Acoustic Sensing and Communication Using Metasurface

Yongzhao Zhang, Yezhou Wang, and Lanqing Yang, *Shanghai Jiao Tong University;*
Mei Wang, *UT Austin;* Yi-Chao Chen, *Shanghai Jiao Tong University and Microsoft Research Asia;* Lili Qiu, *UT Austin and Microsoft Research Asia;* Yihong Liu, *University of Glasgow;* Guangtao Xue and Jiadi Yu, *Shanghai Jiao Tong University*

This paper is included in the Proceedings of the 20th USENIX Symposium on Networked Systems Design and Implementation.

April 17–19, 2023 • Boston, MA, USA

978-1-939133-33-5

Open access to the Proceedings of the
20th USENIX Symposium on Networked
Systems Design and Implementation
is sponsored by

جامعة الملك عبدالله
للعلوم والتقنية
King Abdullah University of
Science and Technology

Acoustic Sensing and Communication Using Metasurface

Yongzhao Zhang[+], Yezhou Wang[+], Lanqing Yang[+], Mei Wang[†]
Yi-Chao Chen[+*], Lili Qiu[†*], Yihong Liu[♮], Guangtao Xue[+], Jiadi Yu[+]
[+]*Shanghai Jiao Tong University,* [†]*UT Austin,* [*]*Microsoft Research Asia,* [♮]*University of Glasgow*

Abstract

Acoustic sensing is increasingly popular owing to widely available devices that support them. Yet the sensing resolution and range are still limited due to limited bandwidth and sharp decay in the signal at inaudible frequencies. Inspired by recent development in acoustic metasurfaces, in this paper, we first perform an in-depth study of acoustic metasurface (AMS) and compare it with the phased array speaker. Our results show that AMS is attractive as it achieves a significant SNR increase while maintaining a compact size. A major limitation of existing AMS is its static configuration. Since our target may be at any possible location, it is important to support scanning in different directions. We develop a novel acoustic system that leverages a metasurface and a small number of speakers. We jointly optimize the configuration of metasurface and transmission signals from the speakers to achieve low-cost dynamic steering. Using a prototype implementation and extensive evaluation, we demonstrate its effectiveness in improving SNR, acoustic sensing accuracy, and acoustic communication reliability over a wide range of scenarios.

1 Introduction

Motivation: Acoustic sensing and communication are becoming increasingly popular due to widely available devices that support it, including smartphones, smart speakers, and many IoT devices. Many interesting sensing systems have been proposed using acoustic signals (e.g., [15, 25, 35–39, 42, 51, 57, 65]). For example, [35, 36, 42, 51, 57, 65] develop smartphone based approaches that transmit inaudible acoustic signals to track a target's distance, position, and movement. [38, 52] enables more accurate sensing by exploiting a microphone array on a smart speaker. [11] develops acoustic communication systems as an NFC alternative, [9] designs an underwater messaging system using acoustic signals

since acoustic signals attenuate slower than RF signals. Refer to [5, 12] for more comprehensive surveys on the underwater acoustic communication systems. Despite significant advances in acoustic sensing, there is a fundamental limit on its sensing range and resolution as shown in the Cramer-Rao bound, which indicates the sensing resolution is limited by SNR and the number of transmitters and receivers. Similarly, acoustic communication also faces similar challenges according to the Shannon capacity.

In order to further improve the performance, one could increase the number of transceivers. However, increasing the number of transceivers increases the cost, size, and energy consumption. In addition, existing sound cards cannot support more than 8 channels. All of these factors significantly limit their applicability in a real-world deployment.

Another option is to adopt an acoustic lens (or acoustic metasurface, AMS). Like optical lenses, acoustic lenses can steer the direction of acoustic wave propagation and focus in a certain region. However, an acoustic metasurface is usually bulky due to the large wavelength of acoustic waves. Recently there emerged some metasurface quantization designs (e.g., [40, 41]). They comprise many sub-wavelength cells, where each cell can act like a mini-antenna and modify the phase and/or intensity of the incident wave so that collectively the AMS can manipulate the wave in an interesting way (e.g., steer the outgoing wave towards a certain direction).

Our approach: Inspired by the potential benefit of AMS, we first compare a passive AMS with beamforming using multiple speakers. We find beamforming using 3 and 6 speakers increases SNR by $4.7dB$ and $7.9dB$, respectively. In comparison, an acoustic metasurface of size 16×16 cells under 1 speaker increases SNR by $15.5dB$. The results suggest AMS is attractive since it can significantly increase the SNR using a compact design without consuming power. To achieve a similar SNR increase, we need 36 speakers spanning $30cm$, which is bulky and challenging to deploy.

While passive AMS is attractive, the existing AMS can support only static configuration (e.g., always beamform towards
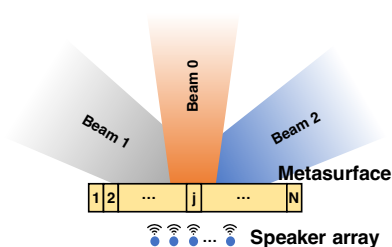

Figure 1: Dynamic beam steering with speaker array and acoustic metasurface (AMS).

a fixed angle). Since the target can be at any location, dynamic beam steering is necessary. One option is through the mechanical movement of the AMS, which not only increases the cost but also limits the speed of adaptation. In this paper, we propose combining beamforming using a small number of speakers with AMS, as shown in Figure 1, to achieve the best of both worlds: the use of AMS allows us to keep the number of speakers low while still achieving sharp beam, large SNR gain, and high resolution; the use of beamforming even using very few speakers can enable dynamic steering without movement. A small number of speakers with a passive AMS can achieve similar beamforming resolution as a large number of speakers. For example, our evaluation shows using an AMS with 16x16 cells and a 6-speaker phased array is comparable to a 9x16=144 phased array in terms of beam width.

To this end, we develop a novel algorithm to jointly optimize the AMS configuration and beamforming weights for the phase array. Specifically, the joint AMS and beamforming design can be formulated as an optimization problem whose objective is to maximize the signal strength along each of the desired angles (e.g., sampled from a range of angles) and minimize the performance variance across these angles and energy in the side lobes. We use the gradient projection method to solve the optimization problem. In addition, we augment our optimization framework to further optimize the speakers' placement and improve the performance.

Based on our designed algorithm, we implement an acoustic system that comprises a 3D-printed AMS, 6 speakers, and a microphone. We apply our algorithm to steer the outgoing beam in real time. We evaluate our design using (i) SNR of the received signal, (ii) sensing performance (i.e., distance estimation using Frequency Modulated continuous Waves (FMCW) and angle estimation using the MUltiple SIgnal Classification (MUSIC) algorithm), and (iii) the communication error.

Our contributions can be summarized as follow:

- Using extensive evaluation and analysis, we shed light on the benefits of phased array versus AMS.

- We jointly optimize AMS and phased array configurations to enable dynamic beam steering and high SNR.

- We further improve the performance by optimizing the speaker placement.

- We develop an acoustic system based on our joint design of AMS and beamforming and apply it to acoustic sensing and communication. Our evaluation shows our system yields a significant improvement in SNR, distance estimation, angle estimation, and communication reliability. In particular, leveraging AMS and phased array allows us to dynamically steer the beam to the desired direction and boost SNR by 18.4dB over a single speaker without AMS. The improved SNR in turn increases the acoustic sensing and communication ranges. Our approach increases the sensing range from $1.5m$ in a single speaker without AMS to $4m$ using 6-speaker with AMS; similarly, it increases the communication range from $0.8m$ to $3.9m$.

Paper outline: We review existing work in Section 2, and introduce acoustic metasurface in Section 3. We describe our algorithm to jointly optimize AMS and speaker array system in Section 4. We present our simulation and testbed experiment results in Section 5. We discuss the limitation and future work in Section 6. We conclude in Section 7.

2 Related Work

Our work is closely related to wireless sensing, acoustic communication, acoustic metasurface, and phased array.

Wireless sensing: Wireless sensing has become increasingly popular due to many important applications. Many algorithms and systems have been developed recently using acoustic [15, 25, 30, 35, 42, 45, 51, 57, 64, 65, 67], WiFi (e.g., [21, 47, 50]), mmWave (e.g., [20, 58, 61]), and RFID signals (e.g., [17, 34, 53–55]).

Among them, acoustic sensing is appealing due to its high accuracy and widely available commodity devices that support it. They use time of flight (e.g., BeepBeep [45]), Doppler shift (e.g., AAMouse [64]), FMCW (e.g., [35,42]), phase (e.g., [57]), correlation (e.g., [43]), channel impulse response (e.g., Strata [65]), and Angle of Arrival (AoA) (e.g., [38, 49, 56]) for sensing. Some works also leverage machine learning for acoustic sensing (e.g., by applying neural networks to either post-processed signals or raw signals) and show ML based sensing is promising.

There are also significant works on sensing using RF signals. Some leverage similar algorithms as in acoustic sensing, while others explore new features and algorithms. For example, [33] use Channel State Information (CSI). Tagyro [59] tracks rotation using an array of passive RFID tags and two orthogonal RFID reader antennas. [46] further exploits polarization to track rotation and translation movement. [22] pushes the tracking accuracy to sub-centimeter level using a large phased array and large bandwidth.

Acoustic Communication: Sound has been a popular way of communicating information. Interestingly, we can also encode and transmit digital data over the acoustic channel. At a high level, it is essentially the same as RF communication but uses a different frequency. A number of interesting systems have been developed for acoustic digital communication (e.g., [11, 23, 66]) as an alternative to Near Field Communication (NFC) owing to the wide availability of speakers and microphones that support acoustic communication. Many of them leverage OFDM due to its robustness to multipath fading. Therefore, we also adopt the OFDM based acoustic communication in our work.

Acoustic metasurface: Ultimately, the accuracy of acoustic sensing depends on the SNR and numbers of speakers and microphones. Acoustic metasurface can boost SNR using a passive 2D structure, which can help improve sensing performance. A metasurface has many unit cells, and each cell can be potentially treated as a mini sound source. In this way, AMS effectively increases the number of speakers, thereby improving sensing resolution. By controlling the phase and/or amplitude of acoustic wave propagation through each unit cell, AMS can manipulate the wave fields. Many designs of acoustic metasurface have been proposed in the literature [6, 32]. Coiling-up space structure [27–29, 40, 41, 60] achieves phase manipulation by forcing acoustic waves to propagate along a coiled path. Helmholtz-resonator-like structure [13, 26] produces a tunable phase velocity and a high transmission efficiency with multiple Helmholtz resonators. Membrane-type structure [14, 19, 44, 62, 63] eliminates reflection with carefully designed membrane resonators. The above designs are not reconfigurable. There are active acoustic metasurfaces [10, 18, 24], as well. They use mechanical structure or emerging materials that can be deformed under the control of a magnetic field or electric current. However, these designs are expensive and bulky to implement.

Our work is inspired by [40, 41]. [40] develops a powerful methodology that assembles many sub-wavelength premanufactured 3D units into an acoustic metasurface. Each unit encodes a specific phase offset. By re-arranging these units, one can produce many different metasurfaces. Since acoustic sensing/communication usually use inaudible sounds with much smaller wavelength to avoid disturbance, the coiling-up metasurface is more compact than Helmholtz-resonator and membrane-type structure. [41] discusses several applications of these metasurfaces, including generating acoustic collimator, acoustic magnifying glasses, and acoustic telescopes. Our work goes beyond [40, 41] by enabling dynamic steering through combining multiple speakers with AMS and applying it to acoustic sensing and digital communication.

[16] proposes using 3D-printed metamaterial to cover the microphone and embed the direction-based signature. During the calibration stage, recordings from all possible angles are collected. During the online usage, the current recording is compared with all recordings collected in the calibration to find the best match, which is used for AoA estimation. [7] develops an acoustic sensing system that uses 3D printed smart surface to embed direction information into the signals for generating a depth map. Our work is related to the above work but goes beyond them by (i) eliminating labor-intensive calibration and (ii) directly increasing SNR and sensing resolution, which can benefit any sensing or communication approaches instead of tailing to one specific sensing scheme. Therefore our design is more general and support more applications. Moreover, our optimization framework for configuring AMS and a speaker array is flexible and can support a range of important what-if analyses. Our adoption of a regular shaped AMS also makes it easier to analyze and optimize its impact on the overall system performance.

Phased array: Multiple transmitters and/or multiple receivers can be used to strengthen the received signals. At the transmitter end, beamforming can be used to generate transmissions that arrive in phase at the receiver so that the multipath signals are added up constructively. At the receiver end, the receiver can compensate for the phase difference of the received signals across different antennas to ensure constructive combining. As mentioned earlier, in order to achieve a comparable gain of AMS, a large phased array is necessary, which increases the size, cost, computation, and power. This motivates our design of AMS based sensing and communication system.

3 Acoustic Metasurface

In this section, We provide background on acoustic metasurface and its properties.

3.1 Background of Acoustic Metasurface

The ability to shape acoustic fields has diverse applications, such as high-quality sound production, particle manipulation, non-invasive therapies, and increasing sensing and communication range and resolution. One way to shape the acoustic fields is to use phased arrays by controlling the phase and amplitude of the transmission signals emitted from each of the speakers. The cost, power consumption, and size of a phased array rapidly increase with the number of speakers.

A few recent research papers show that acoustic metasurfaces could be a promising solution. An acoustic metasurface is a 2D structure that consists of many sub-wavelength cells [40, 41]. By carefully designing each of its cells, we can manipulate acoustic waves in an interesting way. Each unit cell can be viewed as a mini sound source. To perform beamforming in a certain direction, we can ensure the paths going through different cells in the metasurface add up constructively in the desired direction. This can be achieved by
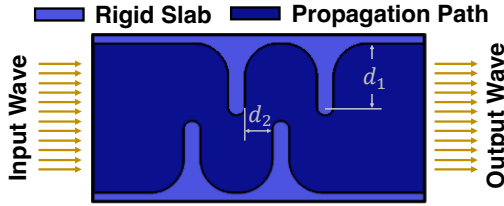
Figure 2: The structure of a unit cell is mainly determined by the parameters d_1 and d_2. Different lengths of the propagation paths induce different phase delay at the output.
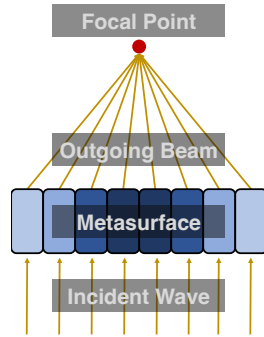

Figure 3: A metasurface consists of multiple unit cells and beamforms towards a focal point by properly configuring the unit cells.

(a) Sound field   (b) 100*mm*   (c) 200*mm*

Figure 4: (a) Sound field simulated in COMSOL with a 16×1 metasurface when we transmit $20kHz$ sound at the focal point (100*mm*). Part of the energy is concentrated in a specific direction. (b) and (c) show the sound field in the zoom-in area when the speaker is placed at 100*mm* and 200*mm*, respectively. When the speaker is at the focal point, the signal coming out of the metasurface is a parallel wave.

letting each unit cell compensate for the phase difference. For example, without special design, in the desired direction, the path going through cell 1 differs from cell 2 by $\phi_{1,2}$. To ensure the signals from these two paths add up constructively, we can design the unit cell 1 and cell 2 to compensate phase difference $\phi_{1,2}$. One way to achieve this is to impose different geometric structure so that the path going through cell 2 is $\frac{\phi_{1,2}}{2\pi}\lambda$ longer than cell 1.

Figure 2 shows an example structure of unit cells used in [40]. Assume the sound waves pass through the unit cell from the left. The curved propagation path will increase the time it takes to penetrate the unit cell, which essentially introduces a phase shift to the outgoing wave. The structures of different unit cells are determined by two dominant parameters: d_1 and d_2 [40], which result in different propagation path lengths and hence different phase delay. One way to determine d_1 and d_2 is through enumeration in a simulator (e.g., COMSOL [2], which is a widely used finite-element-based multiphysics simulator).

We arrange the unit cells in a straight line to form a 1D metasurface, or in a rectangle to form a 2D metasurface. By introducing an appropriate phase shift at each cell, we can achieve beamforming. Figure 3 shows an example. To make it easy to assemble/re-assemble a metasurface, [40] quantizes the types of unit cells into 16 choices, which covers the phase shift from 0 to 2π. So for each unit cell, we can choose one
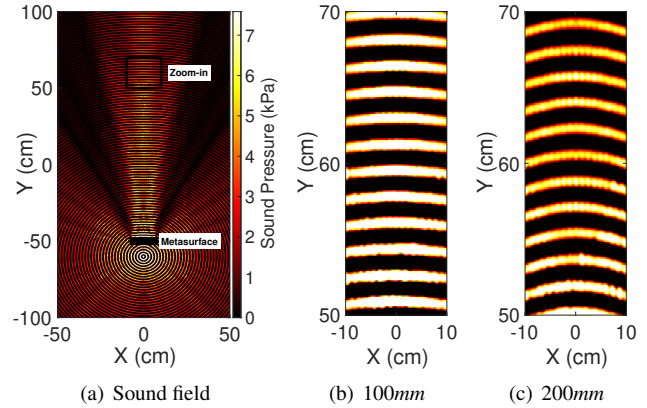
whose phase shift is the closest to our desired shift.

3.2 Properties of Acoustic Metasurface

High transmission efficiency: As shown in Figure 2, the unit cells have intricate maze-like internal structures, with four parallel bars positioned orthogonal to the direction of incoming sound waves. Interestingly, the transmission efficiency is high and reaches 98% on average across all unit cells [40]. This is due to the following two major reasons: i) The sub-wavelength cells produce diffraction and cause the energy of sound to bypass the parallel bars instead of being reflected back; and ii) The bars inside each unit cell are curved instead of sharp angles to reduce acoustic impedance and maintain high transmission efficiency. Overall, the acoustic metasurface has negligible power loss, so we do not consider the power loss when developing AMS.

Focusing behavior: Figure 3 shows that an incident plane wave is focused at a focal point after passing through the metasurface. Due to the reciprocity principle [8], when a point source is placed at the focal point, the signal coming out of the AMS should be a plane wave towards the direction orthogonal to the metasurface. Figure 4(a) and 4(b) show an example scenario, where the source is placed at 100*mm*, which is the focal point. We observe the outgoing wave is nearly parallel in one direction. Second, when the source is not at a focal point, the wave is no longer parallel, as shown in Figure 4(c), and the energy of the signal will be dispersed to nearby directions, making signal strength attenuates faster. Since we want to concentrate the energy in one direction and make the sound wave propagate in a longer range, plane waves are preferred.
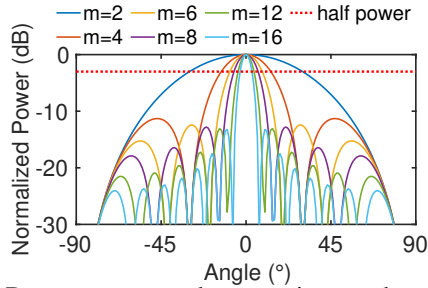
Figure 5: Beam patterns under a varying number of speakers
m. The received sound is normalized by dividing it by the
maximum power and then converted to decibels.

Adaptation: Once a metasurface is printed, the mapping
from the incoming wave to the outgoing wave is fixed. Since
the target can be in any direction, it is important to change
the direction of the outgoing wave. Given the fixed metasur-
face, one way to change the direction of the outgoing wave
is to move the AMS either through translation movement
or through rotation. While movement is feasible, mechani-
cal movement is slow, consumes significant power, causes
wear and tear, and may even require operator intervention.
Therefore, in this paper we seek a software-based approach
to realize fast dynamic adaptation.

4   Phased Array with Metasurface

The passive acoustic metasurface is a fixed 2D structure. Once
designed, it converts from the incident wave to the outgoing
wave in a fixed manner. For practical use, it is desirable to
dynamically adjust the direction of the wave coming out of
the metasurface. We can achieve this using either mechanical
movement or beamforming using a phased array. We take the
latter approach due to its software control and eliminating the
need of movement. An important question is how to config-
ure the metasurface and phased array to realize our desired
beamforming. Below we first introduce phased array and then
describe how to use phased array with metasurface to achieve
dynamic control at a low cost.

4.1   Phased Array

Phased arrays use beamforming to combine signals from
multiple speakers constructively. Beamforming can be applied
to either senders or receivers or both. There are a number
of beamforming algorithms. They vary in the optimization
objectives: some maximize the signal, while others minimize
interference.

In analog beamforming, beamforming is performed on ana-
log signals at the transmitter before sending to the air or at
the receiver before the analog to digital conversion. In digital
beamforming, beamforming is performed on digital signals
at the transmitter before digital to analog conversion or at the
receiver after analog to digital conversion.

The beamforming capability depends significantly on the
number of speakers and their separation. As Figure 5 shows,
the beam width in the desired direction is relatively large and
the sidelobes are significant when the number of speakers
is within 8. The half power beam width (HPBW) at 0° (i.e.,
perpendicular to the speaker array) can be approximated as
follows [48]: $\theta_{0.5} \approx \frac{0.886\lambda}{md}$ where λ is the wave length, m is the
number of speakers, and d represents the speaker separation,
which is usually recommended to be $\frac{\lambda}{2}$. For example, the
HPBW will be 59.6°, 25.5°, 16.9°, and 6.3° when the number
of speakers is 2, 4, 6, and 16, respectively. The beam width
for a general angle can be derived as follows: $\theta_{0.5s} = \frac{\theta_{0.5}}{\cos \theta_s}$
where θ_s is the steering angle and $\theta_{0.5s}$ is the HPBW of the
steered beam. This indicates that the scanning range should
not be too large and usually we let $\theta_s \leq 60°$. These results
show that the acoustic beamforming resolution using a small
phased array is limited.

4.2   Phased Array Coupled with Metasurface

Passive AMS is not reconfigurable on-the-fly once it is assem-
bled. To provide dynamic adaptation while achieving high
resolution and long range, we propose using a small number
of speakers along with an acoustic metasurface. We optimize
the speakers' beamforming so that the outgoing wave from
the AMS is towards our desired angle.

More specifically, phased array can control the direction
of the output signal, which serves as the incoming signal
towards the AMS. By optimizing the transmission signals, we
can potentially generate any shaped waves coming out of the
AMS. The use of multiple speakers allows us to achieve fast
dynamic control without movement. In order to fully realize
this capability, we should carefully design the AMS to cover a
wide range of angles and control the transmission signals from
multiple speakers in order to dynamically generate the desired
signal coming out of the AMS. Below we first formulate the
problem and then present our solution.

4.2.1   Problem Formulation

As shown in Figure 1, there are M speakers. Let w_i denote the
codeword for the i-th speaker, where w_i is a complex number
whose magnitude and phase are the scaling factor and phase
shift for the i-th transmission signal, respectively. There are
N unit cells in AMS. The acoustic signal received by the j-
th AMS cell from the i-th speaker $S_{i,j}$ can be computed as
follow, where t_i is the i-th speaker's transmission signal and
H_{ij} denotes the channel between the i-th speaker and j-th cell.

$$S_{i,j} = H_{ij}w_it_i \tag{1}$$

Since the relative position between the AMS cell j and trans-
mitter i is pre-determined, we can derive $H_{i,j} = F(d_{i,j}) =
a(d_{i,j})e^{-j2\pi f\frac{d_{i,j}}{c}}$, where c is speed of acoustic signal, $d_{i,j}$ is

the distance from the i-th transmitter to j-th cell, $a(d_{i,j})$ is the amount of signal attenuation at the distance $d_{i,j}$, and $F(\cdot)$ is a function that models how the channel attenuates with the distance $d_{i,j}$.

We can take the placement of the phased array into account, denoted as x, and re-write the above relationship in a matrix form as follows:

$$S_{in} = H(x)w \tag{2}$$

We omit the transmission signal t_i before beamforming hereafter because it is the same at each speaker.

Each cell in AMS modifies the incident signal (e.g., by adding a path delay and/or changing the amplitude). Such a modification can be captured using a matrix, denoted as G, and the design of G will be covered in Sec 4.2.3. Then the signal coming out of the AMS becomes

$$S_{out} = GH(x)w \tag{3}$$

Finally, let R_d denote the signal in a given steering direction d from the AMS. R_d can be derived as follow, where K_d denotes the steering vector corresponding to the direction d between the AMS and target.

$$R_d = K_d GH(x)w \tag{4}$$

Our goal is to design the AMS and codeword of the speaker array to maximize the signal strength along each angle of interest. For example, if we want to support a scanning angle from $-60°$ to $60°$, for each angle within the range, we want to maximize the signal strength. Note that the AMS has a fixed configuration across all angles, while the codeword can change for each beamforming angle as in typical beamforming scenarios. Therefore, the signal of interest R can be derived as follow:

$$R = KGH(x)W \tag{5}$$

where R is a matrix of size $d \times d$ (each row represents the received signal from a given direction d and each column represents the steering direction), K is a $d \times N$ matrix specifying the steering vector from the N unit-cell AMS, G is an $N \times N$ matrix and its diagonal elements specify how the N-cell AMS translates the incident signal into outgoing signal, $H(x)$ is an $N \times M$ matrix specifying the channel from M transmitters to N-cell AMS, and W is $M \times d$ codebook for M speakers corresponding to d directions.

The channel H and steering vector K are fixed and can be derived analytically. Given H and K, we want to find the optimal static AMS configuration G and codebook W to perform beamforming across a wide range of angles. Since we optimize the power of the beams in each direction, we use the power P of the received signal R hereafter, which is denoted by $P = |R|^2$.

The structure of our received signals P can be visualized in Figure 6, where we aim to have high signal strength along the
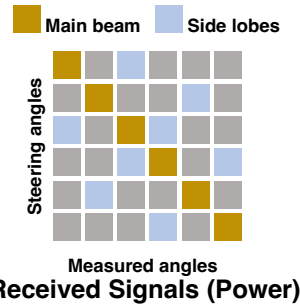

Received Signals (Power)

Figure 6: The structure of the received signal at various angles.

diagonal elements, which indicates our signal is beamformed towards the desired steering angle.

Our objective function comprises the following 3 terms:
Sum Power: Due to the use of a static metasurface design and the need to accommodate a wide range of angles, our goal is to maximize the sum of power across all d directions. This can be derived as follow:

$$L_{power} = \text{tr}(P) \tag{6}$$

where $tr(\cdot)$ is the trace of a matrix (i.e., the sum of the diagonal elements in the trace). This is shown in Figure 6.
Minimum Variance Criterion: Solely maximizing the total power may introduce some dead zones for certain directions. To avoid that problem, we add the variance of P's diagonal as the penalty term L_{var} to ensure all directions are covered:

$$L_{var} = \text{var}(\text{diag}(P)). \tag{7}$$

For generality, we introduce a weight matrix Q, which can put different weights on different angles. As a result, we have:

$$L_{var} = \text{var}(\text{diag}(PQ)) \tag{8}$$

where $Q = \text{diag}(q_1, q_2, \ldots, q_l)$ is a set of weights to control. If we have prior knowledge about the target's (approximate) location, we can increase the entries in Q that correspond to the locations close to the target.
Minimum Sidelobe: Suppressing the side lobes is critical for sensing and communication. In the signal processing literature, extensive works have been done to control sidelobe level (SLL). Sidelobe nullification and minimization are two common methods. Some methods require prior knowledge about the sidelobe's direction, while other methods minimize the maximum sidelobe. We experiment different ways of suppressing sidelobes and find minimizing the average SLL (i.e., minimize the sum of absolute values of all non-diagonal peaks in P) is most effective in our context.

We observe that the non-diagonal peaks shown in Figure 6 are considered as sidelobes and can degrade the overall performance. Therefore, we propose to minimize the sum of non-diagonal peaks as follows:

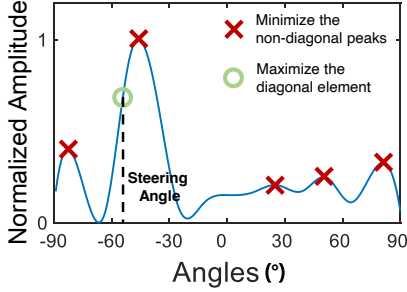$$L_{sidelobe} = \sum \text{non-diagonal peaks} \tag{9}$$

Figure 7: Beam pattern when steered to a specific angle during optimization. Minimizing the non-diagonal peaks helps reduce the side lobes and increase the directivity.

To derive $L_{sidelobe}$, we identify peaks in the P matrix (e.g., using findpeaks() function) and then sum up the peaks that are in non-diagonal entries of the matrix. It not only reduces the sidelobes, but also improves the quality of the main lobe. As shown in Figure 7, if the highest peak is a non-diagonal element, we also minimize it to revise the direction.

Putting together, we have the following optimization model:

$$\min_{W,\Theta,x} -L_{power} + \mu L_{var} + \gamma L_{sidelobe}$$

$$s.t. \begin{cases} |G_{ii}| = 1, \ (i = 1, 2, \ldots, N) \\ |W_{ij}| \leq 1 \ (i = 1, 2, \ldots, M, \ j = 1, 2, \ldots, d). \end{cases}$$

where μ and γ are parameters controlling the importance of the variance and sidelobe terms, respectively. We have two constraints on the magnitude of the metasurface parameters G and codebook W. Both G and W should be no more than 1.

The constraints on the magnitude of metasurface G are called constant modulus constraints (CMC). It is well-known that problems involving CMC are nonconvex and NP-hard [31]. $|G_{ii}| = 1$ refers to the points on the surface of an N dimensional hypercube, which indicates each metasurface cell does not change the magnitude of the incoming signal. These are non-convex constraints. $|W_{ij}| \leq 1$ are constraints on the magnitude of the phased array. The set contains the entire hypercube and includes the interior. Thus, it is a convex set. Therefore, for the phased array codebook, we restrict the amplitude to be within 1 instead of exactly equal to 1 to make the problem easier to solve.

4.2.2 Optimization

Our problem is a non-linear constrained optimization problem. Due to the presence of the constraints, we cannot directly apply the gradient descent scheme. Therefore, we use the gradient projection method, which ensures the solution after each gradient descent update still falls within the feasible set Ω. Specifically, if the $k + 1$-th update (i.e., $x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)}$) makes the solution fall outside the feasible region, where α_k is the learning rate and $d^{(k)}$ is the gradient, we project it to a point inside the feasible set Ω as follows:

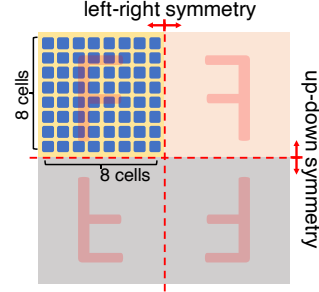$$x^{(k+1)} = \Pi[x^{(k)} + \alpha_k d^{(k)}] \quad (10)$$

Figure 8: The symmetry property of a 16×16 metasurface.

where Π is projection operator, and $\Pi[x]$ is called the projection of x in Ω. To do that, we normalize the amplitude of G_{ii} after each update and normalize W_{ij} if it is larger than 1. As a result, we use Adam optimizer in Pytorch for optimization. Adam is an extended version of stochastic gradient descent that adapts the learning rate for each parameter. We modify the output from the Adam during each iteration using Equation 10 to ensure constraints are satisfied.

4.2.3 Additional Design Details

In this section, we describe how to get the input required for optimization.

Symmetry Property of AMS As mentioned earlier, the diagonal of variable G should represent the phase delay for metasurface cells. Our metasurface is a 2D structure. We observe that the configuration of the AMS should be left-right symmetric and up-down symmetric, as shown in Figure 8, since the scanning performance should be the same in left and right in the azimuth direction and the beam pattern should also be the same in top and bottom in the elevation direction. By utilizing the left-right and up-down symmetry property, we can reduce the search dimension for G by 75%.

Codebook Since the range of the steering angle is from $-60°$ to $60°$, the codebook is also symmetric between the positive angles and negative angles. Therefore, we can optimize half of the codebook (i.e., corresponding to the steering angle in $(-60°, 0)$) and copy them to generate the codebook for $(0, 60°)$.

Channel From Phased Array to Metasurface The channel $H(x)$ can be determined based on the speakers and metasurface cells' positions. Let $x = \{x_1, x_2, \ldots, x_M\}$ denote the speakers' locations, and $g = \{g_1, g_2, \ldots, g_N\}$ denote the metasurface cells' locations. We can derive the channel as follows:

$$H(x) = \begin{bmatrix} F(\|x_1 - g_1\|) & \cdots & F(\|x_M - g_1\|) \\ F(\|x_1 - g_2\|) & \cdots & F(\|x_M - g_2\|) \\ \vdots & \ddots & \vdots \\ F(\|x_1 - g_N\|) & \cdots & F(\|x_M - g_N\|) \end{bmatrix} \quad (11)$$
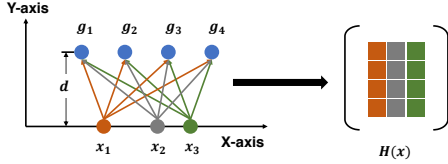
Figure 9: Derive channel matrix H with varying speaker distributions x. Consider 3 speakers and 4 metasurface cells as an example. We can derive the distance between each speaker and metasurface cell to get the channel H between the phased array and metasurface.

where $\|\cdot\|$ denotes the distance between two points (i.e., a speaker and a metasurface cell) and $F()$ denotes the function that maps the distance to the wireless channel, including the amplitude and phase. Figure 9 shows an example.

We can either (i) take a given phased array setup (e.g., uniformly distributed linear array) as the input or (ii) optimize the phased array setup. In the latter case, we treat x_i as the optimization variables along with the other variables. Note that we do not impose any constraints on x because we already consider the symmetry property of metasurface G and codebook W. Equation 11 assumes a single line-of-sight path between the phased array and metasurface, which is realistic since the metasurface is close to the phased array and there is no blockage.

4.2.4 System Design

In this section, we provide further details of our system design, including the AMS design, codebook design, and array placement.

We sample angles from $-60°$ to $60°$ with $1°$ apart. Therefore, for a 6-speaker system, the codebook W is a 121×6 matrix, which contains 121 independent codewords for 121 directions and 6 speakers. Figure 10 shows the amplitude and phase of the optimized codebook, where the first speaker is set as the reference and is aligned to be zero phase. Since our goal is to maximize the sum power of diagonal elements, the amplitude of each element in the codebook is 1 to achieve the maximum transmission power, while the phase is manipulated to generate our desired sound field at the metasurface.

Next, we reconstruct the phase distribution of the metasurface by utilizing the diagonal elements of G and the symmetry property. The results are given in Figure 11(a). This is different from that of [40] due to the presence of a phased array. As mentioned in Sec. 3, the phase shift of each AMS cell is quantized to 16 levels for flexible design and assembly/disassembly. Then the final AMS can be assembled by choosing the unit cells with the closest phase shift, as shown in Figure 11(b), where the color reflects the unit cell index and a higher index indicates a larger phase shift.

We can either (i) place speakers in the phased array uniformly or arbitrarily and feed the placement to our optimization algorithm or (ii) let our algorithm optimize the placement

(a) Magnitude of codebook.
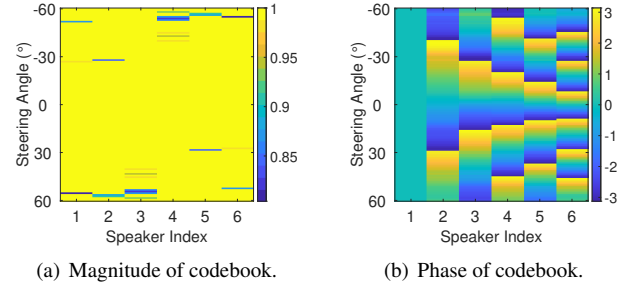

(b) Phase of codebook.

Figure 10: The optimized codebook design, where the first speaker is set as the reference. The magnitudes of the codebook are all close to 1, but some are slightly less than 1 since their constraints are ≤ 1.
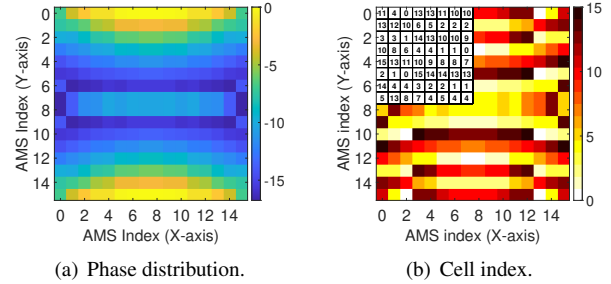

(a) Phase distribution.

(b) Cell index.

Figure 11: Phase distribution and cell indices of the optimized 16×16 metasurface design, where the cell index is the index to one of the 16 quantized phase shifts. The numbers in the upper left corner denote the cell indices for the top left metasurface and we omit the other parts for brevity due to the left-right symmetry and top-bottom symmetry.

along with other configuration parameters. In evaluation, we compare uniform placement and optimized placement.

5 Performance Evaluation

In this section, we first present our evaluation methodology and then describe performance results.

5.1 Evaluation Methodology

We use the experiment setup shown in Figure 12 for our evaluation. The system can be divided into three parts: speakers, microphones, and an acoustic metasurface (also referred to as an acoustic lens or AMS). We use uniform placement as the default configuration. In this case, we have 6 identical miniature speakers (16Ω, $0.25W$) as the transmitter. Each speaker is connected with an operational amplifier THS4001 [4] to amplify the voltage and a power amplifier LM386 [3] to amplify the current. The distance between the centers of adjacent speakers is $8.6mm$, which is a half wavelength of $20kHz$ sound. We used 4 microphones to form a microphone array as a receiver. The distances between the 4 microphones were $3.06cm$, $2.04cm$, and $3.06cm$ to reduce ambiguity and obtain better performance [38]. All speakers and microphones are

(a) Experiment setup

(b) LESN design for difference scenarios

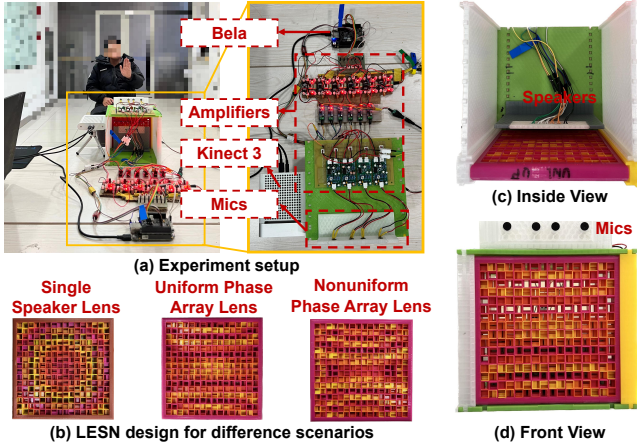**(c) Inside View**

(d) Front View

Figure 12: System setup.

connected to the same Bela board [1] for signal synchronization. We also optimize speaker placement using the approach described in Section 4.2.

We construct an acoustic lens according to our optimization in Section 4.2.2. Our lens consists of 256 (16×16) unit cells, spanning over $15cm \times 15cm$. Since the unit cells are quantized into 16 choices, we 3D print 16 different types of unit cells and assemble them to an acoustic lens according to the evaluation scenario. For example, we assemble an acoustic lens for a 1-speaker setup, a different acoustic lens for 6 speakers with uniform separation, and another one for 6 speakers with non-uniform separation, as we jointly design the metasurface with the speaker array. To ensure most signals coming out of the speakers go through the acoustic lens, we place our lens $2cm$ away from the speaker array. For a single speaker, we follow the setup in [41] where the lens is placed $10cm$ away from the speaker array.

We evaluate our approach in terms of (i) SNR, (ii) sensing accuracy, and (iii) communication performance. For acoustic sensing, we use Kinect V3 to get the ground truth distance and angle of arrival (AoA). We let the speakers transmit the following FMCW signal: $tx(t) = cos(2\pi f_{min}t + \frac{\pi B t^2}{T})$, where $f_{min} = 16kHz$, $B = 4kHz$, and $T = 0.1s$. We quantify the sensing accuracy using distance error and angle of arrival error.

1D MUSIC is a widely used AoA estimation algorithm. It computes the auto-correlation matrix R of the received signals x as $R = x^H x$, where x is a $1 \times N$ vector and x^H is the conjugate transpose of x, and then performs eigenvalue decomposition on R. Let R_N represent the noise space matrix, which is the space spanned by the $N - M$ smallest eigenvectors, where M is the number of signals. The peak in the pseudo spectrum $p(\theta) = \frac{1}{a(\theta)^H R_N R_N^H a(\theta)}$ corresponds to the AoA.

For acoustic communication, we encode the data using OFDM. Each OFDM frame contains 180 BPSK symbols, which are striped onto 12 subcarriers spanning over $18kHz$-$20kHz$. We use CDMA as FEC code to improve resilience

and the code rate is 50%. We quantify the communication performance using bit error rate (BER) and frame error rate (FER). While there are other coding schemes for acoustic communication, the benefit of our approach (i.e., acoustic lens with a speaker array) is likely similar across different acoustic coding schemes.

Unless otherwise specified, all results are from **testbed experiments**: we use a 6-speaker array with an equal separation of $9.4mm$ between the two adjacent speakers and 16×16 acoustic lens; in device-free acoustic sensing experiments, the microphone array is $3cm$ above the acoustic lens to track the distance and AoA of a person's hand so that the signal from the speaker to the target goes through the metasurface and the signal reflected from the target and received by the microphone array does not go through the metasurface; in acoustic communication experiments, the receiver is at $1.5m$ away from the speaker array. We also evaluate the impact of various parameters by varying their values.

5.2 SNR Comparison

We first compare various schemes in terms of SNR.

5.2.1 Beam Pattern

We place a receiver at $1.5m$ away, $0°$ from the speaker(s) and measure the sound field intensity. Figure 13(a)(c)(e) compare the beam patterns of six schemes in COMSOL simulation [2], and Figure 13(b)(d)(f) compare them in testbed. The six schemes include: (i) a single speaker without lens (**w/o PA + w/o lens**), (ii) a single speaker with a lens (**w/o PA + w/ lens**), (iii) a phased array without lens (**w/ PA + w/o lens**), (iv) a phased array with a lens (**w/ PA + w/ lens**), (v) an optimized phased array without lens (**w/ opt-PA + w/o lens**), and (vi) an optimized phased array with a lens (**w/ opt-PA + w/ lens**). Our goal is to focus the transmission signal in any desired direction. As we can see, (vi) yields the highest peak in the desired angle, which is 1.2, 2.9, 10.5, 10.5, and $18.4dB$ higher than (iv), (ii), (v), (iii), and (i), respectively. Optimized array placement yields $1.2dB$ gain over uniform placement. Leveraging acoustic lens yields $15.5dB$ gain when applied to a single lens, but its angle cannot be adapted and is always fixed at $0°$. Combining a phased array having a uniform separation with an acoustic lens allows us to focus the beam in the desired direction while achieving $17.2dB$ gain over a single speaker without lens and $9.3dB$ gain over a phased array without lens.

We further evaluate the signal strength and vary the steering angle, as shown in Figure 14. The improvement of (vi) over (iv) shows the benefit of the optimized array placement, and the improvement of (vi) over the other schemes shows the benefit of combining lens and phased array in the optimized placement. These results show that (vi) yields a high SNR gain across a wide angle from $-60°$ to $60°$.

(a) Simulation: 0°.

(b) Testbed: 0°.

(c) Simulation: −30°.

(d) Testbed: −30°.

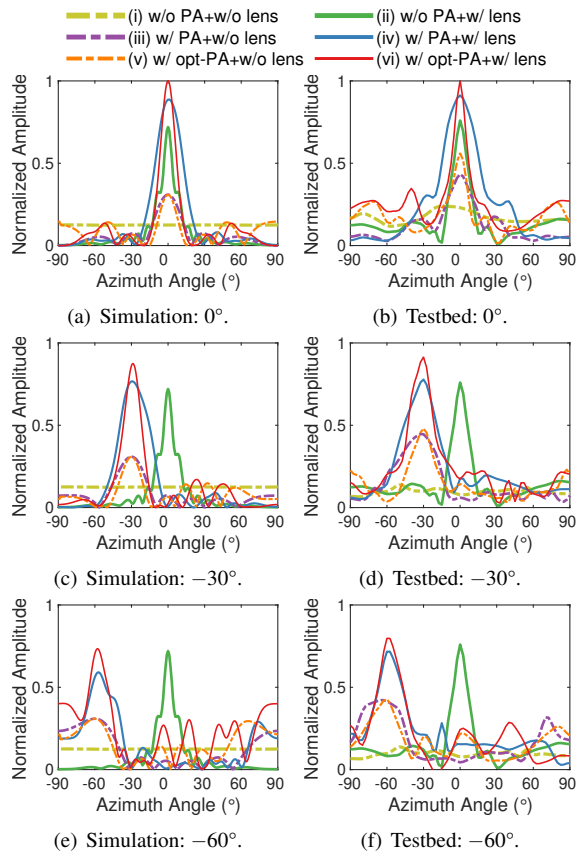(e) Simulation: −60°.

(f) Testbed: −60°.

Figure 13: The amplitude of the acoustic signal at various angles of steering. The amplitude is normalized by dividing it by the maximum amplitude of the signal received at 0°. (a)(c)(e) show the results from the simulation using COMSOL and (b)(d)(f) show the results from the testbed.

5.2.2 Beam Width

The 2D structure of our lens design allows us to focus beams in both the azimuth and elevation directions. To show the impact of the beam width in both directions, we measure the sound field in a far field plane, which is parallel to the surface of the lens. Figure 15 plots the sound field of three different beams steered to 30°, 60°, and 90°, respectively. As we can see, the linear phased array can only focus beams in the azimuth direction. In comparison, the acoustic lens can focus beams in both the azimuth and elevation directions. According to COMSOL simulation shown in Figure 16, the acoustic lens with a 6-speaker phased array generates a comparable beam pattern to a 16×1 array in the elevation direction and a comparable beam pattern to a 9×1 array in the azimuth direction. Therefore, the acoustic lens with a 6-speaker phased array is comparable to a $9 \times 16 = 144$ phased array in terms of beam width. This is a significant reduction in cost, size, energy, and computation. Moreover, we also find that equipping a 6-speaker phased array with acoustic lenses of sizes
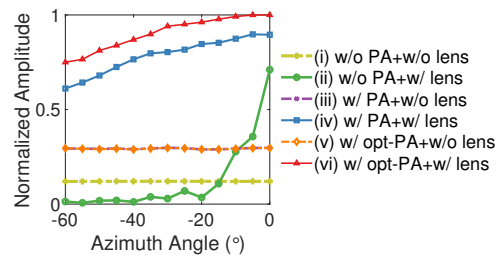

Figure 14: Power gain while steering at various angles. The amplitude is normalized by dividing it by the maximum amplitude of scheme (vi). The testbed uses a 16×16 acoustic lens and 6 speakers to form a phased array.

(a) 90°, w/ lens.

(b) 60°, w/ lens.

(c) 30°, w/ lens.

(d) 90°, w/o lens.
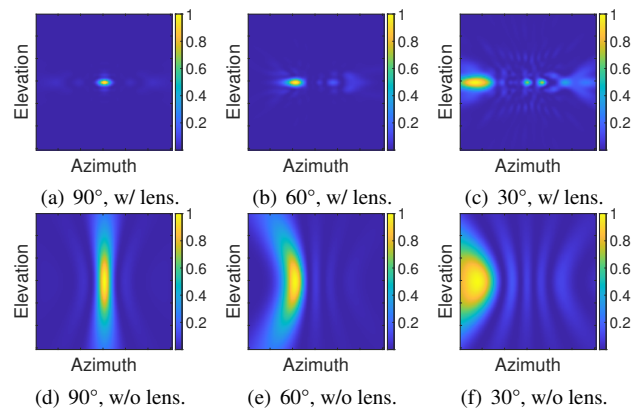
(e) 60°, w/o lens.

(f) 30°, w/o lens.

Figure 15: In the COMSOL simulation, comparing the sound signal strength using a phased array (6 speakers) with and without the acoustic lens in the azimuth direction.

32×32, 48×48, and 64×64 yield similar beam patterns to 16×28, 20×40, and 30×48 phased arrays, respectively.

5.2.3 Frequency Response

Our acoustic metasurfaces are designed for $20kHz$ sound, but we use 16-$20kHz$ and 18-$20kHz$ for acoustic sensing and communication, respectively. To understand how the lens works at a different frequency, we test the frequency response of the lens. We first calibrate the speaker(s) and microphone(s) and use compensation to generate close to a flat frequency response in the received signal. We then transmit a sine wave with a frequency varying from $10kHz$ to $22kHz$ and record the received sound at $1m$ and $0°$ from the speaker. We test all lens configurations, including a single-speaker lens, uniform phased-array lens, and non-uniform phased-array lens. Fig. 17 shows the frequency response. As expected, the peak of the lens frequency response is $20kHz$. It drops rapidly after $20kHz$. Fortunately, the frequency response remains stable in $14kHz$-$20kHz$, which means that we can use this band for sensing and communication.

(a) Elevation Dimension.
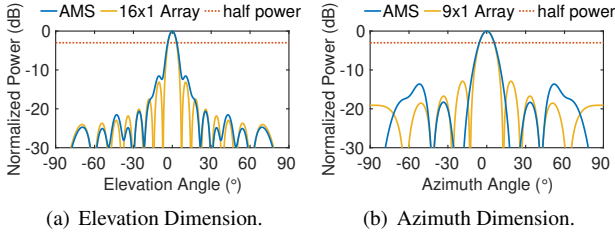
(b) Azimuth Dimension.

Figure 16: Using a 16×16 acoustic lens with a 6-speaker phased array is comparable to a 9×16 phased array in terms of beam width in the elevation and azimuth direction according to COMSOL simulation.
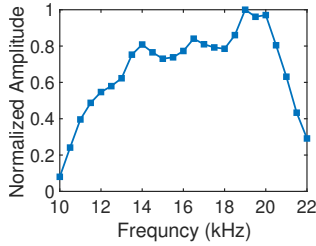

Figure 17: The frequency response of the acoustic lens.

5.2.4 Impact of Lens Size and Phased Array Size

Figure 18(a) plots the received power of (vi) in the desired direction using different lens sizes. We only plot $-60°$ to $0°$ as the result from $0°$ to $60°$ is symmetric. We make several observations. First, as we would expect, increasing the lens sizes improves the normalized power (i.e., the received power is normalized by dividing by the power of 32×32 lens at $0°$ and $1.5m$ away). In particular, the normalized power increases from 0.29 using 8×8 lens to 0.42 using 12×12 lens, to 0.62 using 16×16 lens, to 0.76 using 24×24 lens, and to 0.83 using 32×32 lens at $-60°$. Second, comparing the simulation (solid curves) with the testbed results (dotted curves), we observe high consistency, which validates the fidelity of our simulation.

Figure 18(b) further plots the normalized power (i.e., the received power is normalized by dividing by the power of $m = 16$ phased array at $0°$ and $1.5m$ away) of (iv) as we vary the number of speakers. As expected, increasing the number of speakers from 2 to 4, 6, 8, and 16 increases the normalized power to 1.3x, 1.6x, 1.8x and 2.1x, respectively.

5.3 Distance Estimation Performance

Next we evaluate the impact of our approach on distance estimation accuracy. As described in Section 5.1, we evaluate the benefit of acoustic lens and phased array on the well-known distance estimation techniques: FMCW.

Figure 19(a) plots the errorbar of the distance estimation error, where the center, lower bound, and upper bound of the errorbar correspond to median, 25%, 75% of the distance error, respectively. (vi) performs the best and out-performs
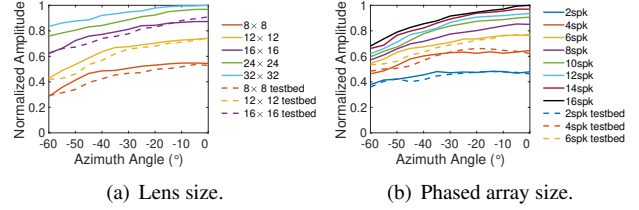

(a) Lens size.

(b) Phased array size.

Figure 18: The impact of lens size and phased array size on the received power of (iv). Dotted curves represent results from the testbed experiments while solid curves represent results from COMSOL simulation.

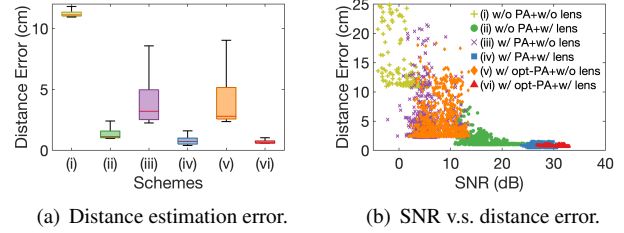(a) Distance estimation error.

(b) SNR v.s. distance error.

Figure 19: The distance estimation performance using acoustic lens.

(i), (ii), (iii), (iv), and (v) by 93%, 25%, 76%, 12%, and 72%, respectively. To better understand where the improvement comes from, we plot the distance estimation error of each FMCW chirp and its corresponding SNR in Figure 19(b). We can see that, as expected, the FMCW chirp with a higher SNR leads to a lower error. The optimized phased array with acoustic lens has the highest SNR, so it yields the lowest distance estimation error.

5.3.1 Impact of Measurement Distance

We evaluate the impact of distance on the distance estimation error as shown in Figure 20. As we can see in Figure 20(a), (vi) performs the best and out-performs (i), (ii), (iii), (iv), and (v) by 94%, 45%, 79%, 18%, and 81%, respectively. (vi) increases the operation distance from $0.5m$ in (i) to $3.5m$. Note that $3.5m$ operation distance is very good considering the total power of our 6 speakers is only $0.1W$. In comparison, [38] achieves $4.5m$ operation distance using a $2.5W$ speaker.

5.3.2 Impact of Measurement Angle

Next we further show the impact of measurement angle on the distance estimation performance. From Figure 20(b)-(e), we can see that (vi) still performs the best in all directions and out-performs (i), (ii), (iii), (iv), and (v) by 93%, 99%, 77%, 21%, and 76%, respectively. However, when we increase the measurement angle from $0°$ to $15°$, $30°$, $45°$, and $60°$, the distance error of (vi) also increases from $1.28cm$ to $1.30cm$, $1.32cm$, $1.86cm$, and $2.47cm$, respectively.
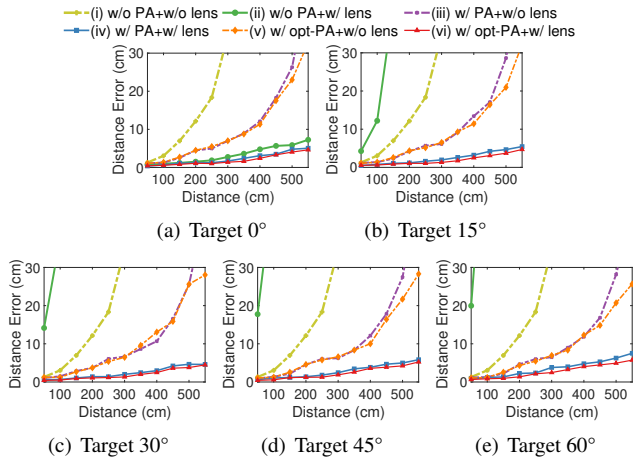
(a) Target 0° (b) Target 15°

(c) Target 30° (d) Target 45° (e) Target 60°

Figure 20: The impact of distance on the distance estimation error while steering in various directions.
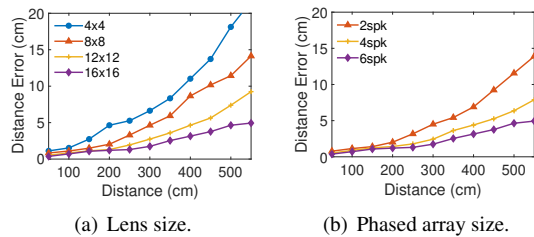

(a) Lens size. (b) Phased array size.

Figure 21: The impact of lens size and phased array size on the distance estimation error of (iv).

5.3.3 Impact of Lens Size and Phased Array Size

We evaluate the impact of lens size on distance estimation performance. We experiment with 4×4, 8×8, 12×12 and 16×16 lenses in our testbed. As shown in Figure 21(a), increasing the lens size improves the distance estimation errors. A 16×16 lens reduces the error over a 4×4, 8×8, and 12×12 lens by 73%, 62%, and 36%, respectively.

We also evaluate the impact of phased array size. As shown in Figure 21(b), increasing the array size improves the distance estimation due to the enhanced SNR. For example, a 6-speaker phased array with AMS reduces the distance estimation error over 1-, 2-, and 4-speaker phased array by 95%, 91%, and 80%, respectively.

5.4 AoA Estimation Performance

In this section, we compare the AoA estimation using 1D MUSIC as introduced in Section 5.1.

5.4.1 Impact of Measurement Distance

Figure 22(a) plots the AoA error versus the measurement distance. In all schemes, the distance estimation errors increase with an increasing distance. (vi) performs the best and reduces the AoA error of (i), (ii), (iii), (iv), and (v) by 92%, 44%, 76%, 16%, and 77%, respectively. The result shows that
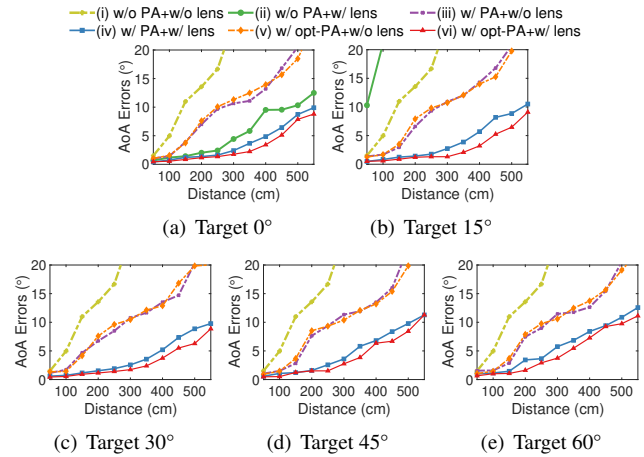

(a) Target 0° (b) Target 15°

(c) Target 30° (d) Target 45° (e) Target 60°

Figure 22: The impact of distance on AoA estimation error while steering in various directions.
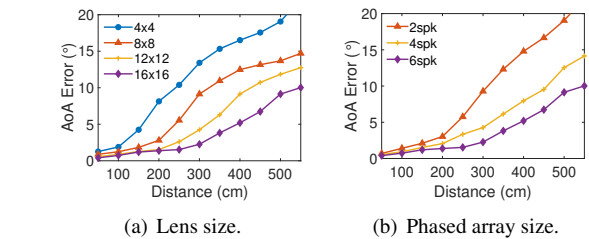

(a) Lens size. (b) Phased array size.

Figure 23: The impact of lens size and phased array size on the AoA estimation error of (iv).

our approach effectively improves AoA sensing accuracy by increasing the SNR.

5.4.2 Impact of Measurement Angle

We further show the impact of measurement angle on AoA estimation. Figure 22(b)-(e) plots AoA estimation error for other directions. When the distance is small, the AoA estimation error remains low across all measurement angles of interest; when the distance is large, the AoA error increases more rapidly with the increasing angle. This is expected because when SNR is sufficiently high, the measurement angle has less impact; but when SNR is low, the measurement angle matters. (vi) out-performs (i), (ii), (iii), (iv), and (v) by 90%, 96%, 81%, 26%, and 82%, respectively.

5.4.3 Impact of Lens Size and Phased Array Size

As shown in Figure 23(a), increasing the lens size effectively reduces the AoA estimation error. For example, increasing the lens size from 4×4 to 8×8 improves AoA estimation by 32%, increasing from 8×8 to 12×12 improves by 53%, and increasing from 12×12 to 16×16 further improves by 46%.

Figure 23(b) further plots the AoA estimation as we vary the array size. Increasing the array size improves SNR and reduces the AoA estimation error. Using a 6-speaker phased array reduces the AoA error by 96%, 74%, and 52% over
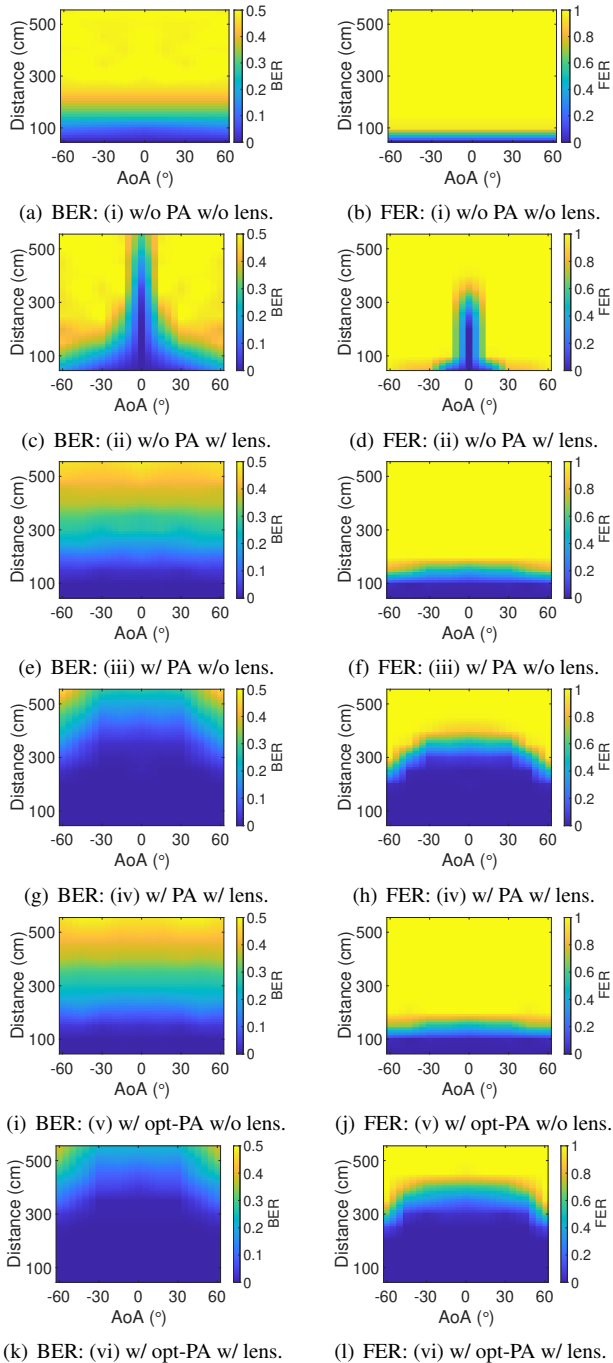
(a) BER: (i) w/o PA w/o lens.

(b) FER: (i) w/o PA w/o lens.

(c) BER: (ii) w/o PA w/ lens.

(d) FER: (ii) w/o PA w/ lens.

(e) BER: (iii) w/ PA w/o lens.

(f) FER: (iii) w/ PA w/o lens.

(g) BER: (iv) w/ PA w/ lens.

(h) FER: (iv) w/ PA w/ lens.

(i) BER: (v) w/ opt-PA w/o lens.

(j) FER: (v) w/ opt-PA w/o lens.

(k) BER: (vi) w/ opt-PA w/ lens.

(l) FER: (vi) w/ opt-PA w/ lens.

Figure 24: The impact of distance and AoA on the communication performance.

using 1-, 2-, and 4-speaker phased array, respectively.

5.5 Acoustic Communication Performance

Finally, we evaluate the impact of our approach on acoustic communication performance.

5.5.1 Impact of Distance

We put the speakers at a fixed location and gradually increase the distance between the microphone and the speakers. We tested the above six scenarios for communication. Figure 24 shows the bit error rate (BER) and frame error rate (FER) across different AoAs at different distances. We can see that both BER and FER increase with the distance. Due to the use of fixed modulation and FEC, the FER rapidly increases after the distance goes beyond a certain point. If we define the communication range as the range corresponding to 50% FER, we observe that (vi) has 3.9m communication range as shown in Figure 24(l). In comparison, (i), (ii), (iii), (iv), and (v) have communication ranges of 0.8m, 3.0m, 1.5m, 3.5m, and 1.5m, respectively.

5.5.2 Impact of AoA

Figure 24 also shows the impact of AoA on BER and FER in our testbed. (vi) achieves low error rates within 60° and 2.8m. This is good coverage considering the total power of our 6 speakers is only 20mW. In comparison, for the same 60° coverage, the other schemes' range is much smaller.

6 Discussion

Acoustic metasurfaces can effectively boost the signal quality, and improve sensing and communication performance. Compared with a large phased array, our approach of using a small phased array and metasurface is more compact, cost effective, and energy efficient. To fully realize the potential of AMS, several challenges remain to be addressed in the future: (i) further reducing the AMS size so that it can be applied to more applications (e.g., mobile devices), (ii) supporting a wider band, and (iii) further simplifying fabrication process. Figure 17 shows that our AMS can support 16-20KHz within 20% amplitude loss. This is sufficient for our purpose but may need further improvement if a wider band is required.

7 Conclusion

In this paper, we develop a novel acoustic system that uses AMS and multiple speakers together to achieve dynamic steering and high SNR. The increase in SNR can be translated into higher accuracy in distance and AoA estimation and larger communication range in acoustic communication. Encouraged by the promising results, we plan to explore more applications that can benefit from our design.

Acknowledgments

References

[1] Bela: create beautiful interaction with sensors and sound. https://bela.io/, 2022.

[2] COMSOL: simulate real-world designs, devices, and processes with multiphysics software from comsol. https://www.ti.com/product/LM386, 2023.2.

[3] LM386: 700-mw, mono, 5- to 18-v, analog input class-ab audio amplifier. https://www.ti.com/product/LM386, 2023.2.

[4] THS4001: 270-mhz voltage-feedback amplifier. https://www.ti.com/product/THS4001, 2023.2.

[5] I. F. Akyildiz, D. Pompili, and T. Melodia. State-of-the-art in protocol research for underwater acoustic sensor networks. In *Proc. of WUWNet*, 2006.

[6] Badreddine Assouar, Bin Liang, Ying Wu, Yong Li, Jian-Chun Cheng, and Yun Jing. Acoustic metasurfaces. *Nature Reviews Materials*, 3(12):460–472, 2018.

[7] Yang Bai, Nakul Garg, and Nirupam Roy. Spidr: Ultra-low-power acoustic spatial sensing for micro-robot navigation. In *Proceedings of the 20th Annual International Conference on Mobile Systems, Applications and Services*, pages 99–113, 2022.

[8] Rainer Bauböck. Migration and citizenship. *Journal of Ethnic and Migration Studies*, 18(1):27–48, 1991.

[9] Tuochao Chen, Justin Chan, and Shyam Gollakota. Underwater messaging using mobile devices. In *Proc. of ACM SIGCOMM*, 2022.

[10] Xing Chen, Peng Liu, Zewei Hou, and Yongmao Pei. Magnetic-control multifunctional acoustic metasurface for reflected wave manipulation at deep subwavelength scale. *Scientific reports*, 7(1):1–9, 2017.

[11] Krishna Chintalapudi, Venkat Padmanabhan, and Ramarathnam Venkatesan. Dhwani: Secure peer-to-peer acoustic nfc. In *Proc. of ACM SIGCOMM*, 2013.

[12] Mandar Chitre, Shiraz Shahabudeen, and Milica Stojanovic. Underwater acoustic communications and networking: Recent advances and future challenges. *Marine Technology Society Journal*.

[13] Nicholas Fang, Dongjuan Xi, Jianyi Xu, Muralidhar Ambati, Werayut Srituravanich, Cheng Sun, and Xiang Zhang. Ultrasonic metamaterials with negative modulus. *Nature materials*, 5(6):452–456, 2006.

[14] Romain Fleury and Andrea Alù. Extraordinary sound transmission through density-near-zero ultranarrow channels. *Physical review letters*, 111(5):055501, 2013.

[15] Zhihui Gao, Ang Li, Dong Li, Jialin Liu, Jie Xiong, Yu Wang, Bing Li, and Yiran Chen. Mom: Microphone based 3d orientation measurement. In *Proc. of ACM/IEEE IPSN*, 2022.

[16] Nakul Garg, Yang Bai, and Nirupam Roy. Owlet: enabling spatial information in ubiquitous acoustic devices. In *Proc. of ACM MobiSys*, 2021.

[17] Unsoo Ha, Junshan Leng, Alaa Khaddaj, and Fadel Adib. Food and liquid sensing in practical environments using rfids. In *17th USENIX Symposium on Networked Systems Design and Implementation (NSDI 20)*, pages 1083–1100, 2020.

[18] Fu-Li Hsiao, Ting-Kuo Li, Pin-Chieh Chen, Su-Chao Wang, Ke-Wei Lin, Wei-Ling Lin, Ying-Pin Tsai, Wen-Kai Lin, and Bor-Shyh Lin. Phase resonance and sensing application of an acoustic metamaterial based on a composite both-sides-open disk resonator arrays. *Sensors and Actuators A: Physical*, 339:113524, 2022.

[19] Yun Jing, Jun Xu, and Nicholas X Fang. Numerical study of a near-zero-index acoustic metamaterial. *Physics Letters A*, 376(45):2834–2837, 2012.

[20] Hao Kong, Xiangyu Xu, Jiadi Yu, Qilin Chen, Chenguang Ma, Yingying Chen, Yi-Chao Chen, and Linghe Kong. m3track: mmwave-based multi-user 3d posture tracking. In *Proc. of ACM MobiSys*, 2022.

[21] Manikanta Kotaru, Kiran Joshi, Dinesh Bharadia, and Sachin Katti. Spotfi: Decimeter level localization using WiFi. In *ACM SIGCOMM Computer Communication Review*, volume 45(4), pages 269–282. ACM, 2015.

[22] Manikanta Kotaru and Sachin Katti. Position tracking for virtual reality using commodity wifi. In *Proceedings of the 10th on Wireless of the Students, by the Students, and for the Students Workshop*, S3 '18, pages 15–17, New York, NY, USA, 2018. ACM.

[23] Hyewon Lee, Tae Hyun Kim, Jun Won Choi, and Sunghyun Choi. Chirp signal-based aerial acoustic communication for smart devices. In *2015 IEEE Conference on Computer Communications (INFOCOM)*, pages 2407–2415. IEEE, 2015.

[24] Kyung Hoon Lee, Kunhao Yu, An Xin, Zhangzhengrong Feng, Qiming Wang, et al. Sharkskin-inspired magnetoactive reconfigurable acoustic metamaterials. *Research*, 2020, 2020.

[25] Dong Li, Jialin Liu, Sunghoon Ivan Lee, and Jie Xiong. Lasense: Pushing the limits of fine-grained activity sensing using acoustic signals. In *Proc. of IMWUT/UbiComp*, 2022.

[26] Jensen Li and Che Ting Chan. Double-negative acoustic metamaterial. *Physical Review E*, 70(5):055602, 2004.

[27] Yong Li, Xue Jiang, Rui-qi Li, Bin Liang, Xin-ye Zou, Lei-lei Yin, and Jian-chun Cheng. Experimental realization of full control of reflected waves with subwavelength acoustic metasurfaces. *Physical Review Applied*, 2(6):064002, 2014.

[28] Yong Li, Bin Liang, Zhong-ming Gu, Xin-ye Zou, and Jian-chun Cheng. Reflected wavefront manipulation based on ultrathin planar acoustic metasurfaces. *Scientific reports*, 3(1):1–6, 2013.

[29] Zixian Liang and Jensen Li. Extreme acoustic metamaterial by coiling up space. *Physical review letters*, 108(11):114301, 2012.

[30] Qiongzheng Lin, Zhenlin An, and Lei Yang. Rebooting ultrasonic positioning systems for ultrasound-incapable smart devices. In *The 25th Annual International Conference on Mobile Computing and Networking*, pages 1–16, 2019.

[31] Yuan Liu, Bo Jiu, and Hongwei Liu. Admm-based transmit beampattern synthesis for antenna arrays under a constant modulus constraint. *Signal Processing*, 171:107529, 2020.

[32] Guancong Ma and Ping Sheng. Acoustic metamaterials: From local resonances to broad horizons. *Science advances*, 2(2):e1501595, 2016.

[33] YONGSEN MA, GANG ZHOU, and SHUANGQUAN WANG. Wifi sensing with channel state information: A survey. *ACM Comput. Survey*, June 2019.

[34] Yunfei Ma, Nicholas Selby, and Fadel Adib. Minding the billions: Ultra-wideband localization for deployed rfids. In *Proc. of ACM MobiCom*, 2017.

[35] Wenguang Mao, Jian He, and Lili Qiu. CAT: high-precision acoustic motion tracking. In *Proc. of ACM MobiCom*, 2016.

[36] Wenguang Mao, Wei Sun, Mei Wang, and Lili Qiu. Deeprange: Ranging via deep learning. In *Proc. of UbiComp*, 2021.

[37] Wenguang Mao, Mei Wang, and Lili Qiu. Aim: Acoustic imaging on a mobile. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*, pages 468–481. ACM, 2018.

[38] Wenguang Mao, Mei Wang, Wei Sun, Lili Qiu, Swadhin Pradhan, and Yi-Chao Chen. Rnn-based room scale hand motion tracking. In *The 25th Annual International Conference on Mobile Computing and Networking*, pages 1–16, 2019.

[39] Wenguang Mao, Zaiwei Zhang, Lili Qiu, Jian He, Yuchen Cui, and Sangki Yun. Indoor follow me drone. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*, pages 345–358. ACM, 2017.

[40] Gianluca Memoli, Mihai Caleap, Michihiro Asakawa, Deepak R. Sahoo, Bruce W. Drinkwater, and Sriram Subramanian. Metamaterial bricks and quantization of meta-surfaces. *Nature Communication*, 2017.

[41] Gianluca Memoli, Letizia Chisari, Jonathan P. Eccles, Mihai Caleap, Bruce W. Drinkwater, and Sriram Subramanian. Vari-sound: A varifocal lens for sound. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–14, New York, NY, USA, 2019. Association for Computing Machinery.

[42] Rajalakshmi Nandakumar, Shyam Gollakota, and Nathaniel Watson. Contactless sleep apnea detection on smartphones. In *Proc. of ACM MobiSys*, 2015.

[43] Rajalakshmi Nandakumar, Vikram Iyer, Desney Tan, and Shyamnath Gollakota. FingerIO: Using active sonar for fine-grained finger tracking. In *Proc. of ACM CHI*, pages 1515–1525, 2016.

[44] Jong Jin Park, KJB Lee, Oliver B Wright, Myoung Ki Jung, and Sam H Lee. Giant acoustic concentration by extraordinary transmission in zero-mass metamaterials. *Physical review letters*, 110(24):244302, 2013.

[45] Chunyi Peng, Guobin Shen, Yongguang Zhang, Yanlin Li, and Kun Tan. BeepBeep: a high accuracy acoustic ranging system using COTS mobile devices. In *Proc. of ACM SenSys*, 2007.

[46] Swadhin Pradhan, Shuozhe Li, and Lili Qiu. Rotation sensing using passive rfid tags. In *Proc. of MobiHoc*, 2021.

[47] Qifan Pu, Sidhant Gupta, Shyam Gollakota, and Shwetak Patel. Whole-home gesture recognition using wireless signals. In *Proc. of ACM MobiCom*, 2013.

[48] Saeed Ur Rahman, Qunsheng CAO, Muhammad Mansoor Ahmed, and Hisham Khalil. Analysis of linear antenna array for minimum side lobe level, half power beamwidth, and nulls control using pso. *Journal of Microwaves, Optoelectronics and Electromagnetic Applications*, 16:577–591, 2017.

[49] Sheng Shen, Daguan Chen, Yu-Lin Wei, Zhijian Yang, and Romit Roy Choudhury. Voice localization using nearby wall reflections. In *Proc. of ACM MobiCom*, 2020.

[50] Deepak Vasisht, Swarun Kumar, and Dina Katabi. Decimeter-level localization with a single wifi access point. In *13th USENIX Symposium on Networked Systems Design and Implementation (NSDI 16)*, pages 165–178, 2016.

[51] Anran Wang and Shyamnath Gollakota. Millisonic: Pushing the limits of acoustic motion tracking. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–11, 2019.

[52] Anran Wang, Jacob E Sunshine, and Shyamnath Gollakota. Contactless infant monitoring using white noise. In *The 25th Annual International Conference on Mobile Computing and Networking*, pages 1–16, 2019.

[53] Ju Wang, Jie Xiong, Xiaojiang Chen, Hongbo Jiang, Rajesh Krishna Balan, and Dingyi Fang. Tagscan: Simultaneous target imaging and material identification with commodity rfid devices. In *Proc. of ACM MobiCom*, pages 288–300. ACM, 2017.

[54] Jue Wang, Fadel Adib, Ross Knepper, Dina Katabi, and Daniela Rus. RF-compass: robot object manipulation using RFIDs. In *Proc. of the 19th annual international conference on Mobile computing and networking*, pages 3–14, 2013.

[55] Jue Wang and Dina Katabi. Dude, where's my card? RFID positioning that works with multipath and non-line of sight. In *Proc. of the ACM SIGCOMM*, pages 51–62, 2013.

[56] Mei Wang, Wei Sun, and Lili Qiu. Mavl: Multiresolution analysis of voice localization. In *Proc. of NSDI*, 2021.

[57] Wei Wang, Alex X Liu, and Ke Sun. Device-free gesture tracking using acoustic signals. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*, pages 82–94. ACM, 2016.

[58] Teng Wei and Xinyu Zhang. mTrack: high precision passive tracking using millimeter wave radios. In *Proc. of ACM MobiCom*, 2015.

[59] Teng Wei and Xinyu Zhang. Gyro in the air: Tracking 3d orientation of batteryless internet-of-things. In *Proceedings of MobiCom*, MobiCom '16, pages 55–68, New York, NY, USA, 2016. ACM.

[60] Yangbo Xie, Wenqi Wang, Huanyang Chen, Adam Konneker, Bogdan-Ioan Popa, and Steven A Cummer. Wavefront modulation and subwavelength diffractive acoustics with an acoustic metasurface. *Nature communications*, 5(1):5553, 2014.

[61] Hongfei Xue, Yan Ju, Chenglin Miao, Yijiang Wang, Shiyang Wang, Aidong Zhang, and Lu Su. mmmesh: towards 3d real-time dynamic human mesh construction using millimeter-wave. In *Proc. of MobiSys*, 2021.

[62] Min Yang, Guancong Ma, Ying Wu, Zhiyu Yang, and Ping Sheng. Homogenization scheme for acoustic metamaterials. *Physical Review B*, 89(6):064309, 2014.

[63] Min Yang, Guancong Ma, Zhiyu Yang, and Ping Sheng. Coupled membranes with doubly negative mass density and bulk modulus. *Physical review letters*, 110(13):134301, 2013.

[64] Sangki Yun, Yi chao Chen, and Lili Qiu. Turning a mobile device into a mouse in the air. In *Proc. of ACM MobiSys*, May 2015.

[65] Sangki Yun, Yi-Chao Chen, Huihuang Zheng, Lili Qiu, and Wenguang Mao. Strata: Fine-grained acoustic-based device-free tracking. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*, pages 15–28. ACM, 2017.

[66] Bingsheng Zhang, Qin Zhan, Si Chen, Muyuan Li, Kui Ren, Cong Wang, and Di Ma. Enabling keyless secure acoustic communication for smartphones. *IEEE internet of things journal*, 1(1):33–45, 2014.

[67] Zengbin Zhang, David Chu, Xiaomeng Chen, and Thomas Moscibroda. Swordfight: Enabling a new class of phone-to-phone action games on commodity phones. In *Proc. of ACM MobiSys*, 2012.