/ADMP401.pdf>. Last accessed 24 September 2017.
- [2] Adafruit. Adafruit edgebadge datasheet. https://media.digikey.com/pdf/Data%20Sheets/Adafruit%20PDFs/4400_Web.pdf, 2021.
- [3] Fadel Adib, Zach Kabelac, Dina Katabi, and Robert C Miller. 3d tracking via body radio reflections. In *11th {USENIX} Symposium on Networked Systems Design and Implementation (NSDI) 14*, pages 317–329, 2014.
- [4] Apollo. Apollo3 blue datasheet. https://cdn.sparkfun.com/assets/c/1/b/7/6/Apollo3_Blue_MCU_Data_Sheet_v0_10_0.pdf, 2021.
- [5] Ehsan Ardestanizadeh, Mahdi Cheraghchi, and Amin Shokrollahi. Bit precision analysis for compressed sensing. In *2009 IEEE International Symposium on Information Theory*, pages 1–5. IEEE, 2009.
- [6] Serdal Ayhan, Mario Pauli, Thorsten Kayser, Steffen Scherr, and Thomas Zwick. Fmcw radar system with additional phase evaluation for high accuracy range detection. In *2011 8th European Radar Conference*, pages 117–120. IEEE, 2011.
- [7] Stanley S Baek, Fernando L Garcia Bermudez, and Ronald S Fearing. Flight control for target seeking by 13 gram ornithopter. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2674–2681. IEEE, 2011.
- [8] Yang Bai, Li Lu, Jerry Cheng, Jian Liu, Yingying Chen, and Jiadi Yu. Acoustic-based sensing and applications: A survey. *Computer Networks*, 181:107447, 2020.
- [9] Andrew T Baisch, Christian Heimlich, Michael Karpelson, and Robert J Wood. Hamr3: An autonomous 1.7 g ambulatory robot. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5073–5079. IEEE, 2011.
- [10] Anna C Bakenecker, Anselm von Gladiss, Hannes Schwenke, André Behrends, Thomas Friedrich, Kerstin Lüdtke-Buzug, Alexander Neumann, Joerg Barkhausen, Franz Wegner, and Thorsten M Buzug. Navigation of a magnetic micro-robot through a cerebral aneurysm phantom with magnetic particle imaging. *Scientific reports*, 11(1):1–12, 2021.
- [11] David K Barton. Modern radar system analysis. *Norwood*, 1988.
- [12] Jorge Beltrán, Carlos Guindel, Francisco Miguel Moreno, Daniel Cruzado, Fernando Garcia, and Arturo De La Escalera. Birdnet: a 3d object detection framework from lidar information. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pages 3517–3523. IEEE, 2018.
- [13] P. Birkmeyer and R. S. Fearing. Dash: A resilient high-speed 15g hexapedal robot. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 418–419, 2009.
- [14] Tara Boroushaki, Junshan Leng, Ian Clester, Alberto Rodriguez, and Fadel Adib. Robotic grasping of fully-occluded objects using rf perception. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 923–929. IEEE, 2021.
- [15] MD Brown, DI Nikitichev, BE Treeby, and BT Cox. Generating arbitrary ultrasound fields with tailored optoacoustic surface profiles. *Applied Physics Letters*, 110(9):094102, 2017.
- [16] Remo Brühwiler, Benjamin Goldberg, Neel Doshi, Onur Ozcan, Noah Jafferis, Michael Karpelson, and Robert J Wood. Feedback control of a legged microrobot with on-board sensing. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5727–5733. IEEE, 2015.
- [17] Nam Bui, Nhat Pham, Hoang Truong, Phuc Nguyen, Jianliang Xiao, Robin Deterding, Thang Dinh, and Tam Vu. ebp: Frequent and comfortable blood pressure monitoring from inside human’s ears. *GetMobile: Mobile Computing and Communications*, 23(4):34–38, 2020.
- [18] Gilles Caprari, Thomas Estier, and Roland Siegwart. Fascination of down scaling-alice the sugar cube robot. In *IEEE International Conference on Robotics and Automation (ICRA 2000): Workshop on Mobile Micro-Robots*, number CONF, 2000.
- [19] E Caroli, JB Stephen, G Di Cocco, L Natalucci, and A Spizzichino. Coded aperture imaging in x-and gamma-ray astronomy. *Space Science Reviews*, 45(3):349–403, 1987.
- [20] Jaebum Choi, Simon Ulbrich, Bernd Lichte, and Markus Maurer. Multi-target tracking using a 3d-lidar sensor for autonomous vehicles. In *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, pages 881–886. IEEE, 2013.
- [21] Romit Roy Choudhury. Earable computing: A new area to think about. In *Proceedings of the 22nd International Workshop on Mobile Computing Systems and Applications*, pages 147–153, 2021.
- [22] David L Christensen, Elliot W Hawkes, Srinivasan A Suresh, Karen Ladenheim, and Mark R Cutkosky. μ tugs: Enabling microrobots to deliver macro forces with controllable adhesives. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4048–4055. IEEE, 2015.
- [23] Gabe Cohn, Sidhant Gupta, Tien-Jui Lee, Dan Morris, Joshua R Smith, Matthew S Reynolds, Desney S Tan, and Shwetak N Patel. An ultra-low-power human body motion sensor using static electric field sensing. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 99–102, 2012.
- [24] John C Curlander and Robert N McDonough. *Synthetic aperture radar*, volume 11. Wiley, New York, 1991.
- [25] David E Dausch, John B Castellucci, Derrick R Chou, and Olaf T Von Ramm. Theory and operation of 2-d array piezoelectric micromachined ultrasound transducers. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 55(11):2484–2492, 2008.
- [26] Christophe De Wagter, Sjoerd Tijmons, Bart DW Remes, and Guido CHE de Croon. Autonomous flight of a 20-gram flapping wing mav with a 4-gram onboard stereo vision system. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4982–4987. IEEE, 2014.
- [27] ARM Developer. Optimizing c code with neon intrinsics. <https://developer.arm.com/documentation/102467/0100/Matrix-multiplication-example>, 2021.
- [28] Roberto Di Leonardo, Francesca Ianni, and Giancarlo Ruocco. Computer generation of optimal holograms for optical trap arrays. *Optics Express*, 15(4):1913–1922, 2007.
- [29] Ivan Dokmanić, Reza Parhizkar, Andreas Walther, Yue M Lu, and Martin Vetterli. Acoustic echoes reveal room shape. *Proceedings of the National Academy of Sciences*, 110(30):12186–12191, 2013.
- [30] Ivan Dokmanić and Ivan Tashev. Hardware and algorithms for ultrasonic depth imaging. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6702–6706. Ieee, 2014.
- [31] Ivan Dokmanić and Martin Vetterli. Room helps: Acoustic localization with finite elements. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2617–2620. Ieee, 2012.
- [32] David L Donoho. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006.
- [33] Marco F Duarte, Mark A Davenport, Dharmal Takhar, Jason N Laska, Ting Sun, Kevin F Kelly, and Richard G Baraniuk. Single-pixel imaging via compressive sampling. *IEEE signal processing magazine*, 25(2):83–91, 2008.
- [34] Julien Dupeyroux, Julien R Serres, and Stéphane Viollet. Antbot: A six-legged walking robot able to home like desert ants in outdoor environments. *Science Robotics*, 4(27), 2019.
- [35] P-C Eccardt, K Niederer, T Scheiter, and C Hierold. Surface micromachined ultrasound transducers in cmos technology. In *1996 IEEE Ultrasonics Symposium. Proceedings*, volume 2, pages 959–962. IEEE, 1996.
- [36] Dalia El Badawy and Ivan Dokmanić. Direction of arrival with one microphone, a few legs, and non-negative matrix factorization. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 26(12):2436–2446, 2018.
- [37] Dalia El Badawy, Ivan Dokmanić, and Martin Vetterli. Acoustic doa estimation by one unsophisticated sensor. In *International Conference on Latent Variable Analysis and Signal Separation*, pages 89–98. Springer, 2017.
- [38] Elegoo. <https://www.elegoo.com/product/elegoo-mars-uv-photocuring-lcd-3d-printer/>, 2020.
- [39] I Elmi, S Zampolli, E Cozzani, M Passini, GC Cardinali, and M Severi. Development of ultra low power consumption hotplates for gas sensing applications. In *SENSORS, 2006 IEEE*, pages 243–246. IEEE, 2006.
- [40] Péter Fankhauser, Michael Bloesch, Diego Rodriguez, Ralf Kaestner, Marco Hutter, and Roland Siegwart. Kinect v2 for mobile robot navigation: Evaluation and modeling. In *2015 International Conference on Advanced Robotics (ICAR)*, pages 388–394. IEEE, 2015.
- [41] Efrén Fernandez-Grande and Angeliki Xenaki. Compressive sensing with a spherical microphone array. *The Journal of the Acoustical Society of America*, 139(2):EL45–EL49, 2016.
- [42] Dariush Forouher, Marvin Große Besselmann, and Erik Maehle. Sensor fusion of depth camera and ultrasound data for obstacle detection and robot navigation. In *2016 14th international conference on control, automation, robotics and vision (ICARCV)*, pages 1–6. IEEE, 2016.
- [43] Raspberry Pi Foundation. Raspberry pi 4. <https://www.raspberrypi.org/products/raspberry-pi-4-model-b/>, 2021.
- [44] Jean-Jacques Fuchs. Recovery of exact sparse representations in the presence of bounded noise. *IEEE Transactions on Information Theory*, 51(10):3601–3608, 2005.
- [45] Nakul Garg, Yang Bai, and Nirupam Roy. Owllet: 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.
- [46] Nakul Garg, Yang Bai, and Nirupam Roy. Demo: Microstructure-guided spatial sensing for low-power iot. In *The 19th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys '21)*, 2021.
- [47] Nakul Garg and Nirupam Roy. Enabling self-defense in small drones. In *Proceedings of the 21st International Workshop on Mobile Computing Systems and Applications*, pages 15–20, 2020.
- [48] Nicolas Gay and Wolf-Joachim Fischer. Ultra-low-power rfid-based sensor mote. In *SENSORS, 2010 IEEE*, pages 1293–1298. IEEE, 2010.
- [49] Benjamin Goldberg, Raphael Zufferey, Neel Doshi, Elizabeth Farrell Helbling, Griffin Whittredge, Mirko Kovac, and Robert J Wood. Power and control autonomy for high-speed locomotion with an insect-scale legged robot. *IEEE Robotics and Automation Letters*, 3(2):987–993, 2018.
- [50] Rafael C Gonzalez, Richard E Woods, et al. Digital image processing, 2002.

- [51] E Farrell Helbling, Sawyer B Fuller, and Robert J Wood. Altitude estimation and control of an insect-scale robot with an onboard proximity sensor. In *Robotics Research*, pages 57–69. Springer, 2018.
- [52] Seth Hollar, Anita Flynn, Colby Bellew, and KSJ Pister. Solar powered 10 mg silicon robot. In *The Sixteenth Annual International Conference on Micro Electro Mechanical Systems, 2003. MEMS-03 Kyoto. IEEE*, pages 706–711. IEEE, 2003.
- [53] Xun Huang. Compressive sensing and reconstruction in measurements with an aerospace application. *AIAA journal*, 51(4):1011–1016, 2013.
- [54] Per Hyberg, Magnus Jansson, and Björn Ottersten. Array interpolation and bias reduction. *IEEE Transactions on Signal Processing*, 52(10):2711–2720, 2004.
- [55] HiLetgo Inc. Hiletgo tct40-16r/t 16mm split ultrasonic sensor ultrasonic transceiver. <https://www.amazon.com/HiLetgo-Ultrasonic-Sensor-Transceiver-Receiver/>, 2021.
- [56] Parallax Inc. Ping ultrasonic distance sensor. <https://www.parallax.com/product/ping-ultrasonic-distance-sensor/>, 2021.
- [57] Texas Instruments. Dual rail-to-rail low-voltage low power operational amplifier. <https://www.ti.com/product/TLV2252>, 2021.
- [58] Texas Instruments. Texas instruments msp430fr5969. <https://www.ti.com/tool/MSP-EXP430FR5969>, 2021.
- [59] Intel. Intel realSense lidar camera l515. <https://www.intelrealsense.com/lidar-camera-l515/>, 2021.
- [60] Intel. Intel realSense lidar camera l515. <https://www.intelrealsense.com/lidar-camera-l515/>, 2021.
- [61] Vikram Iyer, Ali Najafi, Johannes James, Sawyer Fuller, and Shyamnath Gollakota. Wireless steerable vision for live insects and insect-scale robots. *Science robotics*, 5(44), 2020.
- [62] Vikram Iyer, Rajalakshmi Nandakumar, Anran Wang, Sawyer B Fuller, and Shyamnath Gollakota. Living iot: A flying wireless platform on live insects. In *The 25th Annual International Conference on Mobile Computing and Networking*, pages 1–15, 2019.
- [63] Noah T Jafferis, E Farrell Helbling, Michael Karpelson, and Robert J Wood. Untethered flight of an insect-sized flapping-wing microscale aerial vehicle. *Nature*, 570(7762):491–495, 2019.
- [64] Ingmar Jager, Richard Heusdens, and Nikolay D Gaubitch. Room geometry estimation from acoustic echoes using graph-based echo labeling. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE, 2016.
- [65] Mathias Jensen. Coin cells and peak current draw. *White Paper. Texas Instruments*, 2010.
- [66] Ryoga Jinzai, Kouei Yamaoka, Mitsuo Matsumoto, Takeshi Yamada, and Shoji Makino. Microphone position realignment by extrapolation of virtual microphone. In *2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pages 367–372. IEEE, 2018.
- [67] Dong-Won Jung, Zhong-Soo Lim, Byung-Geuk Kim, and Nak-Ku Kim. Multi-channel ultrasonic sensor system for obstacle detection of the mobile robot. In *2007 International Conference on Control, Automation and Systems*, pages 2347–2351. IEEE, 2007.
- [68] Hiroki Katahira, Nobutaka Ono, Shigeki Miyabe, Takeshi Yamada, and Shoji Makino. Virtually increasing microphone array elements by interpolation in complex-logarithmic domain. In *21st European Signal Processing Conference (EUSIPCO 2013)*, pages 1–5. IEEE, 2013.
- [69] Eleni Kelasidi, Signe Moe, Kristin Pettersen, Anna M Kohl, Pål Liljebäck, Jan Tommy Gravdahl, et al. Path following, obstacle detection and obstacle avoidance for thrusted underwater snake robots. *Frontiers in Robotics and AI*, 6:57, 2019.
- [70] Winthrop N Kellogg. Sonar system of the blind. *Science*, 137(3528):399–404, 1962.
- [71] Keysight. Keysight e6313a power supply. <https://www.keysight.com/us/en/assets/9018-04576/user-manuals/9018-04576.pdf>, 2021.
- [72] Young-Joon Kim, Hansraj S Bhamra, Jithin Joseph, and Pedro P Irazoqui. An ultra-low-power rf energy-harvesting transceiver for multiple-node sensor application. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 62(11):1028–1032, 2015.
- [73] Bernhard Klaassen and Karl L Paap. Gmd-snake2: a snake-like robot driven by wheels and a method for motion control. In *Proceedings 1999 IEEE International Conference on Robotics and Automation (Cat. No. 99CH36288C)*, volume 4, pages 3014–3019. IEEE, 1999.
- [74] Nikita Korobov, Oleg Shipitko, Ivan Konovalenko, Anton Grigoryev, and Marina Chukalina. Swap-c based comparison of onboard computers for unmanned vehicles. In *Proceedings of 14th International Conference on Electromechanics and Robotics "Zavalishin's Readings"*, pages 573–583. Springer, 2020.
- [75] Pieter Kruizinga, Pim van der Meulen, Andrejs Fedjajevs, Frits Mastik, Geert Springeling, Nico de Jong, Johannes G Bosch, and Geert Leus. Compressive 3d ultrasound imaging using a single sensor. *Science advances*, 3(12):e1701423, 2017.
- [76] Benjamin Langmann, Klaus Hartmann, and Otmar Loffeld. Depth camera technology comparison and performance evaluation. In *ICPRAM (2)*, pages 438–444, 2012.
- [77] Yoonmyung Lee, David Blaauw, and Dennis Sylvester. Ultralow power circuit design for wireless sensor nodes for structural health monitoring. *Proceedings of the IEEE*, 104(8):1529–1546, 2016.
- [78] Dovid Levin, Emanuel AP Habets, and Sharon Gannot. Maximum likelihood estimation of direction of arrival using an acoustic vector-sensor. *The Journal of the Acoustical Society of America*, 131(2):1240–1248, 2012.
- [79] Dingzeyu Li, David IW Levin, Wojciech Matusik, and Changxi Zheng. Acoustic voxels: computational optimization of modular acoustic filters. *ACM Transactions on Graphics (TOG)*, 35(4):1–12, 2016.
- [80] Lixiang Li, Guoqian Wen, Zeming Wang, and Yixian Yang. Efficient and secure image communication system based on compressed sensing for iot monitoring applications. *IEEE Transactions on Multimedia*, 22(1):82–95, 2019.
- [81] Jongil Lim, Seokju Lee, Girma Tewolde, and Jaerock Kwon. Indoor localization and navigation for a mobile robot equipped with rotating ultrasonic sensors using a smartphone as the robot's brain. In *2015 IEEE International Conference on Electro/Information Technology (EIT)*, pages 621–625. IEEE, 2015.
- [82] Wen-Yan Lin, Siying Liu, Yasuyuki Matsushita, Tian-Tsong Ng, and Loong-Fah Cheong. Smoothly varying affine stitching. In *CVPR 2011*, pages 345–352. IEEE, 2011.
- [83] Xinzhu Ma, Zhihui Wang, Haojie Li, Pengbo Zhang, Wanli Ouyang, and Xin Fan. Accurate monocular 3d object detection via color-embedded 3d reconstruction for autonomous driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6851–6860, 2019.
- [84] Tomás Mantecón, Ana Mantecón, Carlos R del Blanco, Fernando Jaureguizar, and Narciso García. Hand-gesture-based human-machine interface system using compressive sensing. In *2015 International Symposium on Consumer Electronics (ISCE)*, pages 1–2. IEEE, 2015.
- [85] Roummel F Marcia, Zachary T Harmany, and Rebecca M Willett. Compressive coded aperture imaging. In *Computational Imaging VII*, volume 7246, page 72460G. International Society for Optics and Photonics, 2009.
- [86] Mathworks. Mathworks matlab coder. <https://www.mathworks.com/products/matlab-coder.html>, 2021.
- [87] Mitsuharu Matsumoto and Shuji Hashimoto. Multiple signal classification by aggregated microphones. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, 88(7):1701–1707, 2005.
- [88] Mitsuharu Matsumoto and Shuji Hashimoto. Blind source separation of anechoic mixtures in time domain utilizing aggregated microphones. In *2006 14th European Signal Processing Conference*, pages 1–5. IEEE, 2006.
- [89] Mitsuharu Matsumoto and Shuji Hashimoto. A miniaturized adaptive microphone array under directional constraint utilizing aggregated microphones. *The Journal of the Acoustical Society of America*, 119(1):352–359, 2006.
- [90] Kai Melde, Andrew G Mark, Tian Qiu, and Peer Fischer. Holograms for acoustics. *Nature*, 537(7621):518–522, 2016.
- [91] Patrick Kapita Mvumba, Simon Kidiamboko Guwa Gua Band, Aime Lay-Ekuakille, and Nicola Ivan Giannoccaro. Advanced acoustic sensing system on a mobile robot: design, construction and measurements. *IEEE Instrumentation & Measurement Magazine*, 21(2):4–9, 2018.
- [92] Rajalakshmi Nandakumar, Vikram Iyer, and Shyamnath Gollakota. 3d localization for sub-centimeter sized devices. In *Proceedings of the 16th ACM Conference on Embedded Networked Sensor Systems*, pages 108–119, 2018.
- [93] Rajalakshmi Nandakumar, Vikram Iyer, Desney Tan, and Shyamnath Gollakota. Fingero: Using active sonar for fine-grained finger tracking. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 1515–1525, 2016.
- [94] Arye Nehorai and Eytan Paldi. Acoustic vector-sensor array processing. *IEEE Transactions on signal processing*, 42(9):2481–2491, 1994.
- [95] Fangli Ning, Jingang Wei, Lianfang Qiu, Hongbing Shi, and Xiaofan Li. Three-dimensional acoustic imaging with planar microphone arrays and compressive sensing. *Journal of Sound and Vibration*, 380:112–128, 2016.
- [96] P Offermans, HD Tong, CJM Van Rijn, P Merken, SH Brongersma, and M Crego-Calama. Ultralow-power hydrogen sensing with single palladium nanowires. *Applied Physics Letters*, 94(22):223110, 2009.
- [97] Koichi Osuka and Hiroshi Kitajima. Development of mobile inspection robot for rescue activities: Moira. In *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003)(Cat. No. 03CH37453)*, volume 4, pages 3373–3377. IEEE, 2003.
- [98] Josue Pagan, Ramin Fallahzadeh, Mahdi Pedram, Jose L Risco-Martin, Jose M Moya, Jose L Ayala, and Hassan Ghasemzadeh. Toward ultra-low-power remote health monitoring: An optimal and adaptive compressed sensing framework for activity recognition. *IEEE Transactions on Mobile Computing*, 18(3):658–673, 2018.
- [99] Xiaopeng Peng, Garreth J Ruane, and Grover A Swartzlander Jr. Randomized aperture imaging. *arXiv preprint arXiv:1601.00033*, 2016.
- [100] Nhat Pham, Tuan Dinh, Zohreh Raghebi, Taeho Kim, Nam Bui, Phuc Nguyen, Hoang Truong, Farnoush Banaei-Kashani, Ann Halbower, Thang Dinh, et al. Wake: a behind-the-ear wearable system for microsleep detection. In *Proceedings of the 18th International Conference on Mobile Systems, Applications, and Services*, pages 404–418, 2020.

- [101] I Pits and AN Venetsanopoulos. Non linear digital filters, 1990.
- [102] Flavius V Pop, Bernard Herrera, Cristian Cassella, Guofeng Chen, Emrecan Demirors, Raffaele Guida, Tommaso Melodia, and Matteo Rinaldi. Novel pmu-based acoustic duplexer for underwater and intrabody communication. In *2018 IEEE International Ultrasonics Symposium (IUS)*, pages 1–4. IEEE, 2018.
- [103] M Pytasz, G Granosik, and Krzysztof Tchon. Modelling of wheeler-hyper mobile robot. In: *Advances in Robotics. Control, Perception, and Communication*. K. Tchon (Ed.), pages 313–322, 2006.
- [104] Andrea Ricci, Matteo Grisanti, Ilaria De Munari, and Paolo Ciampolini. Improved pervasive sensing with rfid: An ultra-low power baseband processor for uhf tags. *IEEE transactions on very large scale integration (VLSI) systems*, 17(12):1719–1729, 2009.
- [105] Nirupam Roy. Owllet: Insect-scale spatial sensing with 3d-printed acoustic structures. *GetMobile: Mobile Computing and Communications*, 25(2):14–20, 2021.
- [106] Nirupam Roy and Romit Roy Choudhury. Ripple ii: faster communication through physical vibration. In *13th USENIX Symposium on Networked Systems Design and Implementation (NSDI 16)*, pages 671–684, 2016.
- [107] Nirupam Roy, Mahanth Gowda, and Romit Roy Choudhury. Ripple: Communicating through physical vibration. In *12th USENIX Symposium on Networked Systems Design and Implementation (NSDI 15)*, pages 265–278, 2015.
- [108] Nirupam Roy, Haitham Hassanieh, and Romit Roy Choudhury. Backdoor: Sounds that a microphone can record, but that humans can't hear. *GetMobile: Mobile Computing and Communications*, 21(4):25–29, 2018.
- [109] Nirupam Roy and Romit Roy Choudhury. Listening through a vibration motor. In *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*, pages 57–69. ACM, 2016.
- [110] Nirupam Roy, Sheng Shen, Haitham Hassanieh, and Romit Roy Choudhury. Inaudible voice commands: The long-range attack and defense. In *15th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 18)*, pages 547–560, 2018.
- [111] Michael Rubenstein, Christian Ahler, and Radhika Nagpal. Kilobot: A low cost scalable robot system for collective behaviors. In *2012 IEEE international conference on robotics and automation*, pages 3293–3298. IEEE, 2012.
- [112] Sina Sadeghpour, Michael Kraft, and Robert Puer. Highly efficient piezoelectric micromachined ultrasound transducer (pmut) for underwater sensor networks. In *2019 20th International Conference on Solid-State Sensors, Actuators and Microsystems & Eurosensors XXXIII (TRANSDUCERS & EUROSENSORS XXXIII)*, pages 162–165. IEEE, 2019.
- [113] Sriram Sami, Yimin Dai, Sean Rui Xiang Tan, Nirupam Roy, and Jun Han. Spying with your robot vacuum cleaner: eavesdropping via lidar sensors. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*, pages 354–367, 2020.
- [114] Valkyrie Savage, Andrew Head, Björn Hartmann, Dan B Goldman, Gautham Mysore, and Wilmot Li. Lamello: Passive acoustic sensing for tangible input components. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 1277–1280, 2015.
- [115] Ashutosh Saxena and Andrew Y Ng. Learning sound location from a single microphone. In *2009 IEEE International Conference on Robotics and Automation*, pages 1737–1742. IEEE, 2009.
- [116] Irtaza Shahid, Yang Bai, Nakul Garg, and Nirupam Roy. In *Proceedings of the 1st ACM International Workshop on Intelligent Acoustic Systems and Applications*, 2022.
- [117] Sheng Shen, Dagan Chen, Yu-Lin Wei, Zhijian Yang, and Romit Roy Choudhury. Voice localization using nearby wall reflections.
- [118] Sheng Shen, Dagan Chen, Yu-Lin Wei, Zhijian Yang, and Romit Roy Choudhury. Voice localization using nearby wall reflections. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, pages 1–14, 2020.
- [119] Sheng Shen, Nirupam Roy, Junfeng Guan, Haitham Hassanieh, and Romit Roy Choudhury. Mute: bringing iot to noise cancellation. In *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication*, pages 282–296. ACM, 2018.
- [120] Yoni Sher, Lior Cohen, Daniel Istrati, and Hagai S Eisenberg. Low intensity lidar using compressed sensing and a photon number resolving detector. In *Emerging Digital Micromirror Device Based Systems and Applications X*, volume 10546, page 10546J. International Society for Optics and Photonics, 2018.
- [121] Gabriel T Sibley, Mohammad H Rahimi, and Gaurav S Sukhatme. Robomote: A tiny mobile robot platform for large-scale ad-hoc sensor networks. In *Proceedings 2002 IEEE International Conference on Robotics and Automation (Cat. No. 02CH37292)*, volume 2, pages 1143–1148. IEEE, 2002.
- [122] Mandy Lai Yi Sin, Gary Chun Tak Chow, Gary Man Keung Wong, Wen Jung Li, Philip Heng Wai Leong, and Ka Wai Wong. Ultralow-power alcohol vapor sensors using chemically functionalized multiwalled carbon nanotubes. *IEEE Transactions on Nanotechnology*, 6(5):571–577, 2007.
- [123] Kousuke Suzuki, Atsushi Nakano, Gen Endo, and Shigeo Hirose. Development of multi-wheeled snake-like rescue robots with active elastic trunk. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4602–4607. IEEE, 2012.
- [124] Joseph Tabrikian, Reuven Shavit, and Dayan Rahamim. An efficient vector sensor configuration for source localization. *IEEE Signal Processing Letters*, 11(8):690–693, 2004.
- [125] Ying Tan and Zhong-yang Zheng. Research advance in swarm robotics. *Defence Technology*, 9(1):18–39, 2013.
- [126] TI. Texas instruments energytrace++. <https://www.ti.com/tool/ENERGYTRACE>, 2022.
- [127] Chunwei Tian, Lunke Fei, Wenxian Zheng, Yong Xu, Wangmeng Zuo, and Chia-Wen Lin. Deep learning on image denoising: An overview. *Neural Networks*, 2020.
- [128] Xuyu Wang, Lingjun Gao, Shiwen Mao, and Santosh Pandey. Csi-based fingerprinting for indoor localization: A deep learning approach. *IEEE Transactions on Vehicular Technology*, 66(1):763–776, 2016.
- [129] Yan Wang, Wei-Lun Chao, Divyansh Garg, Bharath Hariharan, Mark Campbell, and Kilian Q Weinberger. Pseudo-lidar from visual depth estimation: Bridging the gap in 3d object detection for autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8445–8453, 2019.
- [130] Zhongqin Wang, Fu Xiao, Ning Ye, Ruchuan Wang, and Panlong Yang. A see-through-wall system for device-free human motion sensing based on battery-free rfid. *ACM Transactions on Embedded Computing Systems (TECS)*, 17(1):1–21, 2017.
- [131] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.
- [132] Martin Weinmann, Jens Leitloff, Ludwig Hoegner, Boris Jutzi, Uwe Stilla, and Stefan Hinz. Thermal 3d mapping for object detection in dynamic scenes. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, 2(1), 2014.
- [133] Fang Wu, Akash Vibhute, Gim Song Soh, Kristin L Wood, and Shaohui Foong. A compact magnetic field-based obstacle detection and avoidance system for miniature spherical robots. *Sensors*, 17(6):1231, 2017.
- [134] Yangbo Xie, Chen Shen, Wenqi Wang, Junfei Li, Dingjie Suo, Bogdan-Ioan Popa, Yun Jing, and Steven A Cummer. Acoustic holographic rendering with two-dimensional metamaterial-based passive phased array. *Scientific reports*, 6(1):1–6, 2016.
- [135] Yangbo Xie, Tsung-Han Tsai, Adam Konneker, Bogdan-Ioan Popa, David J Brady, and Steven A Cummer. Single-sensor multispeaker listening with acoustic metamaterials. *Proceedings of the National Academy of Sciences*, 112(34):10595–10598, 2015.
- [136] Donghe Yang, Jinsong Xia, and Liu Xi'ang. The research on intelligent mobile robot's avoiding obstacle by ultrasound. In *2010 International Conference on Artificial Intelligence and Computational Intelligence*, volume 1, pages 77–79. IEEE, 2010.
- [137] Jie Yang, Simon Sidhom, Gayathri Chandrasekaran, Tam Vu, Hongbo Liu, Nicolae Cegan, Yingying Chen, Marco Gruteser, and Richard P Martin. Sensing driver phone use with acoustic ranging through car speakers. *IEEE Transactions on Mobile Computing*, 11(9):1426–1440, 2012.
- [138] Shilong Zhai, Huaijun Chen, Changlin Ding, Fangliang Shen, Chunrong Luo, and Xiaopeng Zhao. Manipulation of transmitted wave front using ultrathin planar acoustic metasurfaces. *Applied Physics A*, 120(4):1283–1289, 2015.
- [139] Tomasz G Zieliński, Kamil C Opiela, Piotr Pawłowski, Nicolas Dauchez, Thomas Boutin, John Kennedy, Daniel Trimble, Henry Rice, Bart Van Damme, Gwenael Hannema, et al. Reproducibility of sound-absorbing periodic porous materials using additive manufacturing technologies: Round robin study. *Additive Manufacturing*, 36:101564, 2020.
- [140] Feryel Zoghalmi, Okan Kamil Sen, Harald Heinrich, Germar Schneider, Emec Ercelek, Alois Knoll, and Thomas Villmann. Tof/radar early feature-based fusion system for human detection and tracking. In *2021 22nd IEEE International Conference on Industrial Technology (ICIT)*, volume 1, pages 942–949. IEEE, 2021.
- [141] Nan Zou and Arye Nehorai. Circular acoustic vector-sensor array for mode beamforming. *IEEE transactions on signal processing*, 57(8):3041–3052, 2009.
- [142] Zhengxia Zou, Zhenwei Shi, Yuhong Guo, and Jieping Ye. Object detection in 20 years: A survey. *arXiv preprint arXiv:1905.05055*, 2019.