Scholars' Mine

Doctoral Dissertations

Student Theses and Dissertations

Spring 2022

Array signal processing for source localization and enhancement

Sohel J. Patel

Follow this and additional works at: https://scholarsmine.mst.edu/doctoral_dissertations

Part of the Electrical and Computer Engineering Commons

Department: Electrical and Computer Engineering

Recommended Citation

Patel, Sohel J., "Array signal processing for source localization and enhancement" (2022). *Doctoral Dissertations*. 3159.
https://scholarsmine.mst.edu/doctoral_dissertations/3159

This thesis is brought to you by Scholars' Mine, a service of the Missouri S&T Library and Learning Resources. This work is protected by U. S. Copyright Law. Unauthorized use including reproduction for redistribution requires the permission of the copyright holder. For more information, please contact scholarsmine@mst.edu.

ARRAY SIGNAL PROCESSING FOR SOURCE LOCALIZATION AND ENHANCEMENT

by

SOHEL JAYESH PATEL

A DISSERTATION

Presented to the Graduate Faculty of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

In Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

in

ELECTRICAL ENGINEERING

2022

Approved by:

Dr. Maciej Zawodniok, Advisor
Dr. Kurt Kosbar
Dr. Joe Stanley
Dr. Jiangfan Zhang
Dr. Venkata Sriram Siddhardh Nadendla

Copyright 2022

SOHEL JAYESH PATEL

All Rights Reserved

PUBLICATION DISSERTATION OPTION

This dissertation consists of the following five articles which have been accepted for publication, or submitted for publication as follows:

Paper I: Pages 6-17 have been accepted by the International Workshop on Acoustic Signal Enhancement (IWAENC).

Paper II: Pages 18-47 are intended for submission to the Journal of the Acoustical Society of America (JASA).

Paper III: Pages 48-66 have been accepted by the International Instrumentation & Measurement Technology Conference (I^2MTC).

Paper IV: Pages 67-81 have been accepted by the International Instrumentation & Measurement Technology Conference (I^2MTC).

Paper V: Pages 82-112 have been accepted by the Transactions on Instrumentation and Measurement Special Section (TIM Special Section).

ABSTRACT

A common approach to the wide-band microphone array problem is to assume a certain array geometry and then design optimal weights (often in subbands) to meet a set of desired criteria. In addition to weights, we consider the geometry of the microphone arrangement to be part of the optimization problem. Our approach is to use particle swarm optimization (PSO) to search for the optimal geometry while using an optimal weight design to design the weights for each particle's geometry. The resulting directivity indices (DI's) and white noise SNR gains (WNG's) form the basis of the PSO's fitness function. Another important consideration in the optimal weight design are several regularization parameters. By including those parameters in the particles, we optimize their values as well in the operation of the PSO. The proposed method allows the user great flexibility in specifying desired DI's and WNG's over frequency by virtue of the PSO fitness function.

Although the above method discusses beam and nulls steering for fixed locations, in real time scenarios, it requires us to estimate the source positions to steer the beam position adaptively. We also investigate source localization of sound and RF sources using machine learning techniques. As for the RF source localization, we consider radio frequency identification (RFID) antenna tags. Using a planar RFID antenna array with beam steering capability and using received signal strength indicator (RSSI) value captured for each beam position, the position of each RFID antenna tag is estimated. The proposed approach is also shown to perform well under various challenging scenarios.

ACKNOWLEDGMENTS

I want to express my gratitude to my doctoral adviser, Dr. Maciej Zawodniok, for his continuous support during my PhD program. I'd like to thank him for presenting me with unique chances that pushed me in unanticipated directions. He continuously fostered a collaborative research atmosphere and decided to join forces with Dr. Daryl Beetner to guide me in one of my most challenging life situations.

Members of my doctoral study committee, Dr. Kurt Kosbar, Dr. Joe Stanley, Dr. Jiangfan Zhang, and Dr. Venkata Sriram Siddhardh Nadendla, deserve special gratitude. I'd like to thank prior members Dr. Yahong Rosa Zheng and Dr. Matt Insall for their contributions.

Dr. Vinay Kothapally, Mr. Dipesh Gaur, Mr. Deepak Melugiri, Mr. Nejav Shah, Mrs. Shreya Almaula and all of my friends, I thank you for your love and never-ending support. My parents, Mr. Jayesh A. Patel and Mrs. Sangita J. Patel, sister Dr. Riya R. Patel as well as my uncle Mr. Rakesh A. Patel and aunt Mrs. Alka R. Patel and all my family members deserve special thanks. Without their help, I would not have gotten this far.

I would like to thank Dr. Jacob Benesty and Dr. Jingdong Chen for their unwavering support and guidance, as well as Dr. Vladimir Tourbabin for providing me with the opportunity to work on intriguing research projects during my internship at Meta Reality Labs (f.k.a Facebook Reality Labs). Thank you Dr. Steven L. Grant, my prior Master's program advisor, for inspiring me to seek a PhD, as well as Mrs. Maria Grant, one of the bravest and generous women I've ever known.

TABLE OF CONTENTS

Page

SECTION

LIST OF ILLUSTRATIONS

Figure Page

PAPER I

PAPER II

PAPER III

PAPER IV

LIST OF TABLES

Table Page

PAPER IV

PAPER V

SECTION

1. INTRODUCTION

Beam former is a processor that performs spatial filtering along with the sensor arrays or typically known as array signal processing. Beamforming is achieved by combining elements in phased array so that it can receive the desired signal from the desired direction and attenuate the signals arriving from other direction. Beamforming provides spatial selectivity both for transmission and reception. It has many applications in RADAR, SONAR, Communications, Seismology, Geology, Biomedical, Acoustics and Radio Astronomy.

1.1. SPATIAL FILTERING

Spatial filtering performs what is known as transmitting or receiving signals preferably in some directions over others. Based on the angle of focus or the angle towards which the beam is steered. Suppose if we want to increase the signal to noise ratio at the output i.e. reducing the output power contribution made by noise, we try blocking/suppressing the noise coming from certain direction even without knowing the frequency of it provided it is non coherent like the FIR filters perform.

Most conventional beamformers consider a specific geometry and optimize weights to operate under certain noise conditions. Consider a situation where more performance is required and the array geometry cannot exceed a certain aperture size. In such cases, a technique that can maximize the DF and WNG with minimal array aperture needs to be developed. We address this issue in this paper using aperiodic arrangement of microphones.

The most basic beamforming technique delay-sum (DS) minimizes white noise but suffers from poor directivity. This can be overcome by maximizing the signal-to-noise ratio (SNR) under diffuse noise conditions known as superdirective beamformer [11, 7]. A beamformer that operates equally well under these noise conditions is desired and extensively studied in this paper.

Robust superdirective beamforming [11, 10] is obtained by putting a constraint on white noise gain (WNG). Recently, closed form solution that enables tuning between WNG and directivity factor (DF) [4, 5] was proposed. However, the optimization of this regularization parameter is not straightforward. Moreover, the frequency independent characteristics can be achieved for either WNG or DF. We address this issue in this study.

Several algorithms have been proposed to obtain frequency independent characteristics including nested arrays [23, 26], differential beamforming [14] among many others. Nested array provides a good performance for both lower and higher frequencies, where each band-passed signal is processed using a specific sub-array. However, array aperture of such arrays quickly increase when higher directivity is desired.

In the past, researchers have demonstrated that aperiodic or irregular microphone geometries outperform arrays with regular geometries [43, 31, 41, 32]. It is also proven that for a fixed number of microphones, the performance of an array is mainly dominated by its geometry [43, 31, 35]. Several techniques have been proposed to optimize the array geometry using parameters like maximum sidelobe level (MSL) [6]. Improvements in beam-forming have been observed with irregular microphone arrays by optimizing geometry parameters, viz., array centroid and dispersion [42]. However, this solution of using optimized irregular arrays for wideband characteristics like frequency invariant beampatterns were rarely examined. Stochastic optimization techniques for the geometries of wideband microphone arrays have been explained in [12]. However, they did not include the combined optimization of tradeoff between DI and WNG. Because, this would result in a nonlinear cost function.

The proposed design utilizes PSO technique [20] to optimize together the above mentioned parameters by developing a non-linear fitness function. We consider differential microphone arrays (DMAs) for smaller array apertures.

1.2. SOURCE LOCALIZATION

With widespread deployment of IoT, an indoor-based localization services (IBLS) are becoming a key requirement in such applications. In particular, the radio frequency identification (RFID) is a promising technology for such systems. In particular, RFID offers low cost, small form-factor, reduced size and cost of passive tags/sensors (battery-free operation), and can be easily deployed. The main challenge is hitherto low localization accuracy.

Several alternatives for an IBLS exist, including an ultra-wideband (UWB) technology, Bluetooth, ZigBee, infrared, wireless local area network (WLAN), RFID, etc. [21]. The UWB [16], infrared, and WLAN technology provide a robust localization performance with up to centimeter-level accuracy but are expensive both in terms infrastructure setup and cost of individual tags/devices. Bluetooth and ZigBee on the other hand provide low cost and low power solution. Whereas Bluetooth has poor accuracy, short range (usually within $10-30$ feet) and carries long latency [13]. ZigBee often performs localization based on the communication between nodes in a network. However, this often requires costly and time-consuming calibration or profiling of the entire network.

Many state of the art approaches have been proposed to accurately locate the RFID tag position through fine-grained localization techniques. Most successful works include [22, 37, 28] that adopt multilateration, hyperbolic based modeling, or an angle of arrival (AoA) estimation to determine the location of tagged objects. Few methods have adopted the concept of synthetic aperture radar (SAR) techniques [24, 27, 33, 36, 40, 39] to achieve localization accuracy in order of centimeters. Tagogram [40] uses a differential augmented hologram technique to suppress the RFID tag's phase shift. MobiTagbot [33] studies the correlation due to changing multipath reflections and carrier frequency channel with antenna in motion. However, the main, remaining challenge is reduction of uncertainties due to environmental factors where theoretical, geometry-based models are not sufficient.

Recently, Deep Learning (DL) based approaches have been proposed [39, 9, 34] to model these complex environmental factors. FaHo [39] employs a fine grained joint hologram technique similar to SAR. The convolutional neural network (CNN) is trained with the holograms as images to achieve *cm* level accuracy. 3DLRA [9] deploys 5 antennas at known locations to sort books in shelves. The RSSI, phase, and timestamp are often collected continuously and correlated with the reference tag positions in order to estimate the absolute position of a tag under test. PRDL [34] is based on single antenna moving in a linear path, parallel to the tags to be located. It uses the measured RSSI and phase values to compute a relative position of tags.

The DL-based techniques discussed so far i.e. [39, 9, 34] operate at a short range. At such distances the multi path effects are not prevalent. Moreover, the techniques discussed in [39]and [34] require moving reader antennas that increases overall cost of a system and makes it impractical in many deployment scenarios. The existing techniques discuss heretofore either require high precision equipment or suffer from poor accuracy when setup using low cost equipment. In addition, several prior assumptions are made such as placing antennas in known locations with multiple readers that often require periodic, tedious calibration or profiling.

Hence, we propose the indoor tag localization based on ML approach with minimal prior assumptions. We use a planar, 2×2 phased antenna array with electrically-steered beam that is connected to a single, off-the-shelf RFID reader. We demonstrate improved accuracy of the DOA estimation by training a DNN that can map complex environment interactions Solely using RSSI values.

RSSI is the most commonly available measurement of the signal received from an RFID tag. This is one of reasons why it has been heavily exploited by several researchers for tag localization. One of the first proposed method using RSSI is SpotOn [17]. Although it has practically not been implemented yet. Another interesting method was proposed in LANDMARC [25]. This method uses multiple reader antennas and reference tags placed

in known locations. Based on the received strength of nearest reference tags, the location of active tag may be estimated. Several alternatives to this localization approach have been proposed [38, 18, 44, 3, 45] in order to reduce the number of reference tags and improved robustness with accuracy up to 1meter. However, RSSI is sensitive to multipath effect and has limited accuracy when compared to TOA or PDOA methods. We address these issues in this paper - including overcoming the RSSI multipath effect.

PAPER

I. ON THE DESIGN OF OPTIMAL LINEAR MICROPHONE ARRAY GEOMETRIES

Sohel J. Patel, Steven L. Grant, Maciej Zawodniok
Department of Electrical & Computer Engineering
Missouri University of Science and Technology
Rolla, Missouri 65409
Email: sjp432@mst.edu, sgrant@mst.edu, mjzx9c@mst.edu

Jacob Benesty
INRS-EMT
University of Quebec, Montreal, Canada
Montreal, Canada
Email: Jacob.Benesty@inrs.ca

ABSTRACT

In this paper, we present a simultaneous optimization model for aperiodic linear microphone array geometry and weights. Desirable properties for Broadband arrays include robust superdirective frequency invariant beampatterns. Our approach is to employ particle swarm optimization (PSO) to search for the optimal geometry while selecting optimal weights for each particle's geometry. The resulting directivity factor (DFs) and white noise gains (WNGs) are used to define the PSO fitness function. The proposed approach also optimizes the trade-off between WNG and DF, to find a geometry within a given aperture, thus, maximizing both these parameters. The proposed method allows the user great flexibility in specifying desired DFs and WNGs over frequency by virtue of the PSO fitness function. The resultant array geometry is smaller and yields greater WNG and DF than conventional approach.

1. SIGNAL MODEL

We consider a plane wave, propagating in the far field at 340 m/s in an anechoic acoustic environment, and impinging on a uniform linear array consisting of M omnidirectional microphones. The steering vector (of length M) is given by

$$\mathbf{d}^H(\omega, \alpha_{P,n}) = \begin{pmatrix} e^{-j\omega\tau_0\alpha_{P,n}} & \dots & e^{-j\omega\tau_{M-1}\alpha_{P,n}} \end{pmatrix}, \quad n = 0, 1, \dots, P \tag{1}$$

where

$$\alpha_{P,n} = \frac{2\pi(n-1)}{P} \tag{2}$$

and $j = \sqrt{-1}$, $\omega = 2\pi f$ is the angular frequency, τ_m is the delay in time from the center of the array to microphone m, and $\alpha_{P,n} = \cos(\theta_{P,n})$ where $\theta_{P,n}$ is the nth angle of the $P+1$ constraint angles.

We consider differential beamforming [1] for this study with the main lobe at angle $\theta_{P,0} = 0^0$(endfire direcction), and the desired signal propagates from the same direction. With the conventional signal model, the observed signal vector (of length M) is

$$\begin{aligned} \mathbf{y}(\omega) &= \begin{bmatrix} Y_1(\omega) & Y_2(\omega) & \dots & Y_M(\omega) \end{bmatrix} \\ &= \mathbf{d}^H(\omega, \alpha_{P,0})\, \mathbf{X}(\omega) + \mathbf{v}(\omega) \end{aligned} \tag{3}$$

where $Y_1(\omega)$ is the mth microphone signal, $\mathbf{X}(\omega)$ is the desired signal, $\mathbf{d}^H(\omega, \alpha_{P,0})$ is the steering vector at $\theta_{P,0} = 0^0$ (direction of source) and $\mathbf{v}(\omega)$ is the additive noise vector. The objective of beamforming is to estimate the desired signal $\mathbf{X}(\omega)$, from the observed signal $\mathbf{y}(\omega)$. To achieve this, a complex linear filter (of length M), $\mathbf{h}(\omega)$, is applied to the observed

signal vector, $\mathbf{y}(\omega)$ to achieve beamformer output

$$\begin{aligned}\mathbf{Z}(\omega) &= \sum_{m=1}^{M} H_m^*(\omega) Y_m(\omega) \\ &= \mathbf{h}^H(\omega)\mathbf{y}(\omega) \\ &= \mathbf{h}^H \mathbf{d}(\omega, \alpha_{P,0})\mathbf{X}(\omega) + \mathbf{h}^H(\omega)\mathbf{v}(\omega)\end{aligned} \tag{4}$$

where $Z(\omega)$ is an estimate of the desired signal, $X(\omega)$, and the superscript ^{H} is the conjugate-transpose operator. In our context, the array gain is expected to be 1 in the look direction, i.e., we should have

$$\mathbf{h}^H(\omega)\mathbf{d}(\omega) = 1. \tag{5}$$

2. PERFORMANCE MEASURES

Some important measures are presented in this section. Taking the first microphone as reference, the input SNR is defined as

$$i\mathrm{SNR}(\omega) = \frac{\phi_X(\omega)}{\phi_{V_1}(\omega)}, \tag{6}$$

where $\phi_X(\omega) = E\left[|X(\omega)|^2\right]$ and $\phi_{v_1}(\omega) = E\left[|V_1(\omega)|^2\right]$ are the variances of $X(\omega)$ and $V_1(\omega)$,respectively, with $E\left[.\right]$ denoting expectation. The output SNR is defined as

$$o\mathrm{SNR}[\mathbf{h}(\omega)] = \frac{\phi_X(\omega)}{\phi_{V_1}(\omega)} \frac{|\mathbf{h}^H(\omega)\mathbf{d}(\omega, \alpha_{P,0})|^2}{\mathbf{h}^H(\omega)\mathbf{\Gamma}_{\mathbf{v}}(\omega)\mathbf{h}(\omega)}, \tag{7}$$

where $\mathbf{\Gamma}_{\mathbf{v}}(\omega) = \frac{E\left[\mathbf{v}(\omega)\mathbf{v}^H(\omega)\right]}{\phi_{V_1}(\omega)}$ is the pseudo-coherence matrix of $\mathbf{v}(\omega)$. Therefore, the gain in SNR is:

$$\mathcal{G}[\mathbf{h}(\omega)] = \frac{o\mathrm{SNR}[\mathbf{h}(\omega)]}{i\mathrm{SNR}(\omega)} = \frac{|\mathbf{h}^H(\omega)\mathbf{d}(\omega, \alpha_{P,0})|^2}{\mathbf{h}^H(\omega)\mathbf{\Gamma}_{\mathbf{v}}(\omega)\mathbf{h}(\omega)}. \tag{8}$$

The array sensitivity to sensor and electronics noise is evaluated using the so-called white noise gain (WNG). This is derived by replacing $\mathbf{\Gamma_v}(\omega)$ with $\mathbf{I}_M$ ($M \times M$ identity matrix).

$$\mathcal{W}[\mathbf{h}(\omega)] = \frac{|\mathbf{h}^H(\omega)\mathbf{d}(\omega, \alpha_{P,0})|^2}{\mathbf{h}^H(\omega)\mathbf{h}(\omega)}. \tag{9}$$

Another important measure is the performance of the beamformer in presence of reverberation known as directivity factor (DF). Considering spherically isotropic (diffuse) noise field, the DF is:

$$\mathcal{D}[\mathbf{h}(\omega)] = \frac{|\mathbf{h}^H(\omega)\mathbf{d}(\omega, \alpha_{P,0})|^2}{\mathbf{h}^H(\omega)\mathbf{\Gamma}_{0,\pi}(\omega)\mathbf{h}(\omega)}, \tag{10}$$

where $\mathbf{\Gamma}_{0,\pi}(\omega) = \frac{1}{2}\int_0^\pi \mathbf{d}(\omega, \cos\theta)\mathbf{d}^H(\omega, \cos\theta)\sin\theta d\theta$. The vector $\mathbf{d}(\omega, \cos\theta)$ is found by replacing $\alpha_{P,n}$ with $\cos\theta$ in equation (1).

$$[\mathbf{\Gamma}_{0,\pi}(\omega)]_{ij} = \frac{\sin[\omega(j-i)\tau_{ij}]}{\omega(j-i)\tau_{ij}} = \mathrm{sinc}[\omega(j-i)\tau_{ij}], \tag{11}$$

where τ_{ij} is the delay between i^{th} and j^{th} sensor and $[\Gamma_{0,\pi}(\omega)]_{mm} = 1, m = 1, 2, \ldots, M$.

3. CONVENTIONAL BEAMFORMERS

In this section, we recall three important fixed beamformers: delay-sum, superdirective and robust superdirective. We also discuss a tunable beamformer [2] for superdirective beamforming.

The maximum WNG beamformer known as delay-sum beamformer is derived by maximizing (9) subject to distortionless constraint (5). We get

$$\mathbf{h}_{DS}(\omega) = \frac{\mathbf{d}(\omega, \alpha_{P,0})}{\mathbf{d}^H(\omega, \alpha_{P,0})\mathbf{d}(\omega, \alpha_{P,0})} = \frac{\mathbf{d}(\omega, \alpha_{P,0})}{M}. \tag{12}$$

The second type of beamformer maximizes the DF (10) known as superdirective beamformer [6] with a hypercardioid beampattern of order M-1 is give by:

$$\mathbf{h}_{SD}(\omega) = \frac{\mathbf{\Gamma}_{\mathbf{0,\beta}}^{-\mathbf{1}}\mathbf{d}(\omega)}{\mathbf{d}^H(\omega)\mathbf{\Gamma}_{\mathbf{0,\beta}}^{-\mathbf{1}}(\omega)\mathbf{d}(\omega)}. \tag{13}$$

The third beamformer is a regularized version of (13). The robust superdirective beamformer [7, 6] is obtained by maximizing DF subject to a constraint on WNG:

$$\mathbf{h}_{\mathrm{R},\epsilon}(\omega) = \frac{[\epsilon\mathbf{I}_M + \mathbf{\Gamma}_{0,\pi}(\omega)]^{-1}\mathbf{d}(\omega)}{\mathbf{d}^H(\omega)[\epsilon\mathbf{I}_M + \mathbf{\Gamma}_{0,\pi}(\omega)]^{-1}\mathbf{d}(\omega)} \tag{14}$$

where $\epsilon \geq 0$ is the Lagrange multiplier. It trades-off the super gain to white noise amplification. i.e., a small ϵ is used when large DF and a low WNG is desired and larger value when small DF and large WNG is desired. Although ϵ trades-off the WNG to DF, it does not have a closed form solution.

To overcome this a tunable beamformer was proposed [2]. Instead of Maximizing the DF subject to WNG ≥ 1, the noise amplification was minimized. The tunable beamformer is as follows:

$$\mathbf{h}_{\mathrm{T},\psi}(\omega) = \frac{[\epsilon_\psi\mathbf{I}_M + \mathbf{\Gamma}_{\psi,\pi}(\omega)]^{-1}\mathbf{d}(\omega)}{\mathbf{d}^H(\omega)[\epsilon_\psi\mathbf{I}_M + \mathbf{\Gamma}_{\psi,\pi}(\omega)]^{-1}\mathbf{d}(\omega)}, \tag{15}$$

where

$$[\mathbf{\Gamma}_{\psi,\pi}(\omega)]_{ij} = \frac{e^{j\omega(j-i)\tau_0\cos\psi} - e^{-j\omega(j-i)\tau_0}}{2j\omega(j-i)\tau_0}, \tag{16}$$

with

$$[\mathbf{\Gamma}_{\psi,\pi}(\omega)]_{mm} = \frac{1+\cos\psi}{2}, m = 1, 2, \ldots, M. \tag{17}$$

4. PROPOSED OPTIMIZATION

The block diagram of the optimization technique is shown in Figure. 1. We first start with random geometry for each particle i.e. the microphone position for each particle are initialized such that all of them are within the prescribed physical dimensions set by the

user. The coefficients are found using equation (50). The WNG and DF is evaluated for each particle and plugged into the fitness function. The PSO algorithm thereafter evaluates the best fit for each iteration and updates the geometry and regularization parameter when another global best particle is found.

The design approach for the weights for a given microphone geometry in a given frequency bin is to minimize a tradeoff in the environmental isotropic noise and the white noise of the microphones and electronics of the array itself while constraining the solution to the differential beamforming constraint [8]. Thus, the problem to be solved in each subband, ω is stated as,

$$\min_{\mathbf{h}(\omega)} \quad \left[\mathbf{h}^H(\omega)\mathbf{\Gamma}_{0,\pi}(\omega)\mathbf{h}(\omega) + \delta_w \mathbf{h}^H(\omega)\mathbf{h}(\omega)\right] \qquad (18)$$
$$\text{subject to } \mathbf{D}(\omega)\mathbf{h}(\omega) = \boldsymbol{\beta},$$

where $\mathbf{h}(\omega)$ is the coefficient vector in subband ω, $\mathbf{\Gamma_{0,\beta}}(\omega)$ is the isotropic noise covariance matrix of subband ω. The first term in the minimization in (18) represents the energy due to isotropic noise, the second term represents the energy due to the white noise in the microphones and electronics, and δ_w is the trade-off parameter between the two.

In the constraint equation of (18), the constraint matrix, $\mathbf{D}(\omega)$ is constructed using (1),

$$\mathbf{D}(\mathbf{!}) = \begin{pmatrix} \mathbf{d}^H(\omega, \alpha_{P,0}) \\ \mathbf{d}^H(\omega, \alpha_{P,1}) \\ \vdots \\ \mathbf{d}^H(\omega, \alpha_{P,P}) \end{pmatrix}, \qquad (19)$$

The constraint vector $\boldsymbol{\beta}$ is,

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 & \dots & \beta_P \end{pmatrix}^T, \qquad (20)$$

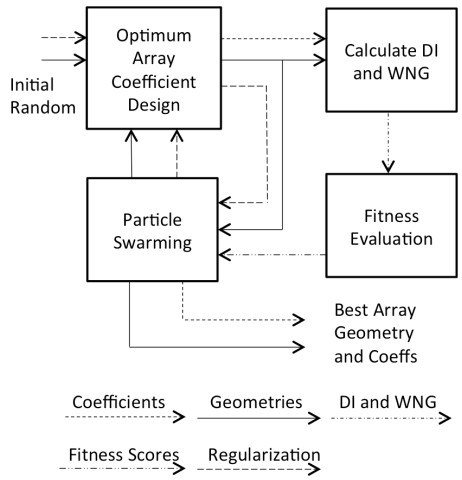


Figure 1. Block diagram of proposed geometry and weights optimization.

usually consists of $1's$ and $0's$. In our case, the first term is 1 and rest all are zero. Using the method of Lagrange multipliers in eq. (18), we define the cost function

$$J = \mathbf{h}^H \mathbf{\Gamma} \mathbf{h} + \delta_w \mathbf{h}^H \mathbf{h} + \boldsymbol{\lambda}^H (\boldsymbol{\beta} - \mathbf{D}\mathbf{h}) + (\boldsymbol{\beta} - \mathbf{D}\mathbf{h})^H \boldsymbol{\lambda} \tag{21}$$

Setting the gradient of J to the all-zero vector and solving the Lagrangian for $\mathbf{h}$, we obtain,

$$\mathbf{h} = \mathbf{\Gamma}_w^{-1} \mathbf{D}^H \left[\mathbf{D} \mathbf{\Gamma}_w^{-1} \mathbf{D}^H + \delta_d \mathbf{I} \right]^{-1} \boldsymbol{\beta}. \tag{22}$$

where $\mathbf{\Gamma}_w = [\mathbf{\Gamma} + \delta_w \mathbf{I}]$.

In our experiments, we use PSO to find the best microphone geometry,

$$\boldsymbol{\tau} = \begin{pmatrix} \tau_0 & \dots & \tau_{M-1} \end{pmatrix}^T, \tag{23}$$

and regularization values, δ_w and δ_d as each particle represents a particular set of microphone array geometry and regularization values, $\boldsymbol{\phi}$, δ_w and δ_d . The algorithm attempts to steer the particles in such a manner as to optimize a *fitness function*, F. In our case this

function includes a non-linear cost function on the weighted directivity factor $\mathcal{D}\left(\mathbf{h}(\omega)\right)$ and the weighted white noise gain, $\mathcal{W}\left(\mathbf{h}(\omega)\right)$. Specifically,

$$\begin{aligned} F &= \sum_{\omega} W_{\mathcal{D}}(\omega)|f\left(\mathcal{D}_{tgt}(\omega) - \mathcal{D}\left(\mathbf{h}(\omega)\right)\right)|^2 \\ &\quad + W_{\mathcal{G}}(\omega)|f\left(\mathcal{W}_{tgt}(\omega) - \mathcal{W}\left(\mathbf{h}(\omega)\right)\right)|^2 \end{aligned} \tag{24}$$

where the target DF, $\mathcal{D}_{tgt}(\omega)$ and the target WNG, $\mathcal{W}_{tgt}(\omega)$ are both functions of frequency, the function $f(x)$ is defined as

$$f(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases}, \tag{25}$$

is designed to not penalize DIs and WNGs that are above the target values and $W_{\mathcal{D}}$ and $W_{\mathcal{G}}$ are the weights assigned to resultant DF and WNG squared error for frequency bins where more performance is desired (for example at lower frequencies for DMAs have high white noise noise amplification at these frequencies). For frequencies where performance is generally good, the weights can ignored i.e. assigned zero.

5. SIMULATION RESULTS

We consider a swarm of 20 particles, where each particle represents a candidate solution to a geometry. The microphones in each particle's geometry move around in the constrained search space of specified distance from the origin. We consider three different situations to compare the proposed array design with various number of microphones and constraints on the array size. We consider a minimum inter-microphone distance of one cm.

First, we consider maximum directivity beamformer with the array length of 64 cm and consisting 21 microphones with desired DF and WNG 20dB and -20dB respectively. The resulting optimized array is shown in Figure. 2(a). The arrays DF and WNG are in Figure. 2(a)-(b). We compare the performance of proposed approach with DS eq.

(12), superdirective eq. (13), robust superdirective $\epsilon = 10^{-4}$ eq. (14) and tunable robust superdirective eq. (15). We choose $\psi = 0.1^0$ for tunable beamformer for it yields maximum directivity. The microphones in beamformers eq. (12)-(15) are uniformly distributed across total length of 64cm. The DF of other arrays is high around 20-25dB only upto 3kHz and falls to around 13dB from 3-8kHz. One possibility for this could be the interelement spacing which should not exceed half the longest wavelength ($\sim$ 2.1cm). However, it is $\sim$ 3cm in case of a 21 element array of length 64cm is. Comparatively, the proposed approach has flat response except for at very low frequencies mitigating the effects of spatial aliasing.

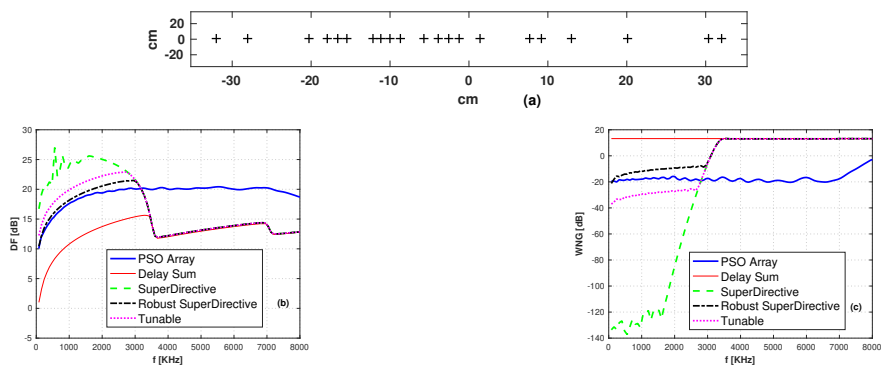


Figure 2. A comparison of PSO array and conventional techniques [19] with 21 elements and length of 64cm: (a) PSO array geometry with 21 elements (b) DF (c) WNG. Array gains for proposed array(solid line), DS(line), superdirective(dashed line), robust superdirective $\epsilon = 10^{-4}$(dash-dot line), and tunable $\psi = 0.1^0$(dotted line).

Next, we consider a nested Figure. 3(b) and a PSO Figure. 3(a) array of length 64cm consisting 13 microphones with desired DF and WNG 20dB and -20dB respectively. The corresponding DF and WNG are shown in Figure. 3(c) and (d) respectively. Under these conditions, we would like to point out two observations. First observation is that both the arrays have same DF and WNG performance. However, only this time the PSO array is smaller in size $\sim$40cm. In nested arrays, it is typical to have $5 + 2S$ number of microphones [9, 10], where S is the number of so-called sub-arrays. Second observation is the maximum length of the nested array. Consider the situation where there are more than $5 + 2S$ microphones available but the array length cannot be increased. What microphone

arrangement would yield the best results in terms of improved directivity over frequency? Under such circumstances, the proposed technique can be used to optimize the array geometry and parameters with more number of microphones and keeping the array size same 64cm. We use the same PSO geometry in Figure. 2(a) and the resulting DF and WNG for comparisons. The resultant directivity is ~ 4dB higher with array length 64cm.

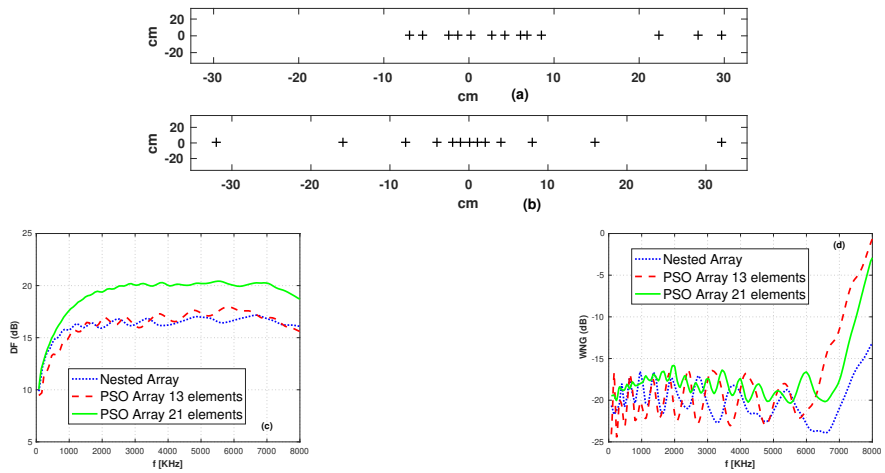


Figure 3. Array geometry and gains for maximum size 64cm (a) PSO array geometry with 13 elements (b) Nested array geometry with 13 elements (c) DF (d) WNG. Array gains for PSO array with 21 elements(solid line), PSO array with 13 elements(dashed line), and nested array(dotted line).

Next, we set the desired DF and WNG to Maximum level 20dB. This time the Proposed approach aims to achieve both the DI and WNG as maximum as possible with array size limited to 40cm and 10 microphones. The PSO array geometry is given in Figure. 4(a). It can be seen that the overall length of the array is ~34cm. The corresponding PSO array and reference DFs and WNGs are in Figure. 4(b) and 4(c) respectively. The response for the stochastic optimized array proposed in [19] are of two categories. Frequency invariant and Max directivity response. The description of algorithm in [19] is beyond the scope of this paper and the reader can refer to for further details. We will use them for comparison with the aperiodic array optimized in this study. In case of frequency invariant response, the DF and WNG responses are flat across the given frequency band 0-12 kHz. In case of

Maximum Directivity beamformer, the DF is as maximum as achievable. But at the cost of deteriorating WNG at low frequencies. The proposed approach using PSO mitigates the above mentioned problem and optimizes the trade-off between DI and WNG i.e. maximizes them both. Also the array aperture is 15% smaller compared to stochastic optimized array. The PSO array DF and WNG is ~10dB and ~9dB respectively and frequency invariant.

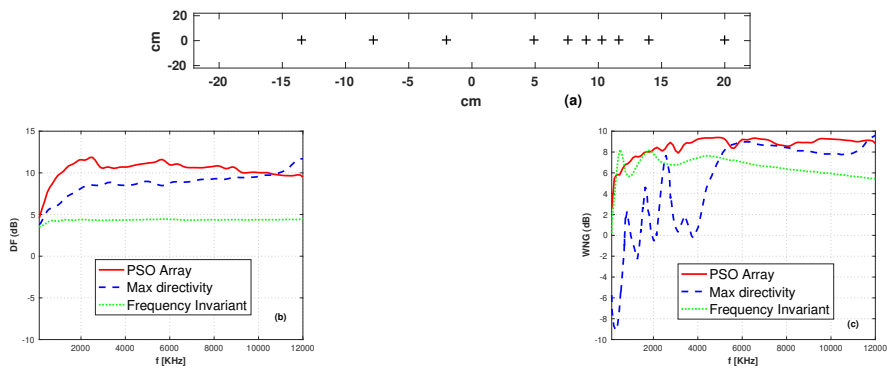


Figure 4. A comparison of PSO array and Sparse aperiodic array [19] with 10 elements and length of 40cm: (a) PSO array geometry (b) DF (c) WNG. Proposed array gains(solid line), max directivity(dashed line) [19], and frequency invariant(dotted line) [19].

6. CONCLUSION

The proposed simultaneous optimization of microphone array geometry ($\boldsymbol{\tau}$ vector) and regularization parameter δ_w using PSO improves the beamformer SNR gains. In comparison to conventional techniques, the proposed approach finds an optimal solution with smaller array aperture for given SNR gains or by adding more microphones while limiting its physical size by giving an overall aperture reduction of 40% in comparison to nested arrays in case higher directivity is desired. Even when the the beamformer was designed to operate equally well under both the noise (white and diffuse) conditions, it's DF and WNG were together higher and 15% smaller aperture compared to stochastic and analytic optimized arrays.

REFERENCES

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

[2] Berkun, R., Cohen, I., & Benesty, J. (2016, September). A tunable beamformer for robust superdirective beamforming. In 2016 IEEE International Workshop on Acoustic Signal Enhancement (IWAENC) (pp. 1-5). IEEE.

[3] Cox, H., Zeskind, R., & Kooij, T. (1986). Practical supergain. IEEE Transactions on Acoustics, Speech, and Signal Processing, 34(3), 393-398.

[4] Cox, H., Zeskind, R., & Owen, M. (1987). Robust adaptive beamforming. IEEE Transactions on Acoustics, Speech, and Signal Processing, 35(10), 1365-1376.

[5] Benesty, J., & Jingdong, C. (2012). Study and design of differential microphone arrays (Vol. 6). Springer Science & Business Media.

[6] Crocco, M., & Trucco, A. (2012). Stochastic and analytic optimization of sparse aperiodic arrays and broadband beamformers with robust superdirective patterns. IEEE Transactions on audio, speech, and language processing, 20(9), 2433-2447.

[7] McCowan, I. (2001). Microphone arrays: A tutorial. Queensland University, Australia, 1-38.

[8] Pal, P., & Vaidyanathan, P. P. (2010). Nested arrays: A novel approach to array processing with enhanced degrees of freedom. IEEE Transactions on Signal Processing, 58(8), 4167-4181.

II. ON THE DESIGN OF OPTIMAL 2D MICROPHONE ARRAY GEOMETRIES

Sohel J. Patel, Steven L. Grant, Maciej Zawodniok
Department of Electrical & Computer Engineering
Missouri University of Science and Technology
Rolla, Missouri 65409
Email: sjp432@mst.edu, sgrant@mst.edu, mjzx9c@mst.edu

ABSTRACT

Microphone arrays are used in wide range of applications such as sound capture and enhancement. Among many factors affecting their performance, geometry is rarely discussed in the literature. There are many techniques available for deriving optimal coefficients for microphone arrays of a given geometry and given criteria such as look-direction and null-steering. However, there are few algorithms available for the simultaneous design of optimal microphone array geometries and coefficients. This paper addresses that problem by optimizing not only the coefficients but geometry as well for non-uniform planar arrays using particle swarm optimization (PSO). The proposed approach achieves desired directivity factor (DF) or white noise gain (WNG) over wide range of frequencies for multiple look directions given the number of microphones and aperture. The resultant optimized planar array geometry achieves steerable, robust, superdirective and frequency invariant characteristics with smaller aperture compared to recently proposed methods in terms of DF, WNG.

1. INTRODUCTION

Microphone arrays find their application in various fields such as sound capture and processing, teleconferenceing, human-to-machine interface, smart handsfree devices and speech enhancement [1, 2, 3, 4]. Beamforming is a common technique to achieve

signal enhancement and noise suppression by applying complex weights to microphones [5]. Conventionally, the beamformers weights are computed for a specific geometry of uniformly spaced microphones. However, depending on the type of application it may require us to design non uniform geometries yet achieve desired DF and WNG.

Classical beamforming techniques include delay-sum beamforming also known as maximum WNG beamformer but suffers form poor directivity. The superdirective beamformer achieves supergain [1, 6] i.e. highest achievable DF but suffers from high white noise amplification. In order to address the regularized superdirective beamformer was proposed that tradesoff the DF to WNG depending on the regularization parameter [1, 7]. However, the tradeoff parameter is difficult if not impossible to optimize to achieve frequency invariant performance. In our study, the frequency invariant characteristics are achieved with respect to DF and WNG. We leave the geometry optimization to achieve frequency invariant beampattern for future work.

Algorithms that achieve frequency invariance performance include differential beamforming [8], nested arrays [9, 10], modal beamforming [11]. Such characteristics are desirable for various microphone arrays signal processing algorithms such as beamforming [1, 2, 4], multichannel denoising [12] and dereverberation [13, 14]. However, their performance given non-uniform geometries is rarely studied.

Earlier geometry optimization algorithms such as [15, 16] have demonstrated their superiority over uniform geometries. Array geometry optimization is achieved using parameters like maximum sidelobe level (MSL) [17] and centroid and dispersion [18]. However, the broadband geometry optimization with Frequency invariance characteristics results in a non-convex cost function which can be extremely difficult to solve if not impossible. Few techniques addressing the above non linear optimization of linear microphone arrays include stochastic and analytic optimization [19], simulated annealing [20] and particle swarm optimization [21]. Algorithms for optimizing planar arrays mainly include genetic algorithm [22, 23, 24, 25, 26, 27, 28, 29].

The differential beamforming techniques [8] usually work well for small microphone arrays. Their performance significantly reduces when the array aperture increases specifically when the microphone placements are sparse. To address this issue, a sparse array optimization technique using compressive sensing was proposed [30]. Other techniques such as [31, 32] optimize the microphone coefficients to obtain frequency invariant beampatterns. However, these techniques require that the array geometry have some kind of symmetry. Moreover, there is no control over the desired DF or WNG. We address these issues in our study to achieve desired frequency invariant DF and WNG for some arbitrary planar array geometries.

The most recent technique [33] employs genetic algorithm to optimize the planar array at f=1kHz and only three look directions. We use this technique as the baseline and compare the proposed approach. The proposed approach uses PSO to optimize the array geometry such that the so called fitness function is minimized. The proposed approach can achieve wideband (0-8) kHz robust, superdirective or frequency invariant response for any given look direction. The frequency invariance is not achieved for beampatterns and only for DF and WNG. The fitness function also gives more flexibility to achieve desired characteristics for multiple look angles across full band of frequencies.

2. SIGNAL MODEL AND PROBLEM FORMULATION

In this study, we consider a source signal propagating in ideal conditions i.e. anechoic environment. The source signal propagates at the speed of sound, $c = 34000\, cm/s$ and is a plane wave. The plane wave impinge on the array at an angle θ measured anti-clockwise form x axis. We consider the center of the sensor array at $(0, 0)$ of Cartesian co-ordinate system. The coordinates for m^{th} microphone is given by $\mathbf{r}_m = [x_m \quad y_m]$. The delay from the microphone to the center of the array for a signal impinging at θ_p degrees in vectorized form can be expressed as

$$\tau_m = \frac{1}{c}\mathbf{r}_m\mathbf{u}_p \tag{1}$$

where $\mathbf{u_p} = \begin{bmatrix} \cos\theta_p & \sin\theta_p \end{bmatrix}^T$, $p = 0, 1, \ldots, P$, $\theta_p = 2\pi(p-1)/P$ and P being the number of constraints.

The constraint angles vector $\boldsymbol{\theta}$ of length P is defined as,

$$\boldsymbol{\theta} = [\theta_1 \ \ldots \ \theta_p \ \ldots \ \theta_P] \tag{2}$$

The steering vector of length M for p^{th} constraint angle is given by,

$$\mathbf{d}(\omega, \theta_p) = [e^{j\omega\tau_1} \ \ e^{j\omega\tau_2} \ \ldots \ e^{j\omega\tau_m} \ \ldots \ e^{j\omega\tau_M}]^T \tag{3}$$

where the subscript ^{T} is the transpose operator, j is imaginary number with $j = \sqrt{-1}$, $\omega = 2\pi f$ is the angular frequency, $f > 0$ is the temporal frequency and M is the number of microphones.

The signal received at m^{th} microphone in frequency domain is

$$\mathbf{Y}_m(\omega) = e^{j\omega\tau_m}\mathbf{X}(\omega) + \mathbf{V}_m(\omega) \tag{4}$$

where $\mathbf{X}(\omega)$ is the desired signal and $\mathbf{V}_m(\omega)$ is the additive noise. The observation signal vector $\mathbf{y}$ of length M for given angle of incidence $\theta_s \in \boldsymbol{\theta}$ is

$$\mathbf{y}(\omega) = \ \mathbf{d}(\omega, \theta_s)\ \mathbf{X}(\omega) + \mathbf{v}(\omega) \tag{5}$$

The spatial filter of length M is

$$\mathbf{h}(\omega) = \begin{bmatrix} h_1(\omega) & h_2(\omega) & \ldots & h_M(\omega) \end{bmatrix}^T \tag{6}$$

The array gain of the look direction is generally unity with a constraint define by

$$\mathbf{h}^H(\omega)\mathbf{d}(\omega, \theta_s) = 1 \tag{7}$$

with θ_s being the look direction. The estimate of the desired signal is achieved by beamforming. This involves applying a complex weight $H_m^*(\omega)$ to the microphone's output $Y_m(\omega)$ and summing them all. The subscript * denotes the complex conjugate

$$\mathbf{Z}(\omega) = \mathbf{h}^H(\omega)\mathbf{y}(\omega) \tag{8}$$

3. PERFORMANCE MEASURES

This section briefly describes the important performance measures. The beampattern is a measure of sensitivity of the beamformer to a plane wave impinging on the array from the direction θ_s

$$\begin{aligned} \mathcal{B}[\mathbf{h}(\omega), \theta] &= \mathbf{h}^H(\omega)\mathbf{d}(\omega, \theta_s) \\ &= \sum_{m=1}^{M} H_m^*(\omega) e^{j\omega\tau_m} \end{aligned} \tag{9}$$

The beamformer performance is evaluated using signal-to-noise ratio (SNR) gain at the reference microphone $m = 1$ [8] i.e. first microphone

$$i\mathrm{SNR}(\omega) = \frac{\phi_X(\omega)}{\phi_{V_1}(\omega)} \tag{10}$$

where $\phi_X(\omega) = E[|X(\omega)|^2]$ and $\phi_V(\omega) = E[|V_1(\omega)|^2]$ are the variances of $X(\omega)$ and V_1, respectively and $E[.]$ is the expectation operator. The output SNR is given by

$$\begin{aligned} o\mathrm{SNR}[\mathbf{h}(\omega)] &= \phi_X(\omega) \frac{|\mathbf{h}^H(\omega)\mathbf{d}(\omega, \theta_s)|^2}{\mathbf{h}^H(\omega)\mathbf{\Phi}_{\mathbf{v}}(\omega)\mathbf{h}(\omega)} \\ &= \frac{\phi_X(\omega)}{\phi_{V_1}(\omega)} \frac{|\mathbf{h}^H(\omega)\mathbf{d}(\omega, \theta_s)|^2}{\mathbf{h}^H(\omega)\mathbf{\Gamma}_{\mathbf{v}}(\omega)\mathbf{h}(\omega)} \end{aligned} \tag{11}$$

where $\mathbf{\Phi_v}(\omega) = E[\mathbf{v}(\omega)\mathbf{v}^H(\omega)]$ and $\mathbf{\Gamma_v}(\omega) = \mathbf{\Phi_v}/\phi_V(\omega)$ are the correlation and pseudo-coherence matrices [8, 1] of $\mathbf{v}(\omega)$, respectively. Therefore, the gain in SNR is

$$\begin{aligned}\mathcal{G}[\mathbf{h}(\omega)] &= \frac{o\mathrm{SNR}[\mathbf{h}(\omega)]}{i\mathrm{SNR}(\omega)} \\ &= \frac{|\mathbf{h}^H(\omega)\mathbf{d}(\omega,\theta_s)|^2}{\mathbf{h}^H(\omega)\mathbf{\Gamma_v}(\omega)\mathbf{h}(\omega)}\end{aligned} \tag{12}$$

The two types of noises significantly important for the design of robust superdirective beamformers are

The temporally and spatially white noise with the same variance at all microphones. In this case, $\mathbf{\Gamma_v}(\omega) = \mathbf{I}_M$, where $\mathbf{I}_M$ is the $M \times M$ identity matrix. Therefore, the WNG is defined as

$$\mathcal{W}[\mathbf{h}(\omega)] = \frac{|\mathbf{h}^H(\omega)\mathbf{d}(\omega,\theta_s)|^2}{\mathbf{h}^H(\omega)\mathbf{h}(\omega)} \tag{13}$$

Using the Cauchy-Schwarz inequality, i.e.,

$$\begin{aligned}|\mathbf{h}^H(\omega)\mathbf{d}(\omega,\theta_s)|^2 &\leq \mathbf{h}^H(\omega)\mathbf{h}(\omega) \\ &\quad \times \mathbf{d}(\omega,\theta_s)^H\mathbf{d}(\omega,\theta_s),\end{aligned} \tag{14}$$

it can be easily deduced that

$$\mathcal{W}[\mathbf{h}(\omega)] \leq M. \tag{15}$$

Therefore the maximum WNG is

$$\mathcal{W}_{max} = M, \tag{16}$$

The DF is given by

$$\mathcal{D}[\mathbf{h}(\omega)] = \frac{|\mathbf{h}^H(\omega)\mathbf{d}(\omega,\theta_s)|^2}{\mathbf{h}^H(\omega)\mathbf{\Gamma}_{\mathrm{dn}}(\omega)\mathbf{h}(\omega)} \tag{17}$$

where

$$\mathbf{\Gamma}_{\mathrm{dn}}(\omega) = \frac{\sin[\omega(j-i)\tau_0]}{\omega(j-i)\tau_0} = sinc[\omega(j-i)\tau_0] \tag{18}$$

Using the Cauchy-Schwarz inequality again, i.e.,

$$\begin{aligned}|\mathbf{h}^H(\omega)\mathbf{d}(\omega,\theta_s)|^2 \leq\ & \mathbf{h}^H(\omega)\mathbf{\Gamma}_{\text{dn}}(\omega)\mathbf{h}(\omega) \\ & \times\mathbf{d}^H(\omega,\theta_i)\mathbf{\Gamma}_{\text{dn}}^{-1}(\omega)\mathbf{d}(\omega,\theta_s),\end{aligned} \tag{19}$$

we find that

$$\mathcal{D}[\mathbf{h}(\omega)] \leq \mathbf{d}^H(\omega,\theta_s)\mathbf{\Gamma}_{\text{dn}}^{-1}(\omega)\mathbf{d}(\omega,\theta_s) \tag{20}$$

Therefore, the maximum DF is given by

$$\begin{aligned}\mathcal{D}_{max}(\omega) &= \mathbf{d}^H(\omega,\theta_s)\mathbf{\Gamma}_{\text{dn}}^{-1}(\omega)\mathbf{d}(\omega,\theta_s) \\ &= \text{tr}[\mathbf{d}(\omega,\theta_s)^H\mathbf{\Gamma}_{\text{dn}}^{-1}(\omega)\mathbf{d}(\omega,\theta_s)] \\ &\leq M\text{tr}[\mathbf{\Gamma}_{\text{dn}}^{-1}(\omega)]\end{aligned} \tag{21}$$

where tr[.] denotes the trace of a square matrix. It can be noted that, unlike the maximum WNG, the maximum DF, also known as supergain, is frequency dependent, which can be achieved at the expense of white noise amplification.

4. CONVENTIONAL BEAMFORMERS

In this section, we recall the conventional fixed beamformers, one that maximize the WNG and the other that maximizes DF. Also, the regularized version of latter will be discussed.

The most commonly used beamformer is delay-and-sum (DS) which also known as maximum WNG beamformer. For it is derived by maximizing the WNG subject to distortionless constraint given in (7)

$$\mathbf{h}_{\text{DS}}(\omega,\theta_s) = \frac{\mathbf{d}(\omega,\theta_s)}{\mathbf{d}^H(\omega,\theta_s)\mathbf{d}(\omega,\theta_s)} = \frac{\mathbf{d}(\omega,\theta_s)}{M}. \tag{22}$$

The superdirective beamformer is derived by maximizing the DF subject to the distortionless constraint given in (7) with hypercardioid beampattern of order M.

$$\mathbf{h}_{\mathrm{S}}(\omega, \theta_s) = \frac{\mathbf{\Gamma}_{\mathrm{dn}}^{-1}(\omega)\mathbf{d}(\omega, \theta_s)}{\mathbf{d}^H(\omega, \theta_s)\mathbf{\Gamma}_{\mathrm{dn}}^{-1}(\omega)\mathbf{d}(\omega, \theta_s)} \tag{23}$$

The regularized superdirective beamformer is derived by maximizing the DF with a constraint on WNG [1, 7]. Using the distortionless constraint in (7), we obtain

$$\mathbf{h}_{\mathrm{R},\epsilon}(\omega, \theta_s) = \frac{[\mathbf{\Gamma}_{\mathrm{dn}}(\omega) + \epsilon\mathbf{I}_M]^{-1}\mathbf{d}(\omega, \theta_s)}{\mathbf{d}^H(\omega, \theta_s)[\mathbf{\Gamma}_{\mathrm{dn}}(\omega) + \epsilon\mathbf{I}_M]^{-1}\mathbf{d}(\omega, \theta_s)} \tag{24}$$

where ϵ is the regularization parameter that finds a good compromise between DF and WNG i.e., a small ϵ leads to large DF and low WNG, while a large ϵ gives low DF and large WNG. Hence, it can be shown that $\mathbf{h}_{\mathrm{R},0}(\omega, \theta_s) = \mathbf{h}_{\mathrm{S}}(\omega, \theta_s)$ and $\mathbf{h}_{\mathrm{R},\infty}(\omega, \theta_s) = \mathbf{h}_{\mathrm{DS}}(\omega, \theta_s)$. Traditionally, the regularization parameter ϵ is same across all subbands. We propose to make the DF to WNG tradeoff parameter ϵ adaptive for different frequencies for more flexibility in achieving desired DF and WNG with additional constraints as will be discussed in Section 4.

5. GRADIENT DESCENT OPTMIZATION

The conventional beamformers discussed in previous section have fixed weights and as such have fixed and rather non-optimal performance characteristics. This section introduces the problem formulation which includes minimizing the array output power subject to a beampattern design. The geometry and coefficients are computed by minimizing noise using gradient descent.

For P constraint angles, θ_p, $p = 1 \ldots P$ and each microphone element, $\mathbf{r}_m$ we express matrix D as,

$$\mathbf{D}(\omega, \theta_s) = \begin{bmatrix} \mathbf{d}^H(\omega, \theta_0) \\ \mathbf{d}^H(\omega, \theta_p) \\ \vdots \\ \mathbf{d}^H(\omega, \theta_P) \end{bmatrix} = \begin{bmatrix} \mathbf{d}_0^H \\ \mathbf{d}_p^H \\ \vdots \\ \mathbf{d}_P^H \end{bmatrix} \tag{25}$$

where

$$\mathbf{d}_p = \mathbf{d}(\omega, \theta_p) = e^{j\omega \frac{1}{c} \mathbf{M} \mathbf{u}_p^T} \tag{26}$$

and $\mathbf{M}$ is a matrix given by

$$\mathbf{M} = \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_1 & \dots & \mathbf{r}_m & \dots & \mathbf{r}_M \end{bmatrix}^T = \begin{bmatrix} \mathbf{x} & \mathbf{y} \end{bmatrix} \tag{27}$$

where $\mathbf{x}$ and $\mathbf{y}$ are the first and second columns of $\mathbf{M}$, respectively. Alternatively, we could express (26) as

$$\mathbf{d}_p = e^{j\omega \frac{1}{c} (\mathbf{x} \cos \theta_p + \mathbf{y} \sin \theta_p)} \tag{28}$$

In this problem both the weight $\mathbf{h}(\omega)$ and position vectors $\mathbf{M}$ should be determined by the minimization,

$$\begin{aligned} \underset{h(\omega), M}{\text{minimize}} \quad & \mathbf{h}^H(\omega) \mathbf{\Phi}_{\mathbf{v}}(\omega) \mathbf{h}(\omega) \\ \text{subject to} \quad & \mathbf{D}(\omega, \theta_s) \mathbf{h}(\omega) = \boldsymbol{\beta} \end{aligned}$$

However, this problem has no closed form solution, so we will take an alternating gradient descent approach. The minimum norm part of the minimization is just a common tap leakage which we shall address later. The standard gradient update is to first calculate an error,

$$\mathbf{e}_i = \boldsymbol{\beta}_i - \mathbf{d}^H(\omega, \theta) \mathbf{h}_{i-1}(\omega) \tag{29}$$

And then update the coefficients, with a gradient descent on holding constant

$$\mathbf{h}_i(\omega) = \mathbf{h}_{i-1}(\omega) - \mu \frac{dJ(e_i)}{d\mathbf{h}} \tag{30}$$

and then a gradient descent update on the delay vector, $\boldsymbol{M}$, holding $\boldsymbol{h}(\omega)$ constant

$$\mathbf{M}_i = \mathbf{M}_{i-1} - \mu \frac{dJ(e_i)}{d\mathbf{M}} \tag{31}$$

Where we will also apply a constraint after each iteration that $M_{i,0}$. In both cases the performance index is,

$$J(e_i) = |e_i|^2 = e_i e_i^* \tag{32}$$

We are interested in a 2-D gradient update on $\mathbf{M}_i$

$$\mathbf{M}_i = \mathbf{M}_{i-1} - \mu \begin{bmatrix} \frac{\partial J(e)}{\partial \mathbf{x}} & \frac{\partial J(e)}{\partial \mathbf{y}} \end{bmatrix} \tag{33}$$

$$\begin{aligned} \frac{\partial e_i}{\partial x_k} &= -\frac{\partial}{\partial x_k} \left\{ \mathbf{d}_i^H \mathbf{h}_{i-1} \right\} \\ &= -\frac{\partial}{\partial x_k} \left\{ \sum_{m=0}^{M-1} e^{j\omega \frac{1}{c}(\mathbf{x}_m \cos\theta_i + \mathbf{y}_m \sin\theta_i)} \right\} h_{m,i-1} \\ &= -j\frac{\omega}{c} \cos\theta_i e^{j\omega \frac{1}{c}(\mathbf{x}_k \cos\theta_i + \mathbf{y}_k \sin\theta_i)} h_{k,i-1} \end{aligned} \tag{34}$$

$$\begin{aligned} \frac{\partial e_i}{\partial y_k} &= -\frac{\partial}{\partial y_k} \left\{ \mathbf{d}_i^H \mathbf{h}_{i-1} \right\} \\ &= -\frac{\partial}{\partial y_k} \left\{ \sum_{m=0}^{M-1} e^{j\omega \frac{1}{c}(\mathbf{x}_m \cos\theta_i + \mathbf{y}_m \sin\theta_i)} \right\} h_{m,i-1} \\ &= -j\frac{\omega}{c} \sin\theta_i e^{j\omega \frac{1}{c}(\mathbf{x}_k \cos\theta_i + \mathbf{y}_k \sin\theta_i)} h_{k,i-1} \end{aligned} \tag{35}$$

In vector form,

$$\begin{aligned} \frac{\partial e_i}{\partial \mathbf{x}} &= -j\frac{\omega}{c} \cos\theta_i e^{j\omega \frac{1}{c}(\mathbf{x} \cos\theta_i + \mathbf{y} \sin\theta_i)} \odot \mathbf{h}_{i-1} \\ &= -j\frac{\omega}{c} \cos\theta_i \mathbf{d}_i^* \odot \mathbf{h}_{i-1} \end{aligned} \tag{36}$$

and

$$\begin{aligned} \frac{\partial e_i}{\partial \mathbf{y}} &= -j\frac{\omega}{c} \sin\theta_i e^{j\omega \frac{1}{c}(\mathbf{x} \cos\theta_i + \mathbf{y} \sin\theta_i)} \odot \mathbf{h}_{i-1} \\ &= -j\frac{\omega}{c} \sin\theta_i \mathbf{d}_i^* \odot \mathbf{h}_{i-1} \end{aligned} \tag{37}$$

So, the gradients of the cost function are,

$$\begin{aligned}&\begin{bmatrix}\frac{\partial J(e)}{\partial \mathbf{x}} & \frac{\partial J(e)}{\partial \mathbf{y}}\end{bmatrix}\\&= -2j\frac{\omega}{c}\begin{bmatrix}\cos\theta_i\mathbf{d}_i^* \odot \mathbf{h}_{i-1} & \cos\theta_i\mathbf{d}_i^* \odot \mathbf{h}_{i-1}\end{bmatrix} e_i^*\end{aligned} \tag{38}$$

Thus, equation (33) becomes,

$$\mathbf{M}_i = \mathbf{M}_{i-1} - 2\mu j\frac{\omega}{c}\begin{bmatrix}\cos\theta_i\mathbf{d}_i^* \odot \mathbf{h}_{i-1} & \cos\theta_i\mathbf{d}_i^* \odot \mathbf{h}_{i-1}\end{bmatrix} e_i^* \tag{39}$$

Defining,

$$\mathbf{B} = \begin{bmatrix}\cos\theta_i\mathbf{d}_i^* \odot \mathbf{h}_{i-1} & \sin\theta_i\mathbf{d}_i^* \odot \mathbf{h}_{i-1}\end{bmatrix} \tag{40}$$

$$\mathbf{M}_i = \mathbf{M}_{i-1} - 2\mu j\frac{\omega}{c}\mathbf{B}\left(\mathbf{B}^T\mathbf{B} + \delta\mathbf{I}\right)^{-1} e_i^* \tag{41}$$

Taking the above equations, the geometry optimization using alternative gradient descent is summarized as:

for $iter \leftarrow 1$ **to** $iterations$ **do**
  for $i \leftarrow 1$ **to** P **do**
    $\boldsymbol{d}_i = e^{-j\omega\frac{1}{c}\mathbf{M}_k\mathbf{v}_i}$
    $e_i = \beta_i - \boldsymbol{d}_i^H\boldsymbol{h}_k$
    $\mathbf{h}_i = \gamma\mathbf{h}_k + \mu_h\mathbf{d}_i e_i^*/(\mathbf{d}_i^T\mathbf{d}_i + \delta)$
    $e_i = \beta_i - \mathbf{d}_i^H\mathbf{h}_i$
    $\mathbf{B} = [\cos\theta_i\mathbf{d}_i^* \circ \mathbf{h}_k, \sin\theta_i\mathbf{d}_i^* \circ \mathbf{h}_k]$
    $\mathbf{M}_i = \mathbf{M}_k - 2\mu j\frac{\omega}{c}\mathbf{B}_i(\mathbf{B}_i^T\mathbf{B}_i + \delta\mathbf{I})^{-1}e_i^*$
    $imag(\mathbf{M}_i) = \gamma\, imag(\mathbf{M}_i)$
    $k = i$
  end
end

Algorithm 1: Algorithm for planar arrays optimized using gradient descent

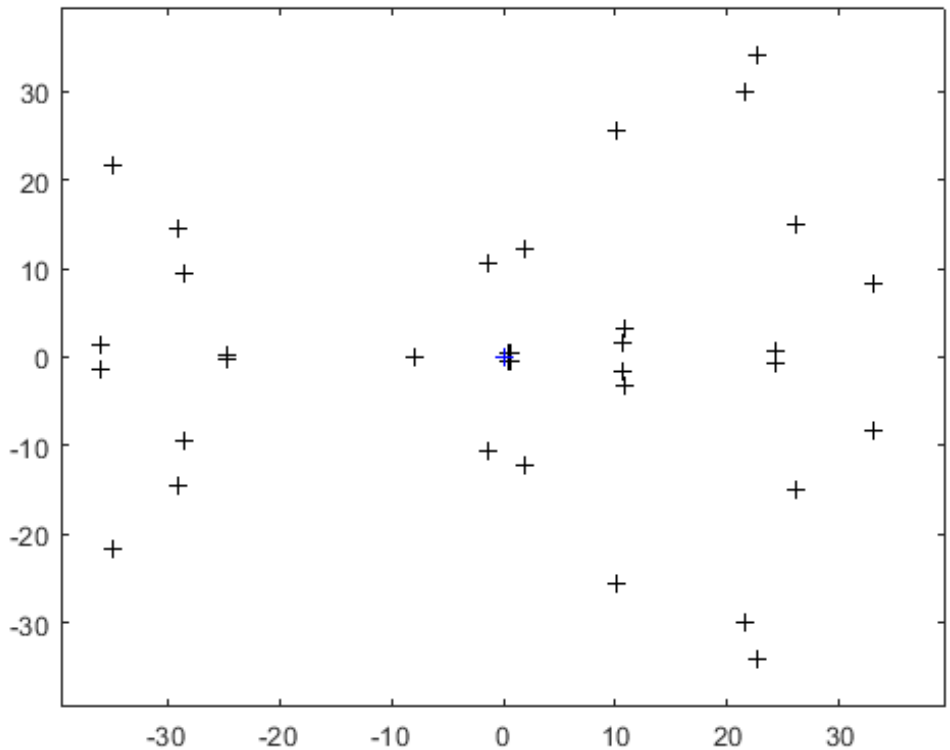
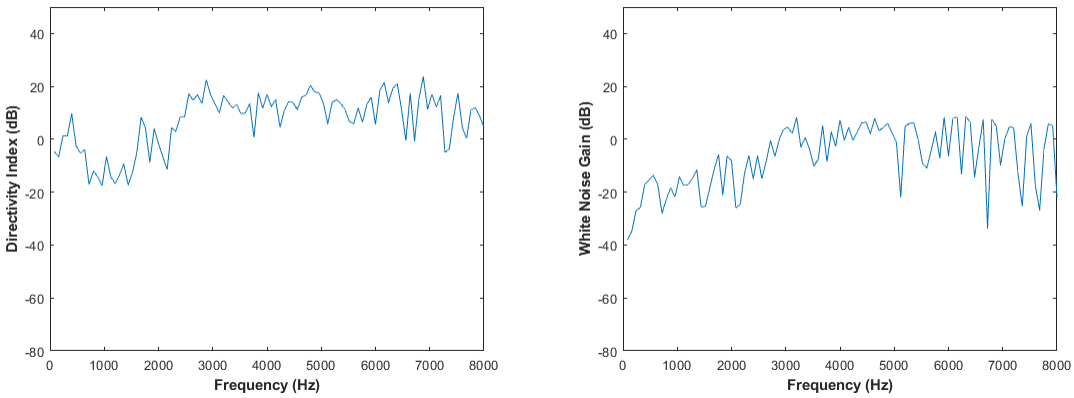


Figure 1. Optimized adaptive planar array geometry with 36 microphones.

So far, we discussed weights and geometry optimization of linear and planar arrays. However, when there are more characteristic desired, the above mentioned methods and geometries have limitations. For example, consider figure(1-3). A planar array is optimized subject to multiple nulls. For 36 microphones, theoretically there may be upto 35 nulls. It can be realized from Figure(3) that the nulls are present at most of the frequencies except for very low frequencies. Which is quite evident form figure(2)(A) plots as well where the directivity index is poor for frequencies below 3000Hz. Besides there is no way to set desired values for these characteristics.

However the beamformer performance is not consistent with respect to frequencies. Hence, frequency invariant responses are desired when optimizing geometry and weights. Also, the above mentioned results are only for one look direction i.e. 0^0. Hence, we



(a) (b)

Figure 2. Performance characteristics of adaptive planar array, (a) DI and (b) WNG.

need to optimize the geometry and weights for multiple look directions. Also the array size keeps getting bigger with iterations and needs to be bounded somehow. Therefore, taking the above properties into consideration, the following properties are desired in a out optimizations steps:

1. Generating geometries and weights for multiple look directions
2. Achieving desired WNG, DF and beampattern
3. Setting boundaries such that the array dimensions does not go beyond maximum size
4. Frequency invariant characteristics.

The weights need to be optimized using a trade-off with energy due to white noise and diffuse noise given by,

$$\begin{aligned} \underset{h(\omega),M}{\text{minimize}} \quad & \left[\mathbf{h}^H(\omega)\mathbf{\Gamma}_{\text{dn}}(\omega)\mathbf{h}(\omega) + \delta_w \mathbf{h}^H(\omega)\mathbf{h}(\omega)\right] \\ \text{subject to} \quad & \mathbf{D}(\omega, \theta_s)\mathbf{h}(\omega) = \boldsymbol{\beta} \end{aligned} \tag{42}$$

The term $\mathbf{h}^H(\omega)\mathbf{\Gamma}_{\text{dn}}(\omega)\mathbf{h}(\omega)$ in (42) represents the energy due to diffuse noise, the latter term $\mathbf{h}^H(\omega)\mathbf{h}(\omega)$ represents the energy due to the white noise and δ_w is the trade-off parameter between the two.

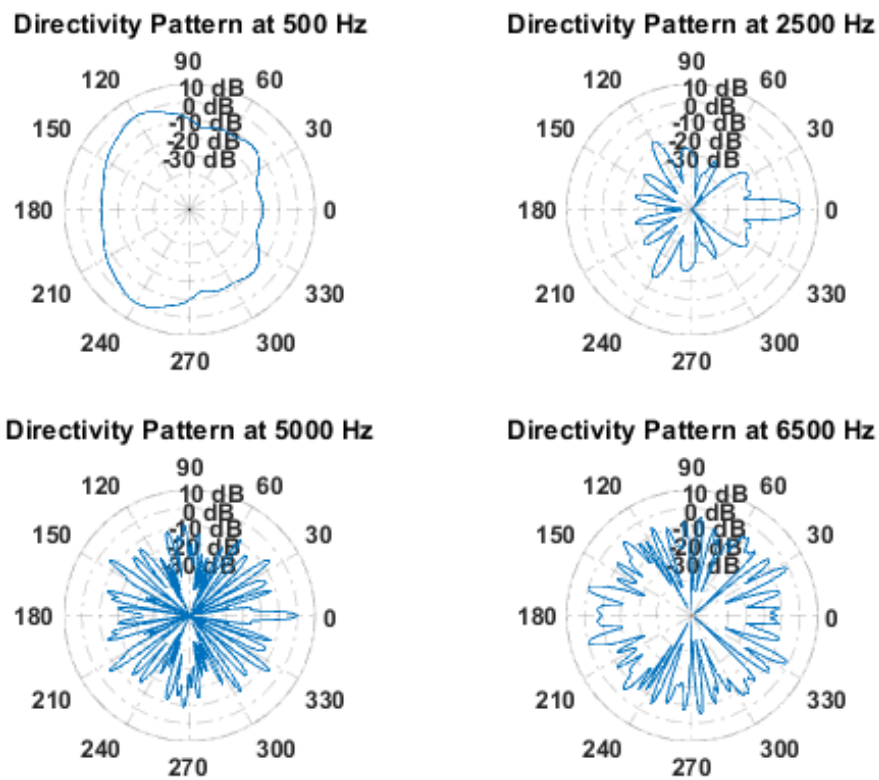


Figure 3. Beampatterns for adaptive planar array.

The optimization of (42) has many complications. Firstly, the minimization term cannot be considered as a common tap leakage because of the trade off parameter between white noise and diffuse noise.

In addition to robustness, the array geometry also needs to be optimized for other look directions as well. So far the constraint matrix, $\mathbf{D}(\omega, \theta)$ considered the first constraint $\mathbf{d}_1^H$ as look direction and $\mathbf{d}_2^H \ldots \mathbf{d}_P^H$ as null steering vectors. Similarly, the geometry needs to be optimized for look directions at the rest of the constraint angles $\theta_p : (p = 1 \ldots P)$ to have unity gain in order the steer the beam in other directions.

$$
\begin{aligned}
\underset{h(\omega),M}{\text{minimize}} \quad & \left[\mathbf{h}^H(\omega)\mathbf{\Gamma}_{\text{dn}}(\omega)\mathbf{h}(\omega) + \delta_w \mathbf{h}^H(\omega)\mathbf{h}(\omega)\right] \\
\text{subject to} \quad & \mathbf{D}(\omega, \theta_s = \theta_0)\mathbf{h}(\omega) = \boldsymbol{\beta}, \\
& \mathbf{D}(\omega, \theta_s = \theta_p)\mathbf{h}(\omega) = \boldsymbol{\beta}, \\
& \vdots \\
& \mathbf{D}(\omega, \theta_s = \theta_P)\mathbf{h}(\omega) = \boldsymbol{\beta}. \\
\text{subject to} \quad & \mathcal{D}(\mathbf{h}(\omega, \theta_s)) = \mathcal{D}_{target} \\
\text{subject to} \quad & \mathcal{W}(\mathbf{h}(\omega, \theta_s)) = \mathcal{W}_{target}
\end{aligned}
\tag{43}
$$

where $\mathcal{D}_{target}$ and $\mathcal{W}_{target}$ are the desired DF and WNG respectively.

The optimization problem in (43) maybe solved if either only the coefficients or geometry need to be optimized. However, with traditional optimization techniques for two parameters like gradient descent might not yield a solution in optimal sense for multiple constraints. This kind of optimization problem can be computationally expensive to solve because of the non linear cost function. We try to address the above mentioned problem by using a combination of PSO and a weights computation technique simultaneously in this section.

6. PROPOSED OPTIMIZATION

We discuss both the steps i.e. coefficients and geometry optimization simultaneously in this section. The proposed approach uses an optimal weights design whose the resulting DF and WNG are used to update the sensor element locations using PSO algorithm such that the fitness function is minimized.

6.1. OPTIMAL WEIGHTS

This section briefly describes the weight update mechanism used in the evaluation of fitness function (53). The minimization problem is the simpler version of (43). Where the minimization term, i.e. the energy due to diffuse noise and white noise with a tradeoff is considered as given below

$$\begin{aligned} \underset{h(\omega)}{\text{minimize}} \quad & \left[\mathbf{h}^H(\omega)\mathbf{\Gamma}_{\text{dn}}(\omega)\mathbf{h}(\omega) + \delta_w(\omega)\mathbf{h}^H(\omega)\mathbf{h}(\omega)\right] \\ \text{subject to} \quad & \mathbf{D}(\omega, \theta_s)\mathbf{h}(\omega) = \boldsymbol{\beta} \end{aligned} \tag{44}$$

Using the method of Lagrange multipliers in eq. (44), we define the cost function

$$\begin{aligned} J = \mathbf{h}^H(\omega)\mathbf{\Gamma}_{\text{dn}}(\omega)\mathbf{h}(\omega) + \delta_w(\omega)\mathbf{h}^H(\omega)\mathbf{h}(\omega) \\ +\boldsymbol{\lambda}^H(\boldsymbol{\beta} - \mathbf{D}\mathbf{h}(\omega)) + (\boldsymbol{\beta} - \mathbf{D}\mathbf{h}(\omega))^H\boldsymbol{\lambda} \end{aligned} \tag{45}$$

Setting the gradient of J to the all-zero vector and solving the Lagrangian for $\mathbf{h}(\omega)$, we obtain,

$$[\mathbf{\Gamma}(\omega) + \delta_w(\omega)\mathbf{I}]\,\mathbf{h}(\omega) - \mathbf{D}^H(\omega)\boldsymbol{\lambda} = \mathbf{0}. \tag{46}$$

For convenience let us define $\mathbf{\Gamma}_w(\omega) = [\mathbf{\Gamma}(\omega) + \delta_w(\omega)\mathbf{I}]$. Then, multiplying from the left by $\mathbf{\Gamma}_w^{-1}(\omega)$

$$\mathbf{h}(\omega) - \mathbf{\Gamma}_w^{-1}(\omega)\mathbf{D}^H(\omega)\boldsymbol{\lambda} = \mathbf{0} \tag{47}$$

and then multiplying again from the left by $\mathbf{D}$ and applying the constraint in (44) we obtain,

$$\boldsymbol{\beta} = \mathbf{D}(\omega)\mathbf{\Gamma}_w^{-1}(\omega)\mathbf{D}^H\boldsymbol{\lambda}. \tag{48}$$

The $P \times P$ matrix $\mathbf{D}(\omega)\mathbf{\Gamma}_w^{-1}(\omega)\mathbf{D}^H(\omega)$ is certainly rank deficient, if the number of constraints, P is greater than the number of microphones, M and even when that is not the case it often has a high condition number. In either case inverting $\mathbf{D}(\omega)\mathbf{\Gamma}_w^{-1}(\omega)\mathbf{D}^H(\omega)$ using the expedient

of diagonal loading, we get,

$$\boldsymbol{\lambda} = \left[\mathbf{D}(\omega)\boldsymbol{\Gamma}_w^{-1}(\omega)\mathbf{D}^H(\omega) + \delta_d(\omega)\mathbf{I}\right]^{-1}\boldsymbol{\beta}, \tag{49}$$

Applying (49) in (47) leads to the solution,

$$\mathbf{h}(\omega) = \boldsymbol{\Gamma}_w^{-1}(\omega)\mathbf{D}^H(\omega)\left[\mathbf{D}(\omega)\boldsymbol{\Gamma}_w^{-1}(\omega)\mathbf{D}^H(\omega) + \delta_d(\omega)\mathbf{I}\right]^{-1}\boldsymbol{\beta}. \tag{50}$$

we can reduce the computational complexity of equation (50) by applying the matrix inversion lemma. This leads to

$$\mathbf{h}(\omega) = \left[\delta_d(\omega)\boldsymbol{\Gamma}_w(\omega) + \mathbf{D}^H(\omega)\mathbf{D}(\omega)\right]^{-1}\mathbf{D}^H(\omega)\boldsymbol{\beta}. \tag{51}$$

Expanding $\boldsymbol{\Gamma}_w(\omega)$ and letting $\delta_T(\omega) = \delta_d(\omega)\delta_w(\omega)$, we obtain

$$\mathbf{h}(\omega) = \left[\delta_d(\omega)\boldsymbol{\Gamma}(\omega) + \delta_T(\omega)\mathbf{I} + \mathbf{D}^H(\omega)\mathbf{D}(\omega)\right]^{-1}\mathbf{D}^H(\omega)\boldsymbol{\beta}. \tag{52}$$

where, $\delta_d(\omega)$ is the diagonal loading constant.

The weights derived in (50) may be computed for $P \geq M$ constraints. This works well even when the look direction is not possible to achieve when traditional approaches require $P \leq M$ to always be true. Although overconstraining the constraint matrix in (51) does not improve the performance, we can increase the number of constraint angles such that the beam can be steered to the desired look direction.

6.2. JOINT OPTIMIZATION OF GEOMETRY AND WEIGHTS

We use PSO to find the best geometry for planar array. In addition to this, the parameters $\delta_w(\omega)$ and $\delta_d(\omega)$ need to be optimized as well to achieve the desired DF and WNG. PSO moves the microphone positions within a confined geometry such that the fitness function is minimized. The fitness function F is computed for multiple look directions by taking the mean squared error for the DF and WNG between the desired and the resulting PSO geometry given by,

$$\begin{aligned}\mathcal{F} = \sum_{\boldsymbol{\theta}} \Big(\sum_{\omega} W_{\text{dn}} \big| \big(\mathcal{D}_{tgt} - \mathcal{D}\left(\mathbf{h}(\omega, \theta_s)\right)\big|^2 \\ + W_{\text{wn}} \big| \big(\mathcal{W}_{tgt} - \mathcal{W}\left(\mathbf{h}(\omega, \theta_s)\right)\big|^2 \Big)\end{aligned} \tag{53}$$

where the $\mathcal{D}_{tgt}$ and $\mathcal{W}_{tgt}$ are the desired DF and WNG respectively. And W_{dn} and W_{wn} in the penalty imposed and derived empirically on desired DF and WNG.

The PSO algorithm is described in Figure 4. For given number of particles, the random geometries within confined area are initialized. The microphone coefficients are then computed using (50). The fitness function given by (53) is evaluated for each particle using the DF and WNG resulting from the microphone coefficients for each geometry. The PSO algorithm updates the microphone positions along with the parameters $\delta_w(\omega)$ and $\delta_d(\omega)$ and keeps track of the best particle with minimum fitness score at each iteration.

6.3. GEOMETRY AND COEFFICIENTS UPDATE USING PSO

The PSO algorithm considers a certain set of particles and moves them in a specific search space to satisfy the fitness function (53). For given number of particles K, each particle corresponds to a geometry (consisting of the radius) and a regularization parameter.

$$\boldsymbol{X} = [\boldsymbol{X}_1 \ \boldsymbol{X}_2 \ \ldots \ \boldsymbol{X}_K] \tag{54}$$

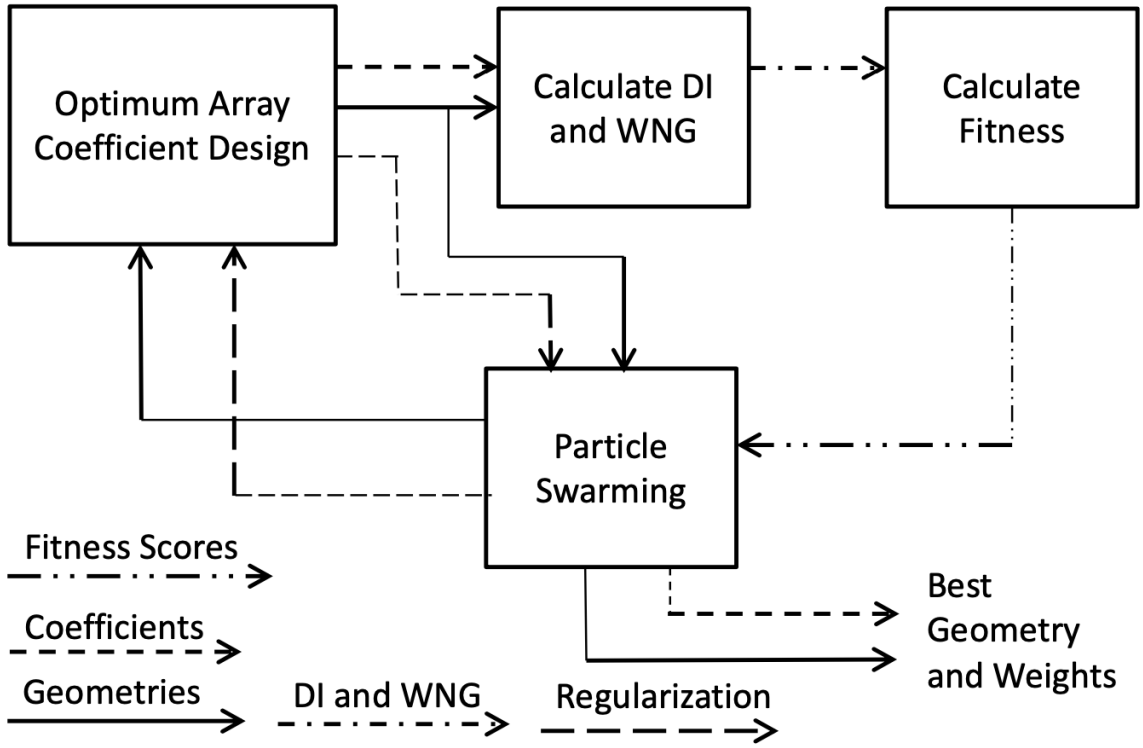


Figure 4. Block diagram of the PSO algorithm for geometry and weights optimization.

where each column

$$\boldsymbol{X}_k = [\mathbf{r}_k \ \boldsymbol{\delta}_w \ \boldsymbol{\delta}_d]^T \tag{55}$$

and

$$\mathbf{r}_k = [r_{k,1} \ r_{k,2} \ \ldots \ r_{k,M}] \tag{56}$$

is the position of the mth microphone and k^{th} particle and the vectors $\boldsymbol{\delta}_w, \boldsymbol{\delta}_d$ consist of the tradeoff parameters $\delta_w(\omega), \delta_d(\omega)$ for all subbands ω.

6.4. PSO UPDATE STEPS

The velocity of the particles is updated as

$$\begin{aligned}\mathcal{V}(i+1) = a\mathcal{V}(i) &+ c_1 \odot (\mathcal{X}_p - \mathcal{X}(i)) \odot \text{rand}() \\ &+ c_2 \odot (\mathcal{X}_s - \mathcal{X}(i)) \odot \text{rand}() \\ &+ c_3 \odot (\mathcal{X}_g - \mathcal{X}(i)) \odot \text{rand}()\end{aligned} \tag{57}$$

where $\mathcal{X}_p$, $\mathcal{X}_s$ and $\mathcal{X}_g$ are the personal, social and global position of particles.

The velocity of particles beyond the maximum weights are negated

$$\begin{aligned}\mathcal{V}(i+1) \ < \ \text{b}_{\text{lo}} &= -\mathcal{V}(i+1) \\ \mathcal{V}(i+1) \ > \ \text{b}_{\text{up}} &= -\mathcal{V}(i+1)\end{aligned} \tag{58}$$

The particles position is updated using the

$$\mathcal{X}(i+1) = b \odot \mathcal{X}(i) + \mathcal{V}(i+1) \tag{59}$$

6.4.1. Personal Best. The personal best fitness function of each particle k is denoted by

$$\boldsymbol{F_p} = [\mathcal{F}_{p_1} \ \ldots \ \mathcal{F}_{p_k} \ \ldots \ \mathcal{F}_{p_K}] \tag{60}$$

and the personal best position is given by

$$\boldsymbol{X_p} = \left[\boldsymbol{X}_{p_1} \ \boldsymbol{X}_{p_2} \ \ldots \ \boldsymbol{X}_{p_K}\right] \tag{61}$$

where $\boldsymbol{X}_{p_k}$ is the personal best position of each particle k. The personal best position $\mathcal{X}_{p_k}$, is computed as

$$\boldsymbol{X}_{p_k} = \begin{cases} \boldsymbol{X}_{p_k}, \ \mathcal{F}_k > \mathcal{F}_{p_1} \\ \boldsymbol{X}_k, \ \mathcal{F}_k < \mathcal{F}_{p_1} \end{cases} \tag{62}$$

6.4.2. Local Best. The local (social) best position for each particle k is given by

$$\boldsymbol{X_s} = [\boldsymbol{X}_{s_1}\ \boldsymbol{X}_{s_2}\ \ldots\ \boldsymbol{X}_{s_K}] \tag{63}$$

where $\boldsymbol{X}_{s_k}$ is the local best position of particle k. The local best position of each particle k is found among it's closest neighbors $\boldsymbol{N}_k$

$$\boldsymbol{X}_{s_k} \epsilon \{\boldsymbol{N}_k | \boldsymbol{X}_{s_k} = \min\{\boldsymbol{F}_p\},\ \forall\ \boldsymbol{F}_p\ \epsilon\ \boldsymbol{N}_k\} \tag{64}$$

where $\boldsymbol{N}_k$ is a matrix with Q neighbors of particle k

$$\boldsymbol{N}_k = [\boldsymbol{X}_{p_1}\ \boldsymbol{X}_{p_q}\ \ldots\ \boldsymbol{X}_{p_Q}] \tag{65}$$

Hence the local best position of each particle k is $\boldsymbol{X}_{p_q}$ with lowest personal best fitness value.

6.4.3. Global Best. The global best fitness value $\mathcal{F}_g$ is the minimum value among all the personal best fitness values and the corresponding particle is denoted by $\mathcal{X}_g$

We find the so called son particle which is approximated as

$$\mathcal{X}_{son} = \mathcal{X}_g + delta \tag{66}$$

where $delta$ is the average sum of random particles.

Once we have $\mathcal{X}_{son}$ we evaluate the fitness for this particle say $\mathcal{F}_{son}$. If the fitness of son particle is less than $\mathcal{F}_g$, then we make son as the new global particle i.e.

$\mathcal{X}_g = \mathcal{X}_{son}\ \ and\ \ \mathcal{F}_g = \mathcal{F}_{son}$

6.4.4. Nudge Function. In our experiments, each particle corresponds to a geometry and tradeoff parameters. We need to avoid the microphone positions getting too close or even on top of one another. For we have a minimum interelement distance between each microphone to avoid severe white noise amplification. To ensure this, we propose the nudging function whose algorithm is as described in Algorithm 1.

```
for k ← 1 to particles do
    X_k = X(:, k)
    for l ← 1 to M do
        r_{k,l} = X_k(l)
    end
    for iterate ← 1 to 10 do
        for i ← 1 to M do
            for j ← i + 1 to M do
                if r_{k,i} − r_{k,j} < 1 then
                    r_{k,i}=r_{k,i} − sign(rand(1) − 0.5)
                    r_{k,j}=r_{k,j} − sign(rand(1) − 0.5)
                end
            end
        end
    end
end
```

Algorithm 2: Algorithm for nudging function

Taking the above equations into consideration, the following algorithm is suggested.

for $i \leftarrow 1$ **to** $iterations$ **do**

  for $k \leftarrow 1$ **to** $particles$ **do**

    for $l \leftarrow 1$ **to** $directions$ **do**

      for $f \leftarrow 1$ **to** $f_s/2$ **do**

        For particle $\boldsymbol{X}_k$, calculate:

        $\mathbf{h}_k(\omega)$, $\mathcal{W}[\mathbf{h}_k(\omega)]$, $\mathcal{D}[\mathbf{h}_k(\omega)]$, $\mathcal{GG}[\mathbf{h}_k(\omega)]$

      end

    end

    Evaluate fitness $\mathcal{F}_k$

    % Personal best

    if $\mathcal{F}_k \leq \mathcal{F}_p$ **then**

      $\mathcal{F}_{p_k}$=$\mathcal{F}_k$

      $\boldsymbol{X}_{p_k}$=$\boldsymbol{X}_k$

    end

  end

  % local best

  Local best particle is chosen form the 5 neighbors with smallest fitness value

  $\boldsymbol{X}_{s_k} = \boldsymbol{X}_{p_k} \epsilon \min\left[k, \epsilon\, j\left[\min_{1\ldots5} \sum (\boldsymbol{X_k} - \boldsymbol{X}_j)^2\right]\right]$

  % Global best

  $\boldsymbol{X}_g = \boldsymbol{X}_{p_k} \epsilon \min\left[\boldsymbol{F}_{p_k}\right]$

  % Update Velocity

  % Reflect velocity when particle hits wall

  % Update particles

  $\mathcal{X}_{i+1} = b \odot \mathcal{X}_i + \mathcal{V}_{i+1}$

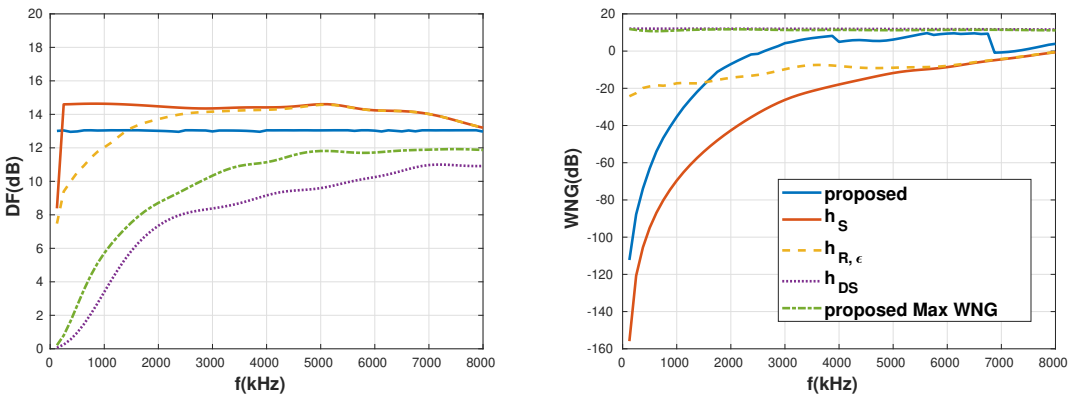
  $nudge(\boldsymbol{X}_{i+1})$

end

Algorithm 3: PSO algorithm for optimal 2D geometry

7. SIMULATION RESULTS

The proposed approach is used to optimize a planar array consisting of 16 microphones placed within $20cm \times 20cm$ area. We consider a swarm of 30 particles with each particle representing a geometry and coefficient parameters $\delta_w(\omega)$ and $\delta_d(\omega)$ for wideband frequencies ranging from (0-8000) Hz and 24 constraint angles at every 15^0 uniformly divided from $(0 - 360)^0$. The minimum inter-element distance between microphones is set to half centimeter.



(a) (b)

Figure 5. Comparison of performance characteristics of PSO optimized planar array and conventional techniques. (a) DF (b) WNG. Array gains for PSO optimized array with $\mathcal{D}_{tgt} = 13dB$ (blue solid line), superdirective $\mathbf{h}_S$ (red line), regularized superdirective $\mathbf{h}_{R,\epsilon}$, $\epsilon = 10^{-4}$ (dashed), delay-sum $\mathbf{h}_{ds}$ (dotted) and PSO optimized array with $\mathcal{W}_{tgt} = 12dB$ (dash-dot).

First, we consider the performance of the proposed approach with conventional approaches such as delay-sum beamformer $\mathbf{h}_{ds}(\omega, \theta_s)$ in (22), superdirective beamformer $\mathbf{h}_S(\omega, \theta_s)$ given in (23) and the regularized superdirective beamformer $\mathbf{h}_{R,\epsilon}(\omega, \theta_s)$ in (24). The proposed approach is configured to robust superdirective and maximum WNG. As for the robust superdirective configuration we set the $\mathcal{D}_{tgt} = 13dB$ and the weights W_{dn} and W_{wn} in the fitness function are set to 100 and 40 respectively. The results are as indicated in Figure 5. The proposed approach achieves quite consistent DF in all subbands. It's performance drops only by about (1-1.5) dB compared to the superdirective and regularized

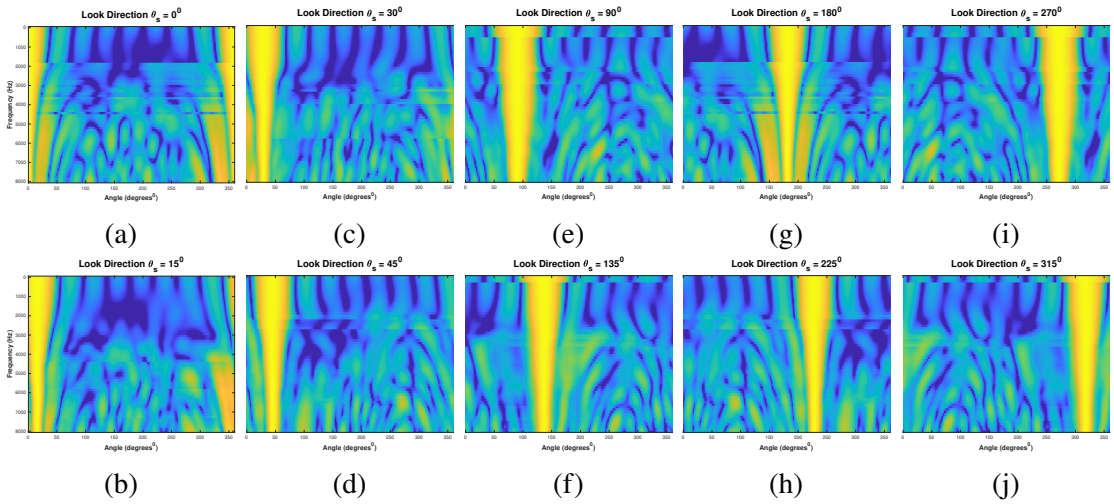


Figure 6. Wide-band beampattern for various look directions θ_s for $\mathcal{D}_{tgt} = 12dB$. (a) $\theta_s = 0^0$, (b) $\theta_s = 15^0$, (c) $\theta_s = 30^0$, (d) $\theta_s = 45^0$ and (e) $\theta_s = 90^0$, (f) $\theta_s = 135^0$, (g) $\theta_s = 180^0$, (h) $\theta_s = 225^0$, (i) $\theta_s = 270^0$ and (j) $\theta_s = 315^0$.

superdirective beamformers. The white noise amplification is severe for both conventional approaches as compared to proposed approach which has good WNG except for frequencies below 1kHz. Similarly, to achieve maximum WNG configuration, the $\mathcal{W}_{tgt}$ is set to $12dB$ and $(W_{\text{dn}}, W_{\text{wn}}) = (0, 100)$. As seen in Figure 5, the proposed approach has WNG very similar to the delay sum beamformer but has DF greater across all frequencies.

The proposed approach has been demonstrated to outperform conventional techniques. The steering capability is also tested. I.e., the performance characteristics of the beamformer need to be evaluated for various look directions to achieve similar results if not exactly same. The constraint angles vector $\boldsymbol{\theta}$ for $P = 24$ constraints is given as $\boldsymbol{\theta} = [0^0\ 15^0\ 30^0\ \ldots\ 345^0]$. Hence, we optimize the planar array geometry within $20cm \times 20cm$ to achieve $\mathcal{D}_{tgt} = 12dB$ for all look directions given by $\boldsymbol{\theta}$. The weights $(W_{\text{dn}}, W_{\text{wn}})$ are set to $(100, 50)$ to achieve robust yet frequency invariant response. The geometry of the optimized array can be seen in Figure 7 (b). We plot the SNR gains for ten different look directions i.e. $\theta_s = [0^0, 15^0, 30^0, 45^0, 90^0, 135^0, 180^0, 225^0, 270^0\ and\ 315^0]$ and the DF shown in Figure 7 (a) is very similar for all positions to which the beam is

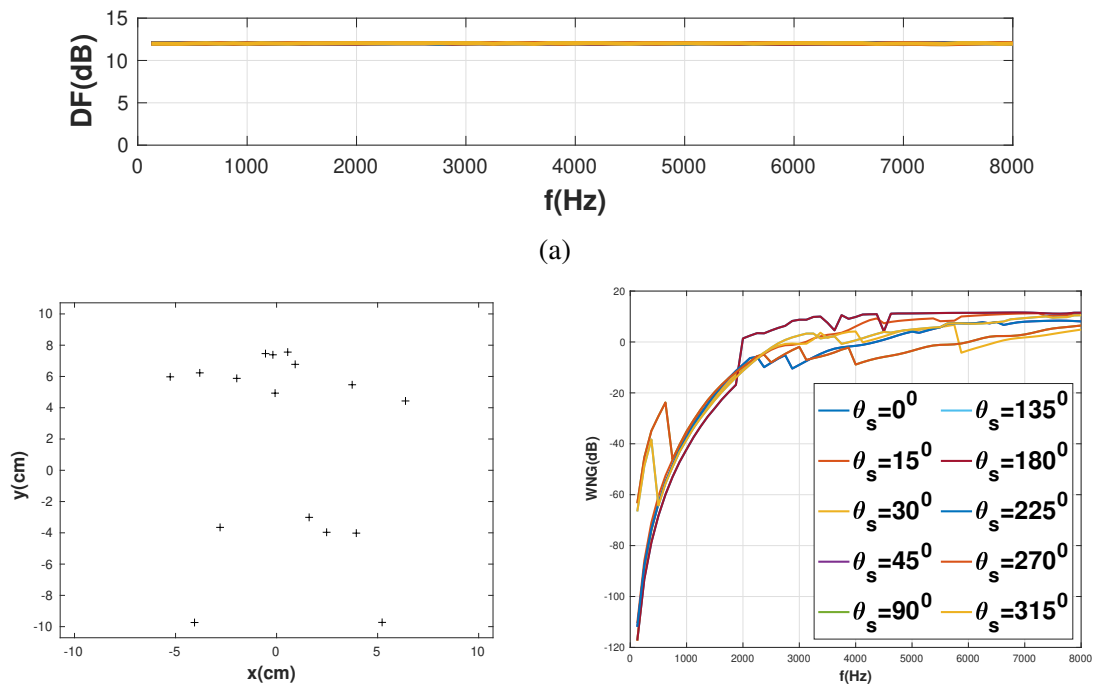
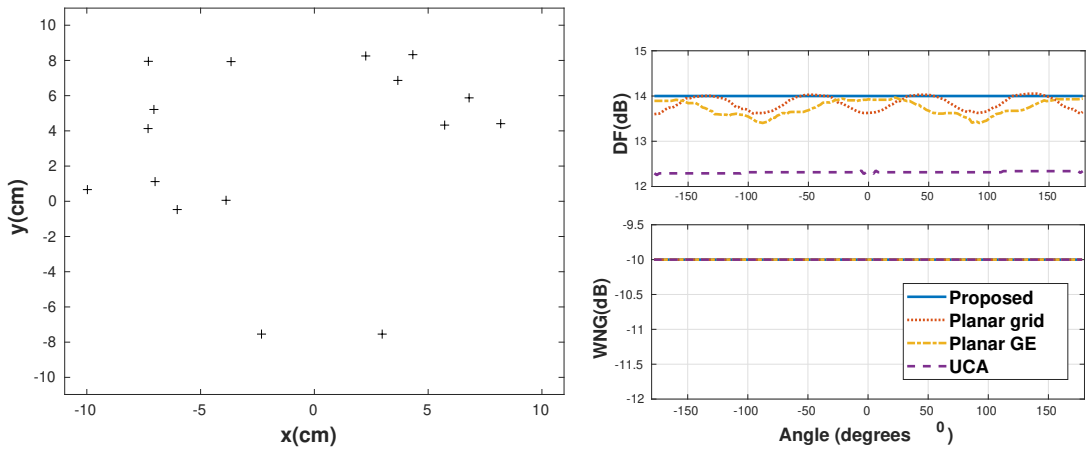


Figure 7. Performance characteristics for various look directions θ_s and $\mathcal{D}_{tgt} = 12dB$, (a) DF , (b) Array geometry and (c) WNG.

steered. Similary, the WNG, although not exactly same for all look directions as DF, as shown in Figure 7 (c) is quite similar as well. The wideband beampatterns for the optimized array at angles θ_s are given in Figure 6. The array is optimized for all 24 look directions of $\boldsymbol{\theta}$. However, we only plot for few angles because of the space constraints.

Finally, we compare the performance for the proposed approach with recently proposed GA based planar array optimization [33] along with other geometries. The GA optimizes the array for superdirective response at $f = 1kHz$ only such that $\mathcal{W}_{tgt} = -10dB$. Moreover, the DF is optimized only for three look directions i.e. $[30^0, 0^0 and - 30^0]$. Whereas the proposed approach optimizes the array for wideband frequencies and any look direction. The other geometries compared are a uniform circular array (UCA) $20cm$ in diameter and a 4×4 uniform planar array of dimensions $20cm \times 20cm$. The baseline methods all have maximum dimension of $20cm$ across both x and y dimension. The proposed

optimized array is given in Figure8 (a). The maximum aperture length across x and y is $18cm \times 16cm$ hence yielding smaller array aperture while maintaining exactly similar higher DF than any of the discussed methods for all possible look directions.



(a) (b)

Figure 8. Performance characteristics for various proposed geometries at all possible look directions θ_s and $\mathcal{D}_{tgt} = 14dB$ at f=1kHz. Top: DF and Bottom: WNG. Proposed (blue solid line), Rectangular grid array(red dotted line), GE based planar array [33] (dashed-dot), UCA (dashed). (a) Array geometry, (b) DF and WNG.

8. CONCLUSIONS

The proposed approach for optimized planar array geometry for a given maximum aperture and the desired DF or WNG for multiple look directions. The approach presented in this paper can achieve robust super directive frequency invariant response. We compared our results with the conventional techniques such as delay-sum, superdirective and regularized superdirective beamformer to achieve a decent compromise between DF and WNG for multiple subbands. We also compare our approach with UCA, uniform grid and the recently proposed GA based array optimization at 1kHz with the proposed approach reducing the aperture by 10% and 20% along x and y dimension respectively. Although the proposed has been demonstrated for wideband, it should be noted that for this particular case proposed approach was only for single frequency.

REFERENCES

[1] Brandstein, M., & Ward, D. (Eds.). (2001). Microphone arrays: signal processing techniques and applications. Springer Science & Business Media.

[2] Elko, G. W., & Meyer, J. (2008). Microphone arrays. In Springer handbook of speech processing (pp. 1021-1041). Springer, Berlin, Heidelberg.

[3] Benesty, J., Chen, J., & Huang, Y. (2008). Microphone array signal processing (Vol. 1). Springer Science & Business Media.

[4] Benesty, J., Cohen, I., & Chen, J. (2017). Fundamentals of signal enhancement and array signal processing. John Wiley & Sons.

[5] Van Veen, B. D., & Buckley, K. M. (1988). Beamforming: A versatile approach to spatial filtering. IEEE assp magazine, 5(2), 4-24.

[6] Cox, H., Zeskind, R., & Kooij, T. (1986). Practical supergain. IEEE Transactions on Acoustics, Speech, and Signal Processing, 34(3), 393-398.

[7] Cox, H., Zeskind, R., & Owen, M. (1987). Robust adaptive beamforming. IEEE Transactions on Acoustics, Speech, and Signal Processing, 35(10), 1365-1376.

[8] Benesty, J., & Jingdong, C. (2012). Study and design of differential microphone arrays (Vol. 6). Springer Science & Business Media.

[9] McCowan, I. (2001). Microphone arrays: A tutorial. Queensland University, Australia, 1-38.

[10] Pal, P., & Vaidyanathan, P. P. (2010). Nested arrays: A novel approach to array processing with enhanced degrees of freedom. IEEE Transactions on Signal Processing, 58(8), 4167-4181.

[11] Teutsch, H. (2007). Modal array signal processing: principles and applications of acoustic wavefield decomposition (Vol. 348). Springer.

[12] Chen, J., Benesty, J., & Huang, Y. (2008). A minimum distortion noise reduction algorithm with multiple microphones. IEEE transactions on audio, speech, and language processing, 16(3), 481-493.

[13] Schmid, D., Enzner, G., Malik, S., Kolossa, D., & Martin, R. (2014). Variational Bayesian inference for multichannel dereverberation and noise reduction. IEEE/ACM transactions on audio, speech, and language processing, 22(8), 1320-1335.

[14] Yang, W., Huang, G., Zhang, W., Chen, J., & Benesty, J. (2018, September). Dereverberation with differential microphone arrays and the weighted-prediction-error method. In 2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC) (pp. 376-380). IEEE.

[15] Yu, J., & Donohue, K. D. (2013). Optimal irregular microphone distributions with enhanced beamforming performance in immersive environments. The Journal of the Acoustical Society of America, 134(3), 2066-2077.

[16] Shanan, S. S., & Pomalaza-Raez, C. A. (1989). The use of nonuniform element spacing in array processing algorithms. The Journal of the Acoustical Society of America, 86(4), 1416-1418.

[17] Bjelić, M., Stanojević, M., Šumarac Pavlović, D., & Mijić, M. (2017). Microphone array geometry optimization for traffic noise analysis. The Journal of the Acoustical Society of America, 141(5), 3101-3104.

[18] Yu, J., & Donohue, K. D. (2012). Geometry descriptors of irregular microphone arrays related to beamforming performance. EURASIP Journal on Advances in Signal Processing, 2012(1), 1-12.

[19] Crocco, M., & Trucco, A. (2012). Stochastic and analytic optimization of sparse aperiodic arrays and broadband beamformers with robust superdirective patterns. IEEE Transactions on audio, speech, and language processing, 20(9), 2433-2447.

[20] Trucco, A. (2002). Weighting and thinning wide-band arrays by simulated annealing. Ultrasonics, 40(1-8), 485-489.

[21] Patel, S. J., Grant, S. L., Zawodniok, M., & Benesty, J. (2018, September). On the design of optimal linear microphone array geometries. In 2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC) (pp. 501-505). IEEE.

[22] Haupt, R. L. (1994). Thinned arrays using genetic algorithms. IEEE transactions on antennas and propagation, 42(7), 993-999.

[23] Caorsi, S., Lommi, A., Massa, A., & Pastorino, M. (2004). Peak sidelobe level reduction with a hybrid approach based on GAs and difference sets. IEEE Transactions on Antennas and Propagation, 52(4), 1116-1121.

[24] Chen, K., Yun, X., He, Z., & Han, C. (2007). Synthesis of sparse planar arrays using modified real genetic algorithm. IEEE Transactions on Antennas and Propagation, 55(4), 1067-1073.

[25] Li, Z., Yiu, K. F. C., & Feng, Z. (2013). A hybrid descent method with genetic algorithm for microphone array placement design. Applied Soft Computing, 13(3), 1486-1490.

[26] Yu, J., Yu, F., & Li, Y. (2013). Optimization of Microphone Array Geometry with Evolutionary Algorithm. J. Comput., 8(1), 200-207.

[27] Le Courtois, F., Thomas, J. H., Poisson, F., & Pascal, J. C. (2016). Genetic optimisation of a plane array geometry for beamforming. Application to source localisation in a high speed train. Journal of Sound and Vibration, 371, 78-93.

[28] Khatami, I., & Abdollahzadeh Jamalabadi, M. Y. (2021). Optimal design of microphone array in a planar circular configuration by genetic algorithm enhanced beamforming. Journal of Thermal Analysis and Calorimetry, 145(4), 1817-1825.

[29] Pessentheiner, H., Kubin, G., & Romsdorfer, H. (2012, September). Improving beamforming for distant speech recognition in reverberant environments using a genetic algorithm for planar array synthesis. In Speech Communication; 10. ITG Symposium (pp. 1-4). VDE.

[30] Hawes, M. B., & Liu, W. (2014). Sparse array design for wideband beamforming with reduced complexity in tapped delay-lines. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 22(8), 1236-1247.

[31] Buchris, Y., Amar, A., Benesty, J., & Cohen, I. (2018). Incoherent synthesis of sparse arrays for frequency-invariant beamforming. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 27(3), 482-495.

[32] Le Son, P. (2020). On the Design of Sparse Arrays With Frequency-Invariant Beam Pattern. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 29, 226-238.

[33] Chen, X., Pan, C., Chen, J., & Benesty, J. (2021, June). Planar Array Geometry Optimization for Region Sound Acquisition. In ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 756-760). IEEE.

III. PASSIVE RFID TAGS FOR METALLIC ENVIRONMENTS USING PHASED ARRAY READER ANTENNAS

Sohel J. Patel, Maciej Zawodniok
Department of Electrical & Computer Engineering
Missouri University of Science and Technology
Rolla, Missouri 65409–0050
Email: sjp432@mst.edu, mjzx9c@mst.edu

ABSTRACT

In this study, we consider the operation of radio frequency identification (RFID) tag antennas in metallic environments. Phased array is built using 2 RFID reader antennas for electronic beam steering and improving overall read range of the RFID antenna tags. The performance of RFID tag antenna is known to degrade in metallic environments. RFID tag antenna with ground plane is designed to improve the performance in harsh metallic environments. In this study, we use double slit antennas with ground plane for tags radiating in challenging metallic environments to achieve a 30 feet read range.

1. INTRODUCTION

Passive RFID applications consider the use of a Reader antenna which powers the RFID antenna tags within its transmit/receiving range. RFID systems have proved to be more useful for applications like automatic identification and data capture especially for inventory control [1]. The RF signal of RFID antenna may be distorted due to various reasons that inhibit the energy transfer to the IC. Some reasons include but not limited to operation of RFID tag antennas in metal environments and impedance mismatch with the IC.

When the antennas are mounted on metallic surfaces [2], the current distribution of the RFID tag antennas is distorted, leading to limited or read range. The performance of RFID tags when mounted on metal surfaces have been extensively studied in [3, 4, 5, 6, 7, 8]. Several RFID antenna tag designs have been proposed with optimized performance in metallic surface environments [9, 10, 11, 12].

The impedance characteristics between the RFID tag antenna and IC also need to be optimized as well. Otherwise it may yield in improper characteristic impedance than that of the IC. For all ICs may not always have 50Ω impedance.

In our experiments, the read range of approximately 10 meters in metal environments is desired. This may be quite difficult to achieve on metal surfaces given the distorted RF signal[5, 6, 7]. For example the designs in [14, 15, 16, 17] are measured to have a read range of 4-6 meters when RFID tag antennas are placed on metal surfaces. In our applications, we consider the antenna design with slots [13]. These slots act as inductor and vary in sizes in order to match with the input impedance of the antenna to that of the IC.

To improve the read range for a RFID antenna tag at a specific location, an array of reader antennas is setup. This is achieved by controlling the scanning angle at any given time using phase shifts between consecutive antennas. This can be implemented in two ways. Either we consider using an array of RFID antenna tags [18, 19]. Alternatively we can use an array of reader antennas [20]. Such techniques involve using multiple antennas installed in a specific geometry to improve the performance characteristics like enhanced data range, coverage and capacity. One commonly used technique is beamforming [21]. Where the phase shifts between consecutive antenna elements is varied in a controlled manner to scan for RFID antenna tags only from specific directions.

Phased array antennas have been utilized in various applications. For example RADAR, SONAR, communications, geophysical imaging and many more. In this study we use upto 4 antennas inline with a voltage controlled phase shifter.

2. PHASED ARRAY ANTENNAS

Phased array antennas have been used in various applications where the direction transmission and reception needs to happen in a confined area may be altered at convenience. This involves altering phase of each antenna element and based on the geometry these phase changes can be different for different directions. However the radiation pattern of the phased array is a combination of individual antenna's electric field and the combined electric field of the array.

Consider two dipoles on a line along the x axis. The total electric field can be given as [22]

$$\mathbf{E}_t = \mathbf{E}_1 + \mathbf{E}_2 \tag{1}$$

Where

$$\mathbf{E}_1 = \hat{\mathbf{a}}_\theta j\eta \frac{kI_0 l e^{-j[kr_1 - \frac{\beta}{2}]}}{4\pi r_1} \cos\theta_1 \tag{2}$$

$$\mathbf{E}_2 = \hat{\mathbf{a}}_\theta j\eta \frac{kI_0 l e^{-j[kr_2 - \frac{\beta}{2}]}}{4\pi r_2} \cos\theta_2 \tag{3}$$

is the electric field of individual dipoles, θ_1 and θ_2 are the angle of the dipoles in the azimuth plane from the radiating source, β is the phase difference between two consecutive elements. If we assume the magnitude of each radiating antenna is same in farfield observations, then:

$$\begin{aligned} \theta_1 &\simeq \theta_2 \simeq \theta \\ r_1 &\simeq r - \frac{d}{2}\cos\theta \\ r_2 &\simeq r + \frac{d}{2}\cos\theta \\ r_1 &\simeq r_2 \simeq r \end{aligned} \tag{4}$$

Using (2) - (4), (1) can be rewritten as,

$$\begin{aligned}\mathbf{E}_t &= \hat{\mathbf{a}}_\theta j\eta \frac{kI_0 l e^{-jkr}}{4\pi r} \cos\theta [e^{+j\frac{kd\cos\theta+\beta}{2}} + e^{-j\frac{kd\cos\theta+\beta}{2}}] \\ &= \hat{\mathbf{a}}_\theta j\eta \frac{kI_0 l e^{-jkr}}{4\pi r} \cos\theta \times \left\{2\cos\left[kd\cos\theta+\beta\right]\right\}\end{aligned} \tag{5}$$

The equation (5) can be decomposed into two parts. element factor and array factor. Since the element factor is constant for any given element which is the first part of the equation (5). Therefore, the array factor is given by,

$$AF_{Linear} = \sin\left[(kd\sin\theta+\beta)\right] \tag{6}$$

It would not be difficult to derive the array factor for N elements

$$AF_{Linear} = \sin\left[\frac{N}{2}(kd\sin\theta+\beta)\right] \tag{7}$$

Hence the total electric field can be given as the product of element factor and array factor.

$$\mathbf{E}_t = [\mathbf{E}_{element}] \times [\mathrm{AF}] \tag{8}$$

A planar array consists of the antenna elements on a plane uniformly placed on vertices of a square say on XY plane. The array factor for such a planar array can be given by

$$\begin{aligned}AF_{Planar} = &\sum_{n=1}^{N} e^{jkd(n-1)\cos\theta\sin\phi+\beta_x} \\ &\times \sum_{m=1}^{M} e^{jkd(m-1)\sin\theta\sin\phi+\beta_y}\end{aligned} \tag{9}$$

where ϕ is the elevation angle. and β_x and β_y being the phase excitation along the x and y axis linear arrays. The array factor in (1) is the product of the array factor of linear arrays along x and y axis. The general form of the array factor for a planar array can be given by

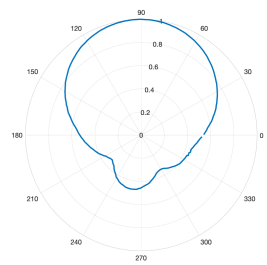


Figure 1. Element factor of the RFID reader antenna.

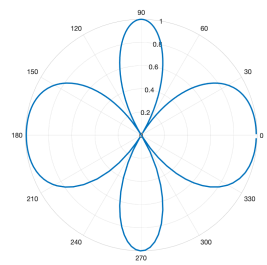


Figure 2. Array factor of the antenna array.

$$AF_{Planar} = \sin(N\frac{kd\cos\theta\sin\phi + \beta_x}{2}) \times \sin(M\frac{kd\sin\theta\sin\phi + \beta_y}{2}) \tag{10}$$

The RFID reader antenna used in our measurement is RFMax S9028PC [23], whose radiation pattern is shown in "Figure. 1". The antenna can be seen to have it's own radiation pattern. It can be observed that 3dB beamwidth of the antenna falls between [45^0 - 135^0] i.e. 90^0. Hence even if we steer the beam electronically, we cannot get directivity out of this range even if the array factor may lead to transmission/reception direction outside this range. Hence individual element factor plays an important role in overall radiation pattern. In "Figure. 2", it can be seen that the array factor for 2 elements when combined with the element factor from "Figure. 1" as given in (8) yields the total electric field and radiation pattern as shown in "Figure. 3"

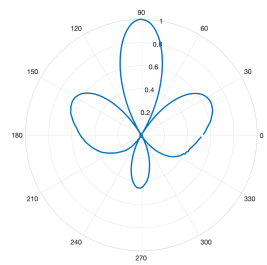


Figure 3. Total electric field of the RFID reader antenna array.

The RFID reader antennas are as shown in “Figure. 13”. The interelement distance between any two consecutive antennas is the distance between the center of each antenna. We chose the interelement distance of 12 inches. The minimum achievable interelement distance is 10 inches for the antennas are of size 10 × 10 inches. The maximum interelement distance for any antenna array has to be the half the wavelength associated with the highest frequency to avoid spatial aliasing/grating lobes. Hence the maximum interelement distance is

$$d = \frac{\lambda}{2} = \frac{c}{2f}. \tag{11}$$

where ’c’ is the velocity of light = 3×10^8 m/s $\simeq$ 118.11×10^8 inches/s and f is the highest frequency = 925 MHz in our case.

The interelement distance has to satisfy (11) in order to avoid spatial aliasing. Four our applications the RFID operating frequency range is 902-928 MHz. This yields in the maximum interelement distance not to exceed 6.36 inches. However, this is not possible practically given the antenna size leading to grating lobes.

Figure 4. Design of dipole RFID antenna tag.

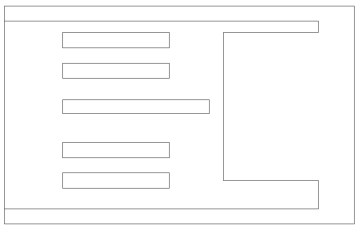


Figure 5. Double slit RFID antenna tag design.

3. RFID ANTENNA TAGS

We use two kinds of RFID antenna tag in our study. A dipole like and a double slit antenna tag [13]. The former works equivalently well as compared to latter in normal environments. However, the performance of dipole antenna degrades when placed on metallic surfaces. The read range of these antennas in absence of metallic environments were measured to be about 80-90 feet. Whereas on a metal surface, about 5 mm above the metal surface in enclosures as seen in “Figure. 6” and “Figure. 7”, these antennas have a read range of 5-10 feet. In our experiments, the desired read range for the RFID antenna tags is about 30feet.

In such environments, the double slit antennas dominate in performance. They have a ground plane separated from the top radiating element connected by a small patch on the side. The radiating element of both antennas are as shown in “Figure. 4” and “Figure. 5”.

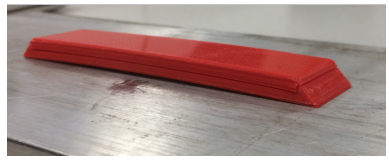


Figure 6. Enclosure for the RFID dipole antenna made from ABS.

Figure 7. Enclosure for the RFID dipole antenna made from black TPU.

The slots made in the antenna in "Figure. 5" act as inductors and their sizes can be altered to match the antenna impedance with that of the chip. The antenna operates in UHF range around 915 MHz. The characteristic impedance of the chip Monza R6 [15] is $Z_c = 11.65 - j117.69\Omega$. The substrate Polyethylene Terphalate (PET) is used for the isolation or ground plane. The readers can refer to [15] for the exact dimensions of the double slit antenna.

The dipole antennas in "Figure. 4" is used only for reference as they perform much better in absence of harsh metallic environments. However for the experiments involving the RFID antenna tags performance on metal surfaces, the double slit antennas will be used for the measurements.

4. RFID ANTENNA ENCLOSURE

We would like to discuss on the enclosures for we found some interesting properties during our measurements with the enclosures made of vrious materials. The antenna tags were packed in small enclosured made of varoius materials based on the applications. The

Figure 8. Enclosure for the RFID double slit antenna made from ninjaflex.

most basic material is Acrylonitrile butadiene styrene (ABS) [24] for it's cheap cost and ease of manufacturing. However, for applications involving placement of the RFID tags on curved or uneven surfaces, it requires the enclosures be made of flexible materials.If not all of it, atleast on the bottom part of the enclosure needs to be flexible such that the complete tag can confirm to any surface. One such material is Thermoplastic polyurethane (TPU) [25]. It has many desirable properties among flexibility, durability. However thses materials are not flexible enough to be bent if the enclosure edge is less than 2-3 cm. During experiments it was observed that the materials made of black ink inhibited the performance of the antennas. This infact could be because the way the black ink is processed [27]. However, there are other classes or derivatives of TPU called ninjaflex [26] that have more flexibility and durability at the same time.

The enclosures made of ABS can be seen in figure "Figure. 6" made for dipole antenna. Also in "Figure. 7" the TPU enclosure can be seen how flexible it is. This is quite desirable for our applications. The material we used for manufacturing our enclosures was ninjaflex. which is type of TPU with shore hardness 85A and 660% elongation. This is specifically attractive to our application for the antenna substrate is made of PET which is form of plastic and cannot be conformed onto any nonuniform surface. Therefore the enclosure needs to be made of material that can confirm to nonuniform surface to some extent.

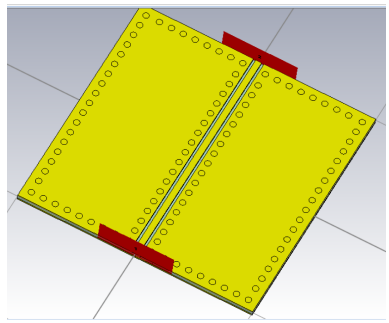


Figure 9. Grounded CoPlanar Waveguide (GCPW) simulated in CST studio with characteristic impedance of 50Ω.

5. PHASE SHIFTER

We use the voltage controlled phase shifter ICs used for feeding the antennas is PS088-315 by skyworks solutions[2]. It features the operating frequencies 700-1100 MHz with a phase shift ranging from 85-105 degrees with 85^0 representing 0V and 105^0 representing 12V. The most important property of this phase shifter is it's bidirectional phase shifting. This is specifically interesting in our applications for we need the phase shift not just for transmission but as well for reception. The readers are requested to refer [2] for more details about the s-parameters performance. And characteristic impedance of the IC is 50Ω.

We simulated a grounded coplanar waveguide (GCPW) in CST as show in "Figure. 9". We used different thicknesses for the trace and the gap between the trace and ground for a 12Mil thick substrate made of Rogers Core RO4003C. The s-parameters and the characteristic impedance of the GCPW are in "Figure. 10"and "Figure. 11". Clearly the s-parameters are as desired with lower reflection loss and the impedance is also matched to that of the IC.

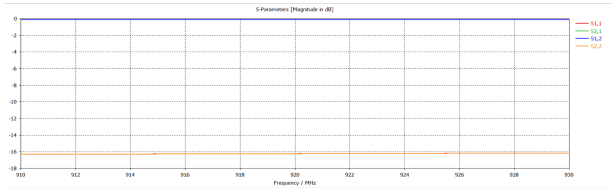


Figure 10. s-parameters for the coplanar waveguide.

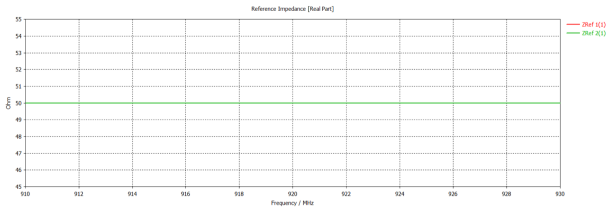


Figure 11. Characteristic impedance of the coplanar waveguide.

6. MEASUREMENTS AND DISCUSSION

The RFID reader speedway R420 rain RFID reader [29] was used in our experiments. The reader is connected to the antennas through a low loss four channel splitter. i.e., the reader output is split euqally among the four channels. each of these phase shifter output is connected to the RFID reader antennas through a phase shifter. As shown in “Figure. 9”, the PCB waveguide is fabricated and it’s s-parameters are shown in “Figure. 12”. It can be clearly seen that the losses due to reflection are very similar to the ones in simulated results in “Figure. 10”. Four such phase shifter boards were built with each phase shifter board excited with a separate voltage source. Based on the (7) and (10), each RFID reader antenna is excited with a certain phase, such that the phase difference between all consecutive elements is same.

To understand how to excite each element we can alternatively consider the voltage difference between consecutive antenna elements is the same as the voltage and phase relation is linear. This implies that for the given minimum phase shift 85^0 and max phase

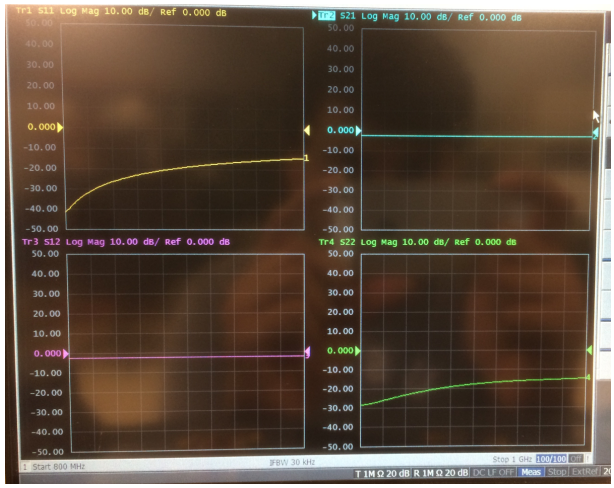


Figure 12. s-parameters of the fabricated coplanar waveguide.

shift 105^0, the maximum achievable phase difference for two elements on the extreme ends of a antenna array is 20^0. Also, the voltage difference of 12 volts between the two reader antennas yields maximum phase shift of 20^0. In case of more than two element the maximum achievable phase shift between any consecutive reader antenna elements is about

$$\left[\frac{20}{N-1}\right]^0 \tag{12}$$

Also, the maximum voltage difference between any consecutive antenna elements is

$$\left[\frac{12}{N-1}\right]\mathrm{V} \tag{13}$$

where N is the number of elements.

Therefore, to steer the beam for all possible directions, the voltage difference between every consecutive reader antennas need to be varied from 0V to Max voltage difference given by (13).

Figure 13. Linear RFID reader Antenna array with 4 elements.

6.1. EXPERIMENTAL SETUP AND RESULTS

The experimental setup consists of three double slit RFID antenna tags as shown in "Figure. 15". The thicknesses of these tags were measured for various thicknesses 3, 4, 5 and 6 mm. The characteristic impedance was unaltered. But for 1 and 2 mm, the tags need to modified by changing the slot lengths. This however resulted in the antenna tags being unusable due to conducting surface being too close to the ground plane. These tags are placed in the room as shown in the "Figure. 16" which is left with respect to the array and "Figure. 17" to the right of the array. The serial number of the RFID antenna ICs end with A4A3, 1223 and E053 placed consecutively from left to right at 30,30 and 15 feet respectively from the array . The linear and planar array as shown in "Figure. 13" and "Figure. 14" respectively. The substrate thicknesses used for these tags were 5 and 6 mm.

In this study, we consider linear array with two elements. Next we change the voltage difference between the two reader antennas from 0V to 12V. We observe if any of the tag is successfully read using the impinj multiread software [30]. When the voltage difference is 0V, we observe the tag 1223 is being read as seen in figure "Figure. 18". Similarly, for 8.7V, tag A4A3 is read indicating the beam pointing about 45^0 to the left of the array and for 11.7V, tag E053 is being read that is 45^0 to the right of the array. The tag ending with serial number E093 was a unknown tag and can be ignored for the moment.

Figure 14. Planar RFID reader Antenna array with 4 elements.

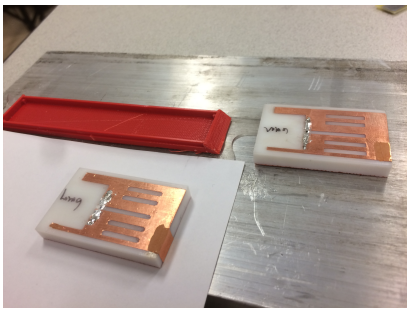


Figure 15. Double slit antenna tags.

Figure 16. The scenario of the room for experimental setup (left to the reader antenna array).

Figure 17. The scenario of the room for experimental setup (right to the reader antenna array).

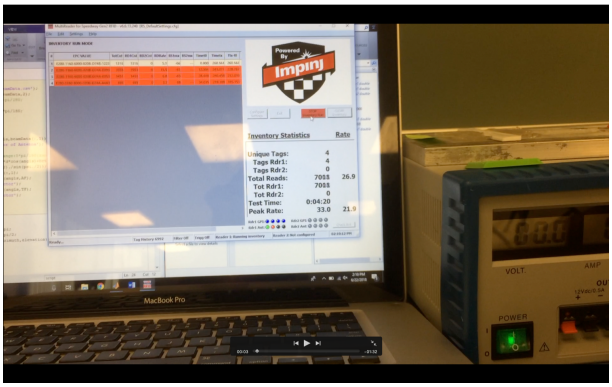


Figure 18. RFID tagg 1223 being read by the reader array.

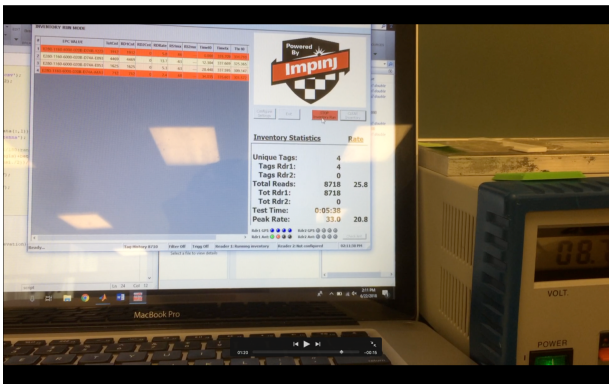


Figure 19. RFID tagg A4A3 being read by the reader array.

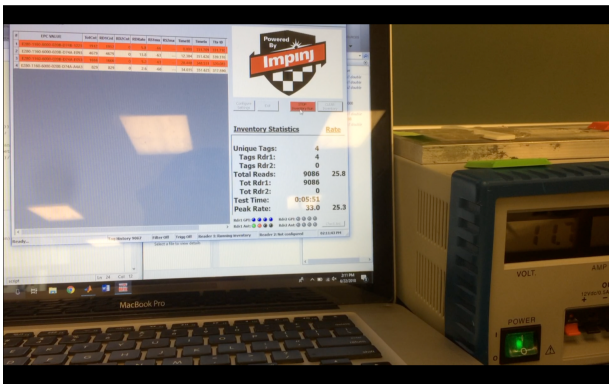


Figure 20. RFID tagg E053 being read by the reader array.

7. CONCLUSIONS

In this study, we designed a double slit RFID tag antenna for operation in metallic environments. Next we assembled a phased array using RFID reader antennas to achieve 30 feet (10 meters) read range. The phase difference between two consecutive antenna was changed by varying voltage from 0-12 Volts to steer the beam of the RFID reader antennas to a specific direction. Every time the beam is steered across all possible angles, a RFID tag antenna was read within the beam's vicinity. We had three RFID tag antennas for the measurements and all three were read by steering the main beam. The spacing between antennas yield more than one main lobe also known as grating lobes. Since it is beyond the scope of this study, more about this can be discussed in future measurements.

However, utilizing all the four antennas and further reducing thickness of the double slit tags upto 3 mm could give us nearly same range with smaller antennas. Also using the planar array would further facilitate the beam in both the azimuth and elevation plane. These measurements were omitted given the size constraints of this paper and will be further studied in future experiments.

REFERENCES

[1] Grabia, M. (2015, September). Assessment of effectiveness of the use of RFID technology in selected processes at DIY retail store. In 2015 IEEE International Conference on RFID Technology and Applications (RFID-TA) (pp. 134-140). IEEE.

[2] Arribas, P.C., Sanchez, P. and Aroca, C., 2010. RFID in Metal Environments: An Overview on Low (LF) and Ultra-Low (ULF) Frequency Systemsled. In Radio frequency identification fundamentals and applications design methods and solutions. InTech.

[3] A. Petrariu, V. Popa, V. Gaitan and I. Finis, "Test results for HF RFID Antenna system tuning in metal environment," Proceedings of the 13th International Carpathian Control Conference (ICCC), High Tatras, 2012, pp. 543-546.

[4] M. Periyasamy and R. Dhanasekaran, "Evaluation of performance of UHF passive RFID system in metal and liquid environment," 2015 International Conference on Communications and Signal Processing (ICCSP), Melmaruvathur, 2015, pp. 0414-0417.

[5] M. Grabia, "Assessment of effectiveness of the use of RFID technology in selected processes at DIY retail store," 2015 IEEE International Conference on RFID Technology and Applications (RFID-TA), Tokyo, 2015, pp. 134-140.

[6] K. D'hoe, T. Hamelinckx, J. Goemaere, N. Stevens, L. De Strycker and B. Nauwelaers, "Design and reliability evaluation of passive HF RFID systems in metal environments," 2011 IEEE International Conference on RFID-Technologies and Applications, Sitges, 2011, pp. 103-108. capitalized," J. Name Stand. Abbrev., in press.

[7] S. F. Wong and Y. Zheng, "The effect of metal noise factor to RFID location system," 2013 IEEE International Conference on Industrial Engineering and Engineering Management, Bangkok, 2013, pp. 310-314.

[8] A. J. Mercer, R. K. James, G. Bennett, P. Patel, C. Johnston and J. Cai, "Performance testing of RFID systems with RF-harsh materials," 2011 IEEE International Conference on RFID-Technologies and Applications, Sitges, 2011, pp. 537-543.

[9] Park, C.R. and Eom, K.H., 2011. RFID label tag design for metallic surface environments. Sensors, 11(1), pp.938-948.

[10] J. Alarcon, T. Deleruyelle, M. Egels and P. Pannier, "A low cost UHF RFID dipole antenna for metallic environments," 2012 IEEE International Conference on RFID-Technologies and Applications (RFID-TA), Nice, 2012, pp. 122-125.

[11] V. Ferro, A. Luz and A. Lucrécio, "Small long range UHF tag for metal applications," 2013 IEEE International Conference on RFID-Technologies and Applications (RFID-TA), Johor Bahru, 2013, pp. 1-6.

[12] Yu, B., Harackiewicz, F.J. and Lee, B., 2010. RFID tag antennas mountable on metallic platforms. In Radio Frequency Identification Fundamentals and Applications Design Methods and Solutions. InTech.

[13] J. B. Ferreira, Á. A. A. de Salles, G. Bulla and E. L. Rhod, "A new structure of UHF RFID tag antenna mountable on metallic surface using double slits," 2017 IEEE 9th Latin-American Conference on Communications (LATINCOM), Guatemala City, 2017, pp. 1-3.

[14] H. Kwon and B. Lee, "Compact slotted planar inverted-F RFID tag mountable on metallic objects," in Electronics Letters, vol. 41, no. 24, pp. 1308-1310, 24 Nov. 2005.

[15] Yu, B., Kim, S.J., Jung, B., Harackiewicz, F.J. and Lee, B., 2007. RFID tag antenna using two-shorted microstrip patches mountable on metallic objects. Microwave and optical technology letters, 49(2), pp.414-416.

[16] Choi, W., Kim, J.S., Bae, J.H., Choi, G.Y., Pyo, C.S. and Chae, J.S., 2008. RFID tag antenna coupled by shorted microstrip line for metallic surfaces. ETRI journal, 30(4), pp.597-599.

[17] Chen, H.D., Tsao, Y.H. and Kuo, C.Y., 2010. Low-profile radio frequency identification tag antenna using a trapezoid patch mountable on metallic surfaces. Microwave and Optical Technology Letters, 52(8), pp.1697-1700.

[18] A. Boryssenko and E. Boryssenko, "HF RFID antenna array for multi-tag enviroment," 2013 IEEE Antennas and Propagation Society International Symposium (APSURSI), Orlando, FL, 2013, pp. 1724-1725.

[19] S. E. Asl, M. T. Ghasr, M. Zawodniok and K. E. Robinson, "Preliminary study of mutual coupling effect on a passive RFID antenna array," 2013 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Minneapolis, MN, 2013, pp. 138-141.

[20] T. Huang, P. Pan and H. Hsu, "Adaptive Beam Steering Smart Antenna System for Ultra-High-Frequency Radio Frequency Identification Applications," 2012 International Symposium on Computer, Consumer and Control, Taichung, 2012, pp. 713-716.

[21] Van Veen, B.D. and Buckley, K.M., 1988. Beamforming: A versatile approach to spatial filtering. IEEE assp magazine, 5(2), pp.4-24.

[22] Balanis, C.A., 1992. Antenna theory: A review. Proceedings of the IEEE, 80(1), pp.7-23.

[23] https://www.lairdtech.com/products/s9-series-rfid

[24] https://en.wikipedia.org/wiki/Acrylonitrile_butadiene_styrene

[25] https://en.wikipedia.org/wiki/Thermoplastic_polyurethane

[26] https://ninjatek.com/ninjaflex/

[27] http://www.primitiveways.com/black_dye.html

[28] https://datasheet.octopart.com/PS088-315-Skyworks-Solutions-datasheet-8614930.pdf

[29] https://www.impinj.com/platform/connectivity/speedway-r420/

[30] https://support.impinj.com/hc/en-us/articles/360000595570-MultiReader-software-utility-end-of-availability

[31] https://www.mouser.com/datasheet/2/472/200247C-474412.pdf

IV. DNN-BASED RFID ANTENNA TAGS LOCALIZATION

Sohel J. Patel, Maciej Zawodniok
Department of Electrical & Computer Engineering
Missouri University of Science and Technology
Rolla, Missouri 65409–0050
Email: sjp432@mst.edu, mjzx9c@mst.edu

ABSTRACT

Radio frequency identification technology (RFID) is increasingly becoming an integral part of the Internet-of-Things (IoT). It offers different advantages including battery-free operation, small form-factor, and low cost. This makes the RFID an enticing technology for an indoor localization-based application and services. Geometry based localization approaches often achieve low accuracy due to errors introduced by a multipath propagation and interference in indoor environments. Many range-based algorithms assume that reader position is known in advance and there are carefully placed reference tags. In contrast, this paper presents a data driven localization methodology for direction-of-arrival (DOA) estimation using a deep neural network processing of signal captured with a reader antenna array. The proposed approach learns the complex mapping of the radio waves interactions in adverse metal environments based on received signal strength indicator (RSSI) values. The RSSI is captured while electrically steering a planar phased array through the area of interest. The proposed methodology is evaluated with multiple tags placed on metallic surfaces. Using readily available measurements, the proposed approach is able to achieve an average DOA error of 5.93 degrees.

1. BACKGROUND

1.1. SYSTEM ARCHITECTURE

The RFID system typically consists of a RFID antenna (antenna) that radiates energy and the RFID transponder (tag) that reflects this energy back to the antenna within a certain range. We use a 2×2 planar phased antenna array as seen in Figure. 1 and a tag is placed at an angle $(Azimuth, Elevation) = (\phi, \theta)$ with respect to the array. Conventional techniques assume these antennas as well as few reference tags placed at various known positions. These kind of approaches often require the calibration and assumptions well in advance. Hence, a system that can predict the DOA (ϕ, θ) of tags with minimal assumptions is desired.

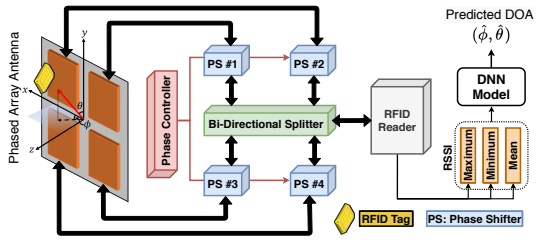


Figure 1. Block diagram of the phased array antenna for tag localization.

We propose the use of phased array antenna as shown in Figure. 1 with beam steering capability. The beam is steered to various positions and the RSSI values are measured each time. Each antenna is connected to the RFID reader through phase shifter and a bidirectional splitter. Each phase shifter is a voltage controlled bi-directional phase shifter IC independentely controlled by a microcontroller. Finally the maximum, minimum and mean RSSI measurements from the reader are computed. These values are used as training features for the neurla network to estimate the DOA for each tag.

1.2. PHASED ARRAY THEORY

The total electric field of the phased array can be analyzed as a product of individual element electric field $E_n = \hat{\mathbf{a}} j\eta \frac{kle^{-jkr}}{4\pi r} \cos\phi I_n$ and the array factor $AF = [e^{+j\frac{kd\cos\phi+\beta}{2}} + e^{-j\frac{kd\cos\phi+\beta}{2}}]$ for a RF signal impinging at an angle ϕ^0. Assuming far-field observations, the total electric field for a 2×2 antenna array is given by [1],

$$\begin{aligned} AF_{Planar} = & \sum_{n=1}^{2} I_n e^{jkd(n-1)\cos\theta\sin\phi+\beta_x} \\ & \times \sum_{m=1}^{2} I_m e^{jkd(m-1)\sin\theta\sin\phi+\beta_y} \end{aligned} \tag{1}$$

where d is the spacing between elements, (β_x,β_y) are the difference in the phase excitation along the x and y axis, I_m, I_n is the amplitude of individual antenna elements, $\hat{\mathbf{a}}$ is the radial unit vector, k is the wave-number, η is the intrinsic impedance of the medium and $j = \sqrt{-1}$, r is the distance from the source. where (ϕ, θ) are the azimuth and elevation angle of impinging RF signal. The phase excitation (β_x,β_y) may be altered to steer the beam to various positions.

1.3. PHASE SHIFTS FOR BEAM STEERING

The phase shifter used in our experiments is PS088-315 by skyworks solutions [2]. The phase could be shifted by upto 20^0. The phase shifts (β_x, β_y) should always satisfy the condition

$$|\beta_x|+|\beta_y| \leq 20^0 \tag{2}$$

Hence the beam can be steered only to certain positions restricted by (1). For example, if the phase shift along y axis is 15^0, then the phase shift along x-axis should be within $|20|^0 - |15|^0$ i.e. $(-5, 5)^0$.

We shift the phases (β_x, β_y) by a fixed value each time. I.e. the phase difference between any consecutive (β_x, β_y) should always be same. For phase increments of 1.8^0, the beam is steered to various positions for 10ms and RSSIs are captured each time. This leads to a total of 313 points as seen in Figure. 2(a). We start from $\beta_y = -20^0, -18.2^0 \ldots 18.2^0, 20^0$. At each phase shift β_y, we shift the phase β_x for all possible shifts satisfying (1).

2. CHALLENGES

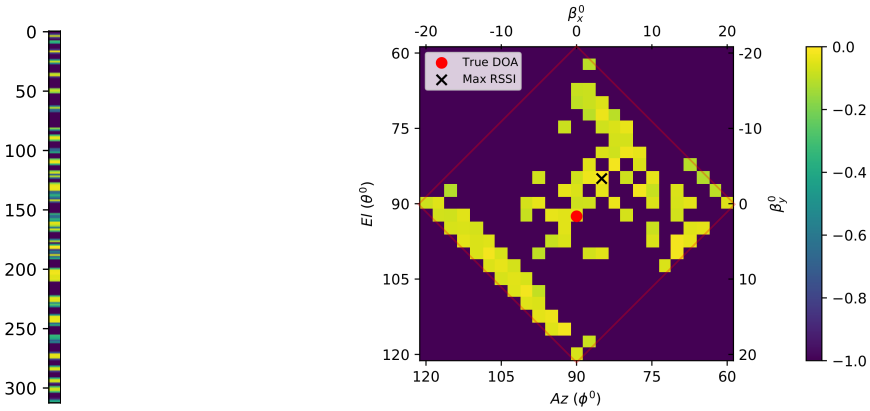
We briefly describe the motivation for using the phased antenna array and the physical RF parameter for DOA estimation. In our study, we mainly focus on the measured RSSI, which is a common indicator supported by most off the shelf RFID readers. The RSSI sensitivity to environmental factors and other challenges posed by the antenna array will be discussed in this section.

2.1. RSSI

The RSSI values measured as a function of (β_x, β_y) is as shown in Figure. 2 (a) and (b). In other words, we shoot the beam at various positions and capture RSSI each time. It may not be possible to achieve all phase shifts for (1) needs to be satisfied. Thus, all the possible phase shifts are indicated by the red box. The RSSi values captured for all points is difficult to visualize in vectorized form as shown in Figure. 2(a). The input to the DNN model is in this vectorized form, however. The RSSI as a function of (β_x, β_y) as seen in Figure. 2(b) is plotted for more detailed understanding. The RSSI measured for tags follow some pattern when placed at various positions. By learning these patterns the tag DOA can be estimated.

The phase shifts $(\beta_x, \beta_y) = (-20^0, 20^0)$ translates the azimuth (ϕ^0) and elevation (θ^0) from $(120^0, 60^0)$ and $(60^0, 120^0)$ respectively which is also the viewing angle of the array. Ideally, the maximum RSSI captured should reflect the position (ϕ, θ) of the tag. Clearly it is not the case in Figure. 2. This is a result of multiple reasons such as grating

lobes, antenna radiation patterns calibration and multipath effect among other reasons. If the antennas are considerably smaller in size, increasing the scale of the array to 3×3 or may be 4×4 might reduce the deviation between Max RSSI and true location. But several other factors would still affect the performance as discussed in Section. 2.2 and 2.3.



(a) (b)

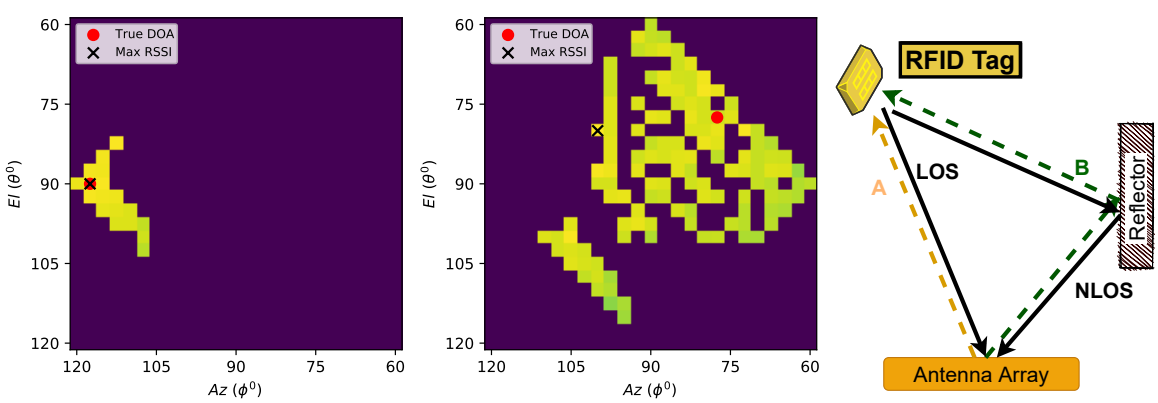
Figure 2. RSSI plots with true and maximum RSSI location for all possible phase shifts β_x, β_y for tag placed at $(\phi, \theta) = (90^0, 93^0)$. And the red box indicates all the phase shifts that possibly steers the beam. The bottom x-axis represent the azimuth angles as a result of β_x given along top x-axis. Similarly, the left y-axis represent steerable elevation angles for phase shift β_y given along right y-axis (a) Vectorized format which is used as input to DNN (b) Matrix format for better visualization.

2.2. MULTIPATH EFFECT

Consider Figure. 3 (a). The RSSI values captured are in the vicinity of the true tag location. Also, the maximum RSSI recorded is close to the tag location. This is desired ideal case. It also gives an idea as to where the tags might be located. Given these maximum RSSI values, the tag location may be estimated by translating the corresponding phase shifts to their azimuth and elevation (ϕ, θ) angles as discussed in Section 3.3. Hence giving the estimated tag location.

The maximum RSSI in Figure. 3 (b) is relatively far from the true tag location unlike Figure. 3 (a). In addition, RSSI values are measured from other beam directions as well. These is the result of multipath effect. For example, consider Figure. 3 (c), two different beam directions towards a tag are indicated by direct and reflected paths given by A and B respectively. Considering path A, the tag radiates energy back to the antenna along the same path as A. This is known as Line of Sight (LOS). The power radiated from the tag travels through the reflected path B as well. Also know as non line of light (NLOS). The energy received by the array has LOS signals much stronger than the NLOS ones. For the reflected signal from the tag is very low to be detected through the sidelobes.

When the beam is directed towards path B, the signal gets reflected from the object towards the tag. Since the NLOS signals suffer significant loss, much of the energy should be received by the LOS signals through path A. However, when the beam is radiating along path B, the power is mainly received through the side lobes. Although the tag is detected in both situations, the RSSI measured when beam is directed towards $'B'$ should be lower when compared to beam directed towards $'A'$. Clearly, it is not the case in Figure. 3 (b). This effect is known as spatial aliasing.

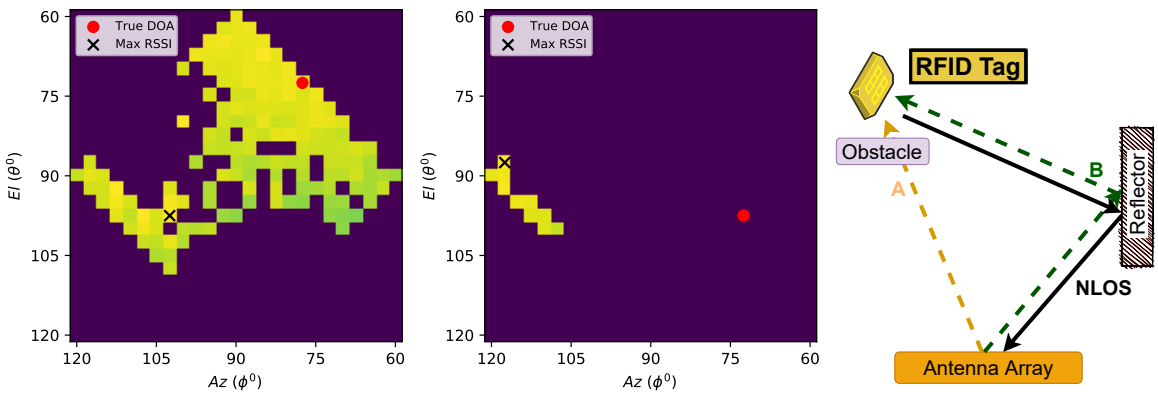


(a) (b) (c)

Figure 3. RSSI plots with true and maximum RSSI location for tag placed at (a)(ϕ, θ) = $(110^0, 90^0)$, (b)(ϕ, θ) = $(77^0, 76^0)$ and (c) Example for multipath effect resulting from strong reflections.

2.3. SPATIAL ALIASING

As indicated in Figure. 4 (a), the tag is located at $(\phi, \theta) \approx (78^0, 74^0)$. The maximum RSSI recorded is however at $(\phi, \theta) \approx (102^0, 99^0)$. As discussed in Section 2.2, given the beam is directed at multiple angles, it is fine for the tag to be read from those directions. The highest RSSI should be measured close to the true tag location which is not the case. This is due to the grating lobes.

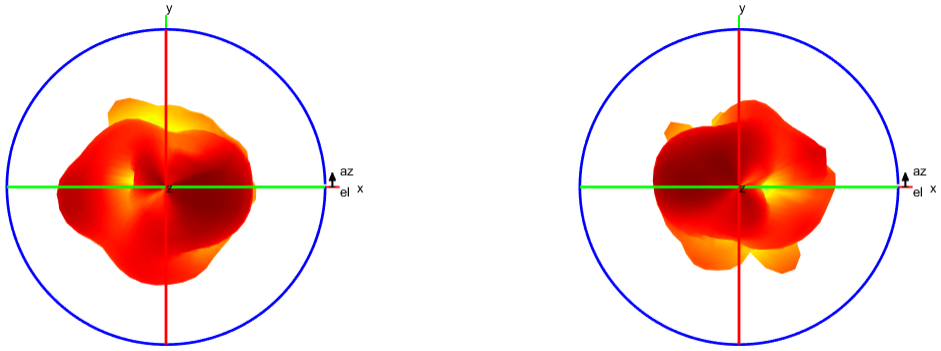


(a) (b) (c)

Figure 4. RSSI plots with true and maximum RSSI location for tag. (a)$(\phi, \theta) = (77^0, 68^0)$, (b)$(\phi, \theta) = (73^0, 97^0)$. (c) Example for NLOS transmission due to obstruction of tag.

To explain the grating lobes effects, we plot the phased array radiation patterns. The array radiation patterns with beam pointing at $(\phi, \theta) = (120^0, 90^0)$ *and* $(60^0, 90^0)$ are as indicated in Figure. 5 (a) and (b). The main lobe is indicated by darkest red-black portion. Such main lobes appear at more than one position. For example in Figure. 5 (a), although the main beam is pointed at $(\phi, \theta) = (120^0, 90^0)$, a small portion of equivalent energy is radiated at $(\phi, \theta) = (60^0, 90^0)$ as well. Similarly this effect can be observed in Figure. 5 (b).

Grating lobes appear when the spacing between consecutive antenna elements is greater than half the smallest wavelength. The highest for UHD RFID antennas is 928MHz with the smallest wavelength being $\approx$ 6 inches. Since the antenna dimensions are $\approx$ 10 inches, grating lobes are unavoidable in this particular case. Antennas with smaller aperture



(a) (b)

Figure 5. Planar array radiation patterns for beam steered with corresponding phase shifts and positions given by $(\beta_x, \beta_y), (\phi, \theta)$ respectively. (a) $(-20^0, 0^0), (120^0, 0^0)$ (b) $(20^0, 0^0), (60^0, 0^0)$.

may be designed to avoid spatial aliasing. This increases the over all cost of the system. The proposed approach is able to model this problem and estimate the tag position with considerable accuracy.

In addition to aforementioned challenges, when the tag is blocked by an obstacle LOS path may not possible. As seen in Figure. 4 (b), the tag located at $(\phi, \theta) = (71^0, 98^0)$ has RSSI measured for $\phi = 105^0 \rightarrow 120^0$. But the tag is not detected for any beam radiated nearby the tag. This situation is special case of multipath effect where the tag power radiated through LOS path $'A'$ is blocked by some obstacle. Therefore only NLOS transmission through path $'B'$ is possible as seen in Figure. 4 (c).

Although the NLOS based localization is not the main focus of this paper, we will briefly discuss few scenarios where tags were placed behind an obstacle and how did it affect the overall localization accuracy.

3. PROPOSED APPROACH

To overcome the problems mentioned heretofore, we propose a ML based approach to estimate tag location. We use DNNs in our experiment for tag localization. The input features to the DNN network will be the RSSI readings captured at all possible phase shifts as described in Section 3.3. Specifically, we calculate the maximum, minimum and mean for these RSSI values and feed into the DNN model as shown in Figure. 6 along with the ground truth DOA i.e. $\text{DOA}_{GT} = (\phi_{GT}, \theta_{GT})$.

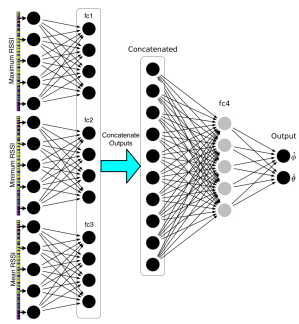


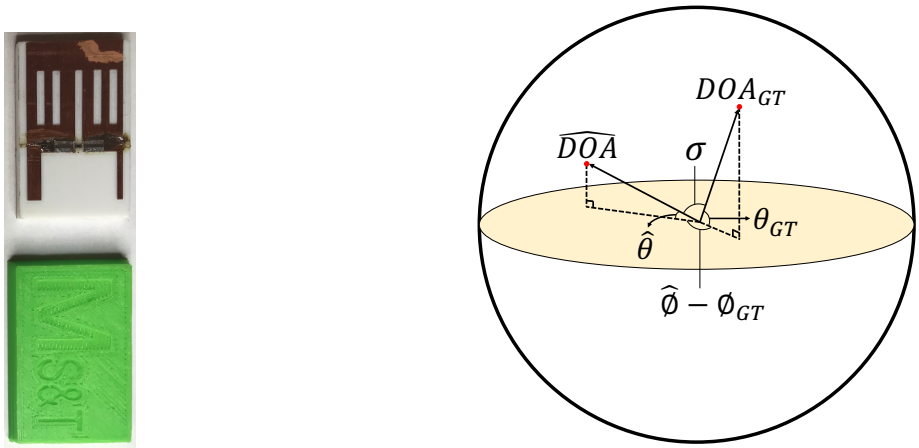
Figure 6. DNN model for RFID tag localization.

Once trained, these model is used to predict the unknown tag locations $\widehat{\text{DOA}} = (\hat{\phi}, \hat{\theta})$ given their RSSI values. As seen in Figure. 6, there are a total of 5 NNs with each layer named fc1, fc2, fc3, fc4 and output. The input maximum, minimum and mean RSSIs are fed into the layers fc1, fc2 and fc3 respectively with each layer connected to 50 output neurons. The three outputs are then concatenated to form a 150 dimension vector which serves as input to fc4 with output size 40. Finally, the fc4 output is fed into the output layer that yields two predicted DOA outputs i.e. $\widehat{\text{DOA}} = (\hat{\phi}, \hat{\theta})$. To avoid overfitting, we apply dropouts [31] in each layer with $p = 0.4$. I.e. in each NN layer, we randomly drop 40% of the neurons.

4. EXPERIMENTAL SETUP

4.1. DATASET

We use a total of 7 tags [32] designed to operate in harsh metallic environments as seen in Figure. 8 (a). Depending on the thicknesses of these tags their RSSI values captured by the reader may be different from one another. The thicknesses range from (3 – 6)mm. Further, the tags were placed $(\phi, \theta) = 60^0 - 120^0$ w.r.t the phased array randomly at distances ranging from (2 – 30) feet and heights (1 – 8) feet. The tags are randomly placed with closest distance between neighboring tag about 1cm. The movable furniture in the room were shifted a few times to analyze their effect on measured RSSI.



(a) (b)

Figure 7. (a) RFID tag used in experiments for metallic environments [32, 29] (b) Calculate the angle between two points on a sphere.

The tags are placed at 738 locations with about 313 RSSI values captured for 0.8^0 phase shifts (β_x, β_y) as described in Section 3.3. This yields a dataset totaling 738×7 = 5166 tag locations. We remove data collected for tag locations that were not detected for any combination of phase shifts. There were in total 2363 valid data samples for training and validating tag localization.

Figure 8. Office space used in experiments for RFID tag data collection. The tags are indicated by red boxes.

As for the tag location labelling, i.e. preparing the DOA_{GT} for each tag, a camera was placed at the center of the array. Depending on the distance of each tag with respect to the array the actual (ϕ_{GT}, θ_{GT}) were calculated from the tag position w.r.t the center of the image. We collect all the described data in our office space as shown in Figure. 8. Unlike sound source localization, where localization and tracking is done using wideband signal [34, 35], we perform in multiple narrow bands. Hence we focus on spot on localization rather than tracking tag movements and we leave tracking for future work.

5. RESULTS

The DNN is trained for the above described dataset. From the valid data samples collected, about 70% is used for training and the rest 30% is used for validating purposes. The DNN is trained for 100 epochs, using Adam optimizer, learning rate 10^{-4} and batch size 4. Hence the DNN as given in Figure. 6 is trained for tag localization using the training dataset such that the output loss is minimized. The loss function used is mean square error (MSE) given by:

$$\mathcal{L}_{mse} = \frac{1}{N}\sum_{n=1}^{N}(\hat{\theta} - \theta_{GT})^2 + (\hat{\phi} - \phi_{GT})^2 \tag{3}$$

where N is the batch size, (ϕ_{GT}, θ_{GT}) represent the ground truth azimuth and elevation angles and $(\hat{\phi}, \hat{\theta})$ are the predicted azimuth and elevation angles.

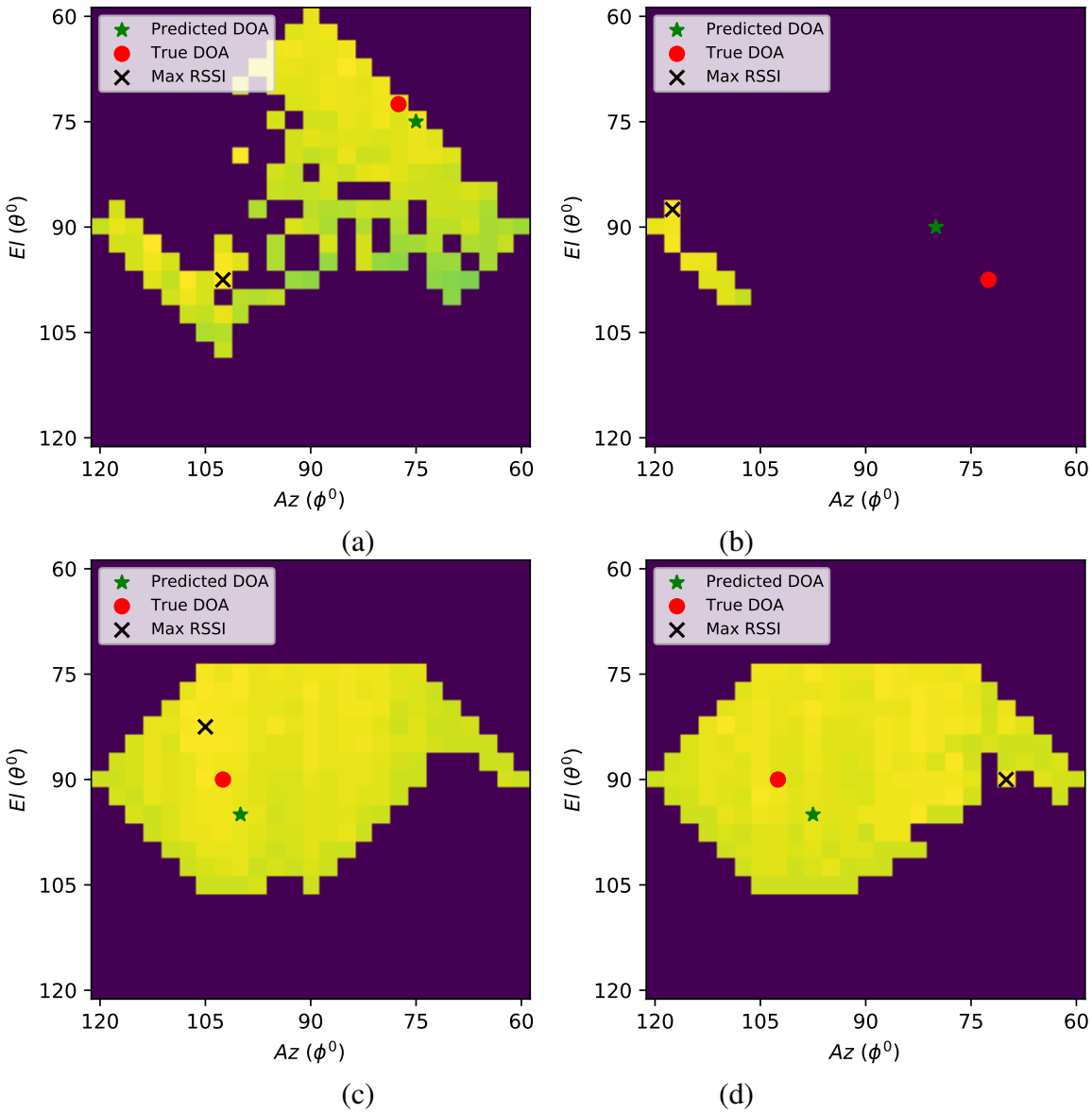


Figure 9. RSSI plots for predicted, true and maximum RSSI location for tag. (a)$(\phi, \theta) = (77^0, 68^0)$, (b)$(\phi, \theta) = (73^0, 97^0)$, (c)&(d)$(\phi, \theta) = (103^0, 90^0)$.

We use the angle between the predicted $\widehat{\text{DOA}} = (\hat{\phi}, \hat{\theta})$ and the ground truth $\text{DOA}_{GT} = (\phi_{GT}, \theta_{GT})$ as the DOA error metric. As seen in Figure. 8 (b) the DOA error is measured by calculating the distance between two points along the surface of the sphere given by:

$$\sigma_e = \cos^{-1}(\sin\hat{\theta}\sin\theta_{GT} + \cos\hat{\theta}\cos\theta_{GT}\cos(\phi_{GT} - \hat{\phi}))\frac{180}{\pi} \tag{4}$$

The metric as described in (4), is computed for every testing tag and the average σ_e is computed as seen Table 1. In addition to this we also decrease the resolution of RSSI values captured by increasing consecutive phase differences while steering beams as discussed in Section 3.3.

Table 1. DOA errors for different consecutive phase differences using basic DNN model. With left entries corresponding to localization achieved based on maximum RSSI value and right entries indicated by bold character the proposed DNN based location.

Phase Resolution	**#Points**	**Latency (s)**	**Average DOA Error** (°/Degrees)
0.8°	313	3.13	10.69/**6.30**
1.6°	85	0.85	17.72/**6.23**
2.4°	41	0.41	22.55/**5.93**
3.2°	25	0.25	25.13/**6.06**

The sensitivity of the proposed approach to the environmental factors discussed in Section. 2.2 and 2.3 can be observed in Figure. 9. Figure. 9 (a) and (b) represents the predicted DOA to Figure. 4 (a) and (b). It can be observed that the DOA error for grating lobes as seen in Figure. 9 (a) is quite satisfactory. Similarly, the predicted DOA for a tag detected through NLOS is given in Figure. 9 (b). To test the sensitivity of the proposed approach to environmental factors we moved the furniture in room while the tags remain at same position. One such scenario can be seen in Figure. 9(c) and (d). Although the tag is placed at same positions, the RSSI distribution is a little different after shifting furniture. Moreover the Max RSSI values is measured from a completely different location. The proposed approach however, generalizes well even with change in surroundings. The DNN is trained for 100 epochs, using Adam optimizer, learning rate 10^{-4} and batch size 4. Hence the DNN as given in Figure. 6 is trained for tag localization using the training dataset such that the output loss is minimized.

Along with 0.8^0 phase difference, we also analyze the performance when the RSSI values are measured by directing fewer beams for low latency applications. The phase increments are in order $0.8^0, 1.6^0, 2.4^0$ and 3.2^0 respectively. The total number of possible beam shifts are $313, 85, 41$ and 25 with latencies $3.13, 0.85, 0.41$ and 0.25 seconds respectively. As the consecutive phase differences increase, lower number of RSSI values are available to estimate tag DOA with corresponding accuracies-10.69^0, 17.72^0, 22.55^0 and 25.13^0. The DOA estimated using maximum RSSI value is clearly dependent on the number of RSSI values captured. The proposed DNN based method is quite consistent in each case and performs similar in all four cases as indicated in Table 1 with accuracies 6.30^0, 6.23^0, 5.93^0 and 6.06^0. Depending on the thicknesses of these tags their RSSI values captured by the reader may be different from one another. The thicknesses range from $(3-6)$mm. Further, the tags were placed $(\phi, \theta) = 60^0 - 120^0$ w.r.t the phased array randomly at distances ranging from $(2-30)$ feet and heights $(1-8)$ feet. The tags are randomly placed with closest distance between neighboring tag about 1cm. The movable furniture in the room were shifted a few times to analyze their effect on measured RSSI. Hence the DNN as given in Figure. 6 is trained for tag localization using the training dataset such that the output loss is minimized.

6. CONCLUSION

RFID tag DOA estimation is performed with minimal assumptions with lowest latency of 0.25 seconds and average error of $\approx 6.06^0$. The proposed approach is setup using commonly available off the shelf equipment and does not require multiple antennas and refrence tag locations to be known in advance. Also, we place tags on metal plates which makes environmental factors realistic and challenging. The proposed approach is able to model the multi path, spatial aliasing and NLOS effects with minimal assumptions. The future work would include localization in 3D probably for multiple indoor locations and cross validating their performances.

REFERENCES

[1] Balanis, C. A. (1992). Antenna theory: A review. Proceedings of the IEEE, 80(1), 7-23.

[2] https://www.mouser.com/datasheet/2/472/200247C-474412.pdf

[3] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research, 15(1), 1929-1958.

[4] Patel, S. J., & Zawodniok, M. (2019, May). Passive rfid tags for metallic environments using phased array reader antennas. In 2019 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) (pp. 1-6). IEEE.

[5] Zawodniok, M. J., Patel, S., Iyyer, N., Goel, K., & Phan, N. (2019). Rfid solution for tracking components life for rotor head assembly.

[6] Politis, A., Adavanne, S., & Virtanen, T. (2020). A dataset of reverberant spatial sound scenes with moving sources for sound event localization and detection. arXiv preprint arXiv:2006.01919.

[7] Patel, S., Zawodniok, M., & Benesty, J. (2020). Dcase 2020 task 3: A single stage fully convolutional neural network for sound source localization and detection. DCASE2020 Challenge.

V. 3D LOCALIZATION OF RFID ANTENNA TAGS USING CONVOLUTIONAL NEURAL NETWORKS

Sohel J. Patel, Maciej Zawodniok
Department of Electrical & Computer Engineering
Missouri University of Science and Technology
Rolla, Missouri 65409
Email: sjp432@mst.edu, mjzx9c@mst.edu

ABSTRACT

With the Internet of Things becoming widespread, there has been a rising demand for indoor location-based services. In recent trends, radio frequency identification has become an integral part of the production of IoT. Conventional methods use prior knowledge of antenna and tag positioning along with high-precision equipment capable of collecting phase or time-of-arrival data for robust estimation of three-dimensional location. In this work, we propose a three-dimensional localization method based on deep learning that relies on the phase and received signal strength indicator (RSSI) captured by steering beams to various locations using a phased array antenna. We evaluate the efficiency of this system by estimating three-dimensional location of 7 RFID tags mounted on metallic surfaces placed in a naturalistic environment. To evaluate the generalization of the proposed approach we crossvalidate the localization performance in different environments. The localization performance of the proposed approach is also tested on different formfactor of the RFID tag. With no prior information of either the tags or environment, the proposed system was able to achieve an average localization error as low as 1.33 cm with better system stability.

1. INTRODUCTION

Demand for indoor location-based services (LBS) have grown since rapid development of Internet-of-Things (IOT) especially indoor based. The common RF based localization techniques include global positioning system (GPS), wireless local area network (WLAN), Bluetooth, infrared, ultrasonic, ultra-wideband (UWB), ZigBee, radio frequency identification (RFID) etc [1]. Although bluetooth, ZigBee and RFID offer lower accuracy, their low cost far outweighs the advantages offered by other techniques for indoor LBS. RFID is useful in applications such as to track goods in the supply chain, in retail industry to inventory products, anti-theft management, authenticate products and wireless device configuration. For example, consider a huge ware house with a huge inventory and a specific product needs to be located. Scanning each product using traditional barcode system can be time consuming. This increases the overall time consumption if multiple products needs to be tracked. Instead RFID tags can be mounted on each product and read from farther distance while tracking multiple items simultaneously thus improving the business model.

We consider passive ultra high frequency (UHF) RFID tag based localization in our study operating at 915 MHz. For their low cost, higher read range $\approx 10m$, real time localization and no-contact communication they are ideal for IoT development. RFID have become promising key technology for the IoT. Expected to replace the traditional barcode system, UHF RFID could potentially serve the new generation of electronic tag. We will be referring to RFID reader antennas as antennas and RFID tag antennas as tags for simplicity hereafter.

Traditional techniques assume they have the location of the antennas and reference tags [2, 3] in order to triangulate and estimate the distance of test tags. These approaches dramatically increase the overall cost of the localization system as opposed to the requirement for low cost equipment. Moreover, the localization accuracy depends on the calibration of the setup thus increasing overall complexity.

To predict the tag position, RF signal parameters such as received signal strength indicator (RSSI), phase, time of arrival (TOA), time difference of arrival (TDOA)[4], phase difference of arrival (PDOA) are commonly used. RSSI and phase are easiest to capture among the mentioned parameters.

Typically 2D localization is performed using phase information. The phase can be obtained using complex demodulation. PDOA method is mainly divided into three types time domain TD-PDOA, frequency domain PDOA [5]-[6] and spatial domain PDOA. Spatial domain PDOA is useful in estimating DOA. Classical DOA algorithms include multiple signal classification (MUSIC) [7] that uses noise subspace, estimation of signal parameters via rotational invariance technique (ESPRIT) [8], combination of both ESPRIT-MUSIC [9] and Maximum Likelihood based estimation [10]. These are the most commonly used DOA estimation algorithms. DOA estimation has also been performed for various type of antenna arrays [11]-[12]. DOA measurements for RFID have been discussed in [13] with only the azimuth angle estimated.

A DL based technique is discussed in in [14] where a DNN is trained to estimate DOA i.e. azimuth and elevation angle. The DOA estimation offers 2D localization only. However, for applications requiring exact coordinates of the tag, 3D location of tags needs to be approximated. This paper extends the Deep Neural Network based DOA estimation proposed in [14] to estimate the accurate 3D location i.e. elevation, azimuth and radial distance of RFID tags using Convolutional Neural Network with RSSI and phase captured by steering beam at various positions.

To the best of our knowledge, this is a first time that an approach based on passive phase shifts on RFID antenna elements has been proposed as a viable tech for RFID technology. We further combine the beamsteering with deep learning based localization that requires to be trained only once and can generalize to any environment with different room parameters such as dimensions and furniture in the room. By delaying the RF signals on each antenna, the beam may be steered to desired position. Instead of modifying the

reader firmware for delaying signals which would be expensive, we adopt a beamsteering technique by delaying the signals using analog phase shifters. Moreover, this approach utilizes only single antenna port of the reader. This instrumentation approach can achieve beamsteering similar to software based approach with reasonable accuracy and inexpensive hardware thus advancing the state of the art for the measurement of tag location.

2. RELATED WORK

Many state of the art approaches have been proposed using fine-grained localization techniques to accurately locate the RFID tag position. Some of the successful works include [20, 21, 22] that adopt trilateration or hyperbolic based modeling to determine the location of tagged objects. Few techniques implement the concept of synthetic aperture radar (SAR) techniques [23, 24, 25, 26, 17, 19, 27] to achieve cm level localization accuracy. Tagogram [17] suppresses the RFID tag's phase shift using a differential augmented hologram technique. MobiTagbot [25] performs localization by studying the correlation with antenna in motion as a result of changing multipath reflections and carrier frequency channel. However, the reduction in uncertainties still remains a challenge where geometry-based models are not sufficient. ML based techniques have the capability of capturing vast probabilistic models in highly complex settings.

Several deep learning based approaches have been proposed [19, 18, 16, 14] to overcome limitations of the theoretical models. FaHo [19] employs a technique similar to SAR using fine grained joint hologram. The holograms are used as images to train the convolutional neural network (CNN). 3DLRA uses RSSI, phase and timestamps from 5 antennas to sort books in shelves by correlating with reference tags. PRDL uses RSSI and phase to find relative position of tags. The 2D location of tags is estimated by training a DNN solely using RSSI values captured by a 2×2 planar array [14] to achieve DOA error of about 5.93^0. The DL-based techniques discussed so far either operate at short range where multipath effects are not dominant or can only achieve 2D location of tags. In addition

Table 1. Comparisons of various RFID localization methods.

Positioning Method	**Antenna Movement Required**	**DL Based**	**Known Antenna Locations**	**range (feet)**	**3D Positioning**	**Positioning Accuracy**
LandMarc[2]			✓	**150**		**1**m
PRDL[16]	✓	✓		**1**		**-**
Tagogram[17]			✓	**13**	✓	**6.35**cm
3DLRA[18]		✓	✓	**2**	✓	**10.02**cm
FAHO[19]	✓	✓	✓	**3**	✓	**4.25**cm
DNN[14]		✓		**30**		**5.93**^{0}
Proposed		✓		**30**	✓	**1.33**cm

to limitations mentioned heretofore, the existing techniques have multiple RFID antennas placed at various positions [18] known in prior. In addition few techniques require moving parts [19] or only measure relative position [16]. Other RSSI based localization techniques such as [28] involve use of ZigBee which have battery powered modules placed at various known locations. We address these limitations in our study i.e. with no moving antennas, reference tags and measure the absolute position of passive tags that do not involve any battery powered parts thus lowering the overall costs.

We address problems discussed heretofore by proposing a ML based 3D localization which is enhanced to the previously proposed DOA estimation method using DNN [14]. The proposed approach requires that the data collected from the RFID tags be optimized for the CNN model such as reducing the overall latency for estimating the tag location and designing the hidden layers for the CNN. The localization problem is assumed to be performed on stationary tags and the antenna array for this paper. The CNN is trained to estimate accurate 3D location of tags using RSSI and phase values captured by means of beamforming similar to RADAR to achieve cm level accuracy. The comparison of the proposed approach can be seen in Table 1 with some state-of-the-art methods. The

required properties for the localization such as if the method required moving antenna, pre calibration,positioning type are compared. In addition to this, the range and accuracy are also provided.

3. SYSTEM DESIGN

This section briefly discusses the functioning of the proposed RFID system. Based on the given RFID system, we discuss how to capture the RSSI or phase in a specific format for training the CNN. We also discuss as to how to design the feature matrix.

3.1. SYSTEM ARCHITECTURE

In our experiments we use single reader for all 4 antennas. As seen in Figure. 1, the reader is connected to a bidirectional 4 way splitter. The splitter is then connected to antennas via a voltage controlled phase shifter. The voltage of each phase shifter is independently controlled by a microcontroller [29]. Both the reader and phase controller are operated by a computer. The computer then retrieves the data from the reader and gives the corresponding RSSI and phase values for all possible beams for each tag. The beam is steered to an appropriate azimuth ϕ^0 and elevation θ^0 angle by varying the phase shift along x and y directions given by β_x^0 and β_y^0 respectively.

The beam is as indicated in Figure. 1, a red vector pointing towards the tag originating from the center of the array. Once the features i.e. RSSI and phase are captured, they are preprocessed as will be discussed in Section 5.3 and fed into the CNN which then learns how to map the tag location to the input features by minimizing the loss between estimated and true location of the tag.

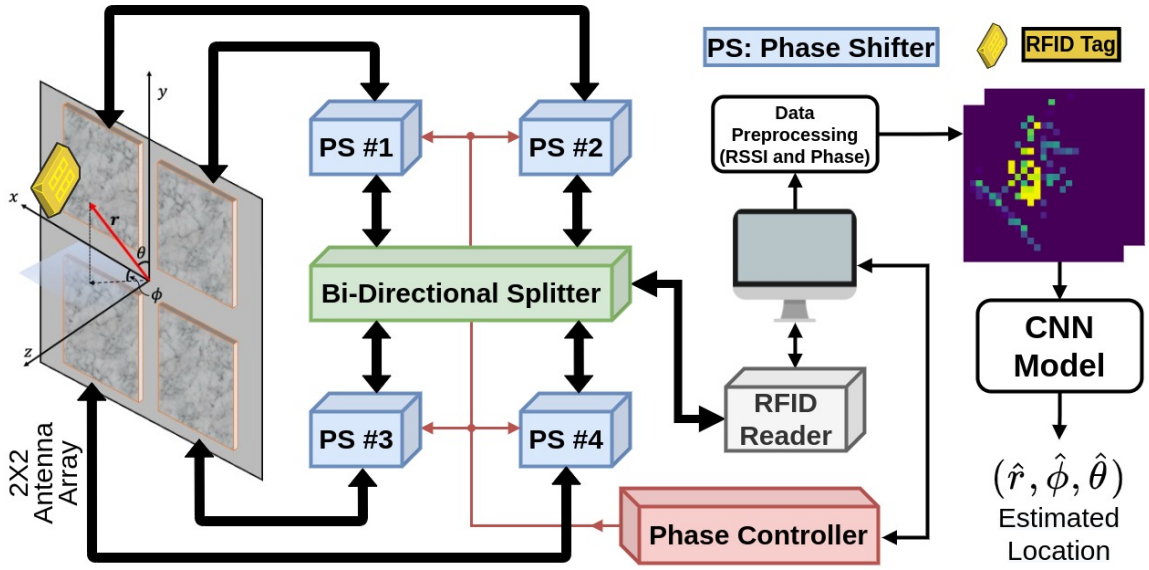


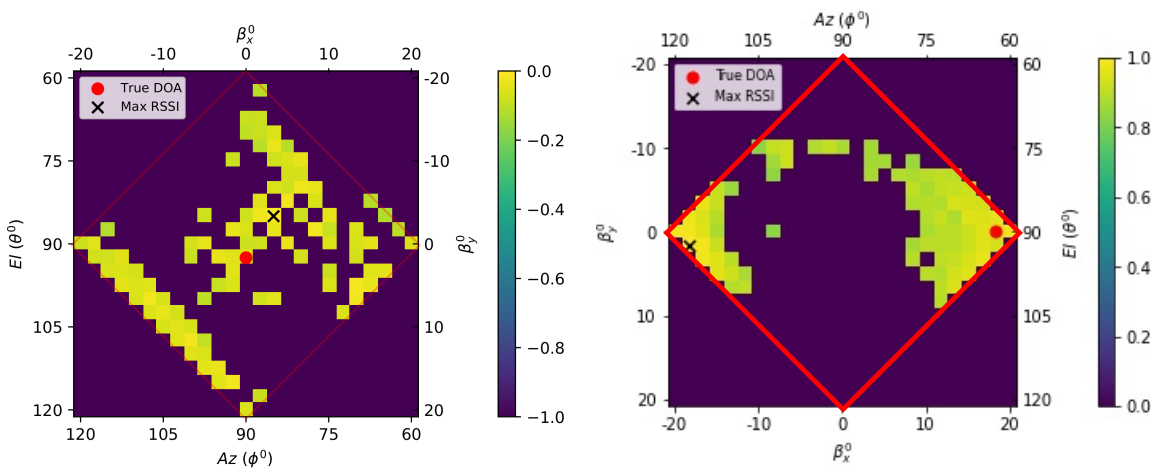
Figure 1. Block diagram of the phased array antenna for tag localization.

Unlike conventional approaches, where multiple antennas are placed at different locations connected to respective readers, our setup has all the antennas placed in a 2×2 array connected to a single reader as seen in Figure. 1. One advantage of having the antennas placed at different locations is the RSSI or phase captured by each antenna is unique to a specific location can be used for triangulating the tag coordinates.

3.2. FEATURES

This paper presents a modified approach where by steering the beam to different directions and capture the RSSI, phase and timestamp for each beam direction. Although, we use only the RSSi and phase as features to estimate the tag location in our experiments. Once we have enough features for different spatial locations, the RF signature pertaining to the tag locations can then be modelled. We show that this approach is both cost effective and can be practically implemented without any calibration or high precision equipment for tag localization.

For example, the RSSI and phase collected for all possible beam shifts is as shown in Figure. 2 (a) & (b) respectively. The lower x-axis and left y-axis indicate the the required phase shifts (β_x^0, β_y^0) respectively to point the beam to that direction. The angle at which the beam is directed is given by the azimuth ϕ^0 and the elevation θ^0 as indicated by the upper x-axis and right y-axis. The region outside the red box represents the phase shifts that cannot be achieved. This is not the limitation of the proposed approach but the limit imposed due to the phase shifter IC used that can shift the phase of the signal upto 20^0. If the hardware supports greater phase shifts, then the beam may be steered to wider angles.



(a) (b)

Figure 2. Input feature matrix plot for different beam directions for respective phase excitation along x and y directions given by (β_x^0, β_y^0) respectively. The tag position and maximum RSSI is indicated by red dot and $'\times'$ mark. The red box represents the possible look direction for the phased array. (a) RSSI. RSSI originally has a range (-120,0)dB and has been scaled to (-1,0). (b) Phase. Phase originally has a range $(0,2\pi)$ and has been scaled to (0,1).

3.3. PHASE SHIFTS RESOLUTION FOR BEAM STEERING

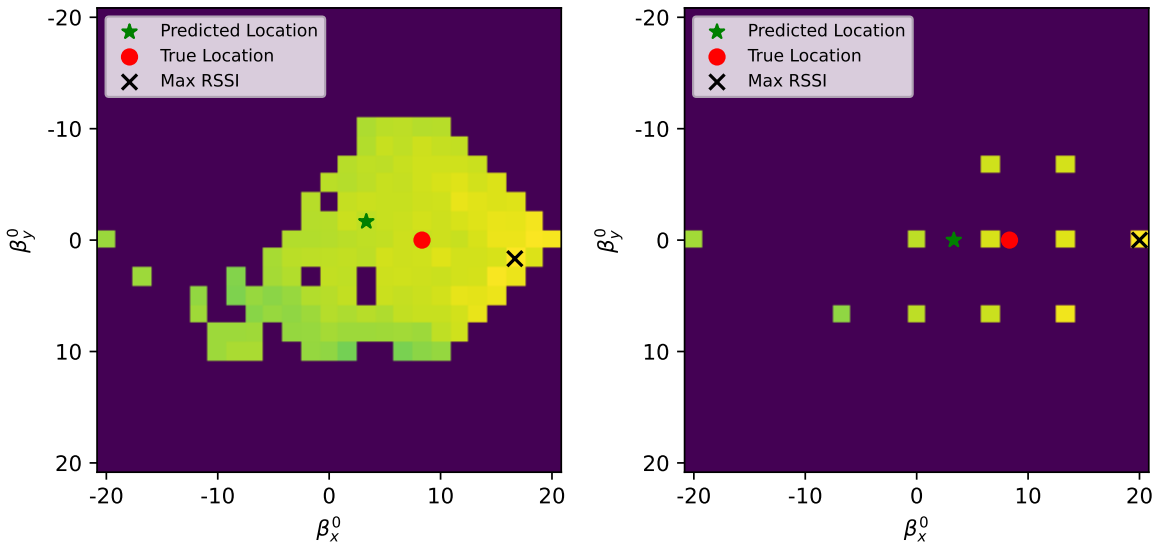
The phase shifts (β_x^0, β_y^0) as discussed in 3.2 need to satisfy the condition:

$$|\beta_x| + |\beta_y| \leq 20^0 \tag{1}$$

For simplicity, we set the step size to a fixed value for any consecutive phase shifts [14]. For example, the step size of 0.8^0 leads to 313 possible phase shifts as seen in Figure. 3 (a). For the beam pointed at 313 directions, the resolution of the measured features is quite dense. In Figure. 3 (b), the maximum RSSI matrix is shown for step size of 3.2^0. From this matrix, we only choose the features that would be measured if the step size was 3.2^0 that yields upto 25 beam locations. This can be seen in Figure. 3 (b).

The predicted location in both cases seems to be quite consistent and this was the case with all other samples as well. One possible explanation for this could be because of the beamwidth. As described in [14], the beamwidth of the resultant 2×2 antenna array is quite wide. As a result, even with greater beam directions, the change in the beam directions does not result in significant change in the magnitude of the captured parameters. One way to increase the resolution to achieve more variance among the captured parameters could be to increase the number of antenna elements. This would result in the increase of the overall cost of the system. However, increasing the step size beyond 3.2^0 would result in even fewer points thus reducing the localization error.

Hence, we reshape this matrix by taking the features that represent the step size of 3.2^0 and end up with a smaller 7×7 matrix instead of a 25×25 feature matrix. The 7×7 matrix is derived from the 25×25 feature matrix by only considering those values that correspond to phase increments by 3.2^0. For example, phase increments of 0.8^0 along x-axis and fixed phase along y-axis would yield 25 samples. Then to construct the 7×7 matrix we consider the values for every 4th sample starting from 1st value thus leading 7 values. this process is repeated for every phase shift along y-axis as well to yield a reduced 7×7 feature matrix. This approach has two advantages. It reduces the localization delay to about 250 ms [14] and reduces the neural training parameters and complexity as well. Therefore, we consider the features captured with step size of 3.2^0 for consecutive phase shifts.



(a) (b)

Figure 3. RSSI plots for a tag placed at $(\phi, \theta) = (76^0, 90^0)$ with true, maximum RSSI and CNN based tag location as a function of the different phase increment along x and y direction. The green marker indicates the predicted location of the tag. For simplicity and visualization purpose, we set the location to 2D in this example. (a) 0.8^0 (b) 3.2^0.

4. PROPOSED APPROACH

The goal of this study is to predict the absolute 3D location of the tag i.e. $\hat{\mathcal{L}} = (\hat{r}, \hat{\phi}, \hat{\theta})$ with respect to the antenna array. The proposed approach uses data driven machine learning technique. Therefore, using RSSI and phase as input features to the localization algorithm, the approximate location of the tag needs to be calculated.

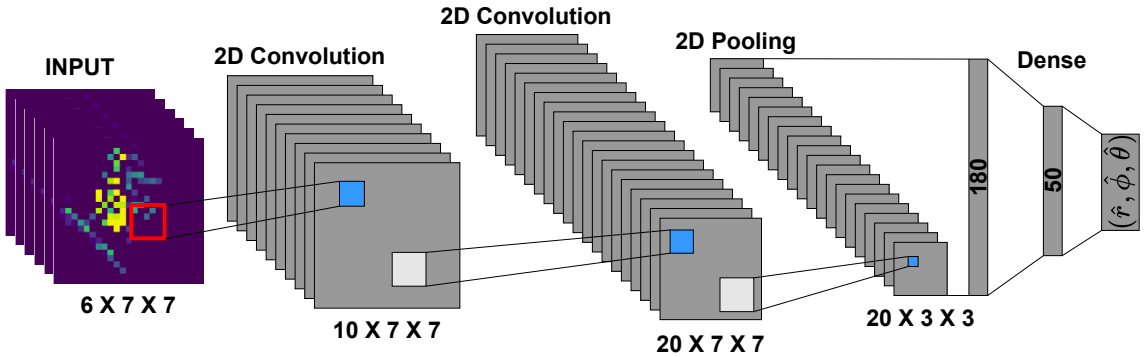


Figure 4. CNN model architecture used for tag localization.

We propose tag localization using a convolutional neural network (CNN) [30]. CNN are designed by neurons connectivity that resembles the organization of visual cortex. This approach resembles similar to as an image as viewed by human eye. CNNs require relatively lesser pre-processing compared to other image processing algorithms. Moreover, the total number of trainable parameters are also lower as compared to the baseline DNN. For example, the total number of trainable parameters in CNN model for the proposed approach are 11303 as opposed to 13663 learnable parameters in DNN. The CNN architecture needs to be optimized for the current setting where an RFID tag needs to be located. This requires choosing the right model parameters such as the desired hidden layers, number of convolutional and dense layers and their input/output dimensions. Moreover, The input features to the network i.e. phase and RSSI as captured by the antenna array needs to be arranged with a specific dimension such that it is compatible with model used.

Specifically the maximum, minimum and mean of the features for each beam are recorded. The CNN model architecture is as described in Figure. 4. Consider the feature matrix that is function of subsequent phase shifts β_x, β_y. The size of this matrix is $(6\times7\times7)$. Where for each β_x, β_y the beam is pointed for a specified amount of time. During that time the tag may be read any number of times with corresponding RSSI and phase values. We calculate maximum, minimum and mean for each feature RSSI and phase leading to 6 values which will serve as input features to our network. Each feature matrix is labelled with the true location of the corresponding tag $\mathcal{L} = (r, \phi, \theta)$.

We use this feature matrix as input to the CNN model to predict the output $\hat{\mathcal{L}} = (\hat{r}, \hat{\phi}, \hat{\theta})$. The CNN model is as described in Figure. 4. And the convolutional and hidden layers with parameters are tabulated in Table. 2. The input feature matrix with 3 channels is fed in to the first 2D convolutional layer (conv1) with kernel size (3×3) followed by relu activation and max pooling. From this 10 output channels are generated of size $(10\times7\times7)$. We apply padding to the input image to maintain same output image size.

Table 2. CNN layer parameters.

Stage	Output	Layers
input	6 x 7 x 7	max,min and mean {RSSI, Phase}
conv1	10 x 7 x 7	convolution 2D (3,3), padding=1
bn1	10 x 7 x 7	batch normalization 2D
conv2	20 x 7 x 7	convolution 2D (3,3), padding=1
pool2	20 x 3 x 3	max pool 2D (2,2)
bn2	20 x 3 x 3	batch normalization 2D
flatten	180	flatten output from bn2 layer
fc1	50	full connections layer 1, dropout=0.4
fc2	3	full connections layer 2, dropout=0.4
output	3	Distance, Azimuth, Elevation (r, ϕ, θ)
Trainable Parameters		11303

The above operations are repeated as described in Table. 2. The output from last pooling layer is flattened into a single vector of size 180. This is followed by dense connections also known as full connections that output values, i.e. 3 predicted $\hat{\mathcal{L}} = (\hat{r}, \hat{\phi}, \hat{\theta})$ in the final dense layer. In order to avoid overfitting we apply dropouts [31] in the dense connections. Dropout randomly selects neurons and ignores them while calculating gradients.

5. EXPERIMENTAL SETUP

This section focuses on the data collection for tag localization. The environment and lab spaces regarding the dataset being collected along with several other details will be discussed.

5.1. ENVIRONMENT

We use in total 3 different rooms including corridor for collecting data. Room 1 is divided into four different zones as seen in Figure. 5. This is done to validate the localization accuracy in the same room when the antenna array is placed at different locations. The

space shared by zone 1 and 2 has an overelap of ≈ 50%. However, zone 3 and 4 do not have any overlap with one another or with zone 1 or 2. The zones reflect the space where the tags were placed during data collection.

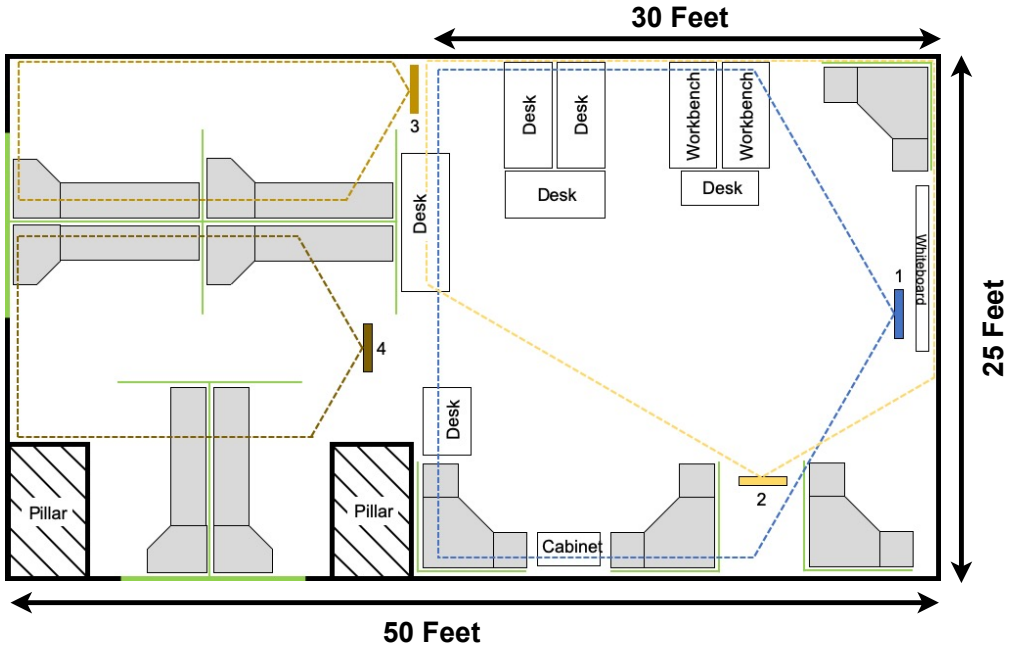


Figure 5. Schematic of room 1 with four different zones for data collection. Each zone, 1-4 is marked by a dashed line originating from the rectangle box with number (1-4) representing the antenna array placement. Each zone represents a specific set of tags collected read by the antenna array when placed at that position. The green lines represent panels that separate cabins and grey boxes represent desks with metal cabins.

In addition to the lab space described in Figure. 5, the data from the tags was collected in one more room (roon 2) and a corridor (room 3) as seen in Figure. 6(a) & (b). The antenna placement was at multiple spots. However, for simplicity the array and tag position as seen in Figure. 6(a) & (b) represent one such instant when measurements were in progress. Therefore, it can be assumed there are in total six different sets of data collected. The end goal of this paper is then to train six individual CNN models for each set of data. The localization error for tags is also tested in a large auditorium as given in Figure. 7. The

data collected in this room is strictly for testing purposed and not included while training to experiment the localization performance in large spaces which are different from the six type of rooms discussed so far.

(a) (b)

Figure 6. (a) Antenna array placed on left in Room 2 $(25 \times 30) feet$ with tags mounted on metal plates on the right and (b) Antenna array placed in corridor (Room 3) with a width of $15 feet$.

5.2. DATASET

Total 7 tags [32],[33] are used in our experiments as shown in Figure .8 (a). These tags were custom manufactured in the lab and soldered with Monza R6 energy harvesting IC. These tags also have different thicknesses 3,4,5 and 6 mm. There were in total 1 - 3mm, 2 - 4mm, 2 - 5mm and 2 - 6mm tags. Previously these tags were demonstrated to work in adverse metallic environments [29, 33] where most of the tags fail to perform. These tags have a read range upto $\approx$ 50 feet. Although, we limit the use of these tags to 30 feet in our measurements.

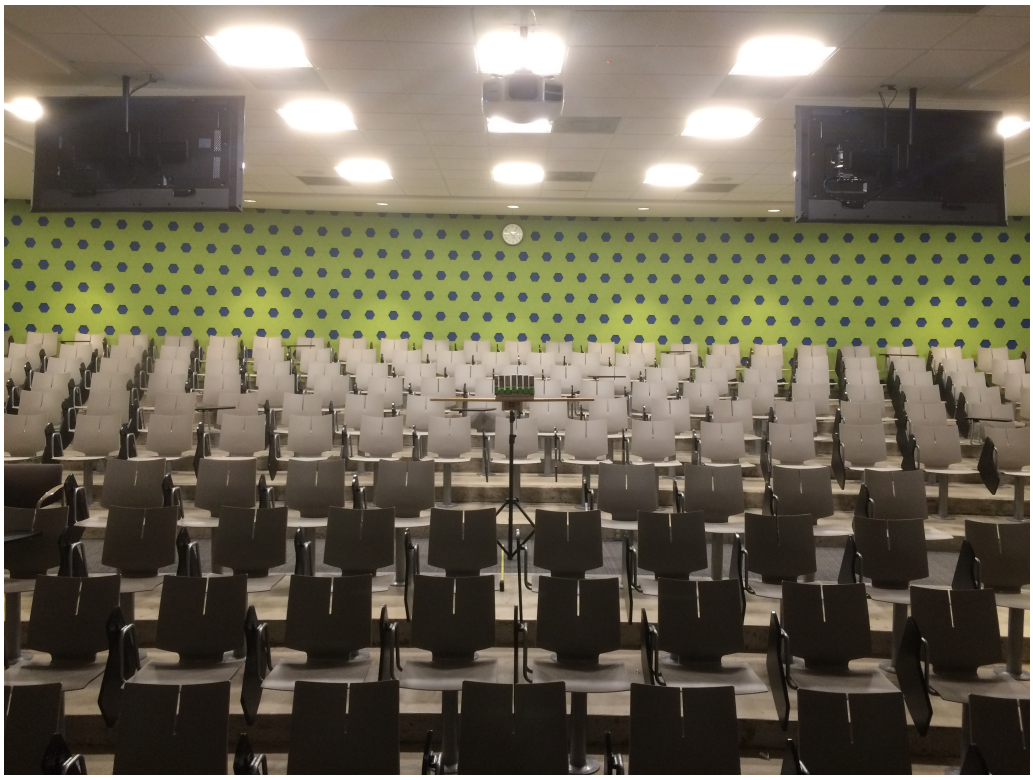
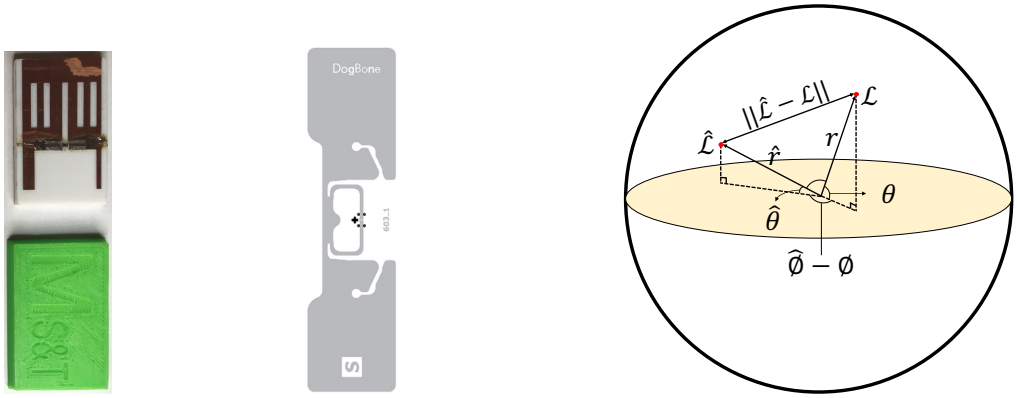


Figure 7. Schematic of a large auditorium used to test tags localization for large distances and in large room.



(a) (b) (c)

Figure 8. (a) RFID tag used in experiments for metallic environments [32, 29] (b) Smartrac Dogbone, Monza R6 RFID tag form factor (c) Calculate the angle between two points on a sphere.

For data collection, we use the university lab spaces as described in 5.1. The tags were placed within $(60-120)^0$ both azimuth and elevation which is the viewing angle of the phased array. The tags were also placed at different distances ranging from (2-30) feet from the antenna array. Also the tags were placed far and very close ($\approx 0.5cm$ in Figure .6(a)) from each other to capture the interference from neighbouring tags.

The tags were randomly placed at 2000 different locations. For each tag location, the beam was steered at all possible locations with a 3.2 degree phase shift along both x and y directions as described in 3.3. While satisfying condition as described in (1), there were a total of 25 features captured for each run. These features values are embedded into a $7 \times 7 = 49$ matrix as seen in Figure. 3 (b). For those beam directions where the tags were not detected by the beam are filled with zeros.

The other entries in the matrix which correspond to the phase shifts where the condition (1) is not satisfied are filled with zeros. The feature matrix can be constructed by ignoring these values. But we filled them with zeros to make them compatible for the CNN network to recognize patterns.

As for labelling the location of the tags, we place a camera exactly at the center of the antenna array and capture the images and recorded their distance perpendicular to the array. The camera requires pre-calibration in order to extract the exact coordinates of the tags. Although this technique would make the data collection much faster, it increases the over all costs. In order to address this issue we measure perpendicular distance and the height of the tags manually with simple measuring tools such as measuring tape with markers placed on the floor. The height of the tags from the floor can be used to measure the elevation angle θ. Using the perpendicular distance, we use camera to calculate the azimuth angle ϕ with respect to the array. Once we have the azimuth angle ϕ, the radial distance r can be calculated using the perpendicular distance and azimuth angle ϕ by solving simple trigonometric equations. Ideally we would want to use solutions like OptiTrack motion capture which have been previously used for sound localization purposes [34, 35].

OptiTrack give a very accurate 3D location of any object using reflectors that have been calibrated using a reference point. However, in our case since the tags are stationary the current approach works just fine.

5.3. DATA PREPROCESSING

The features RSSI theoretically ranges from $(-90, 0)$ and phase from $(0, 2\pi)$. The neural network train well when the inputs are within $(-1, 1)$. The normalized phase (P_s) can be normalized by dividing measured phase P with 2π:

$$P_s = \frac{P}{2\pi} \tag{2}$$

However, the feature matrix might have lot of zeros at positions where the tag is not read. For example in a 7×7 matrix, if the tag is read for only one beam position, then the matrix with 49 points has only one non-zero value. Usually the features are scaled by making the feature matrix zero mean and unit variance. But this results in lot of non-zero values. Hence we normalize the features by scaling and translating such that the maximum absolute value of each feature in each matrix to be 1 thus preserving sparsity.

5.4. TRAINING AND EVALUATION PROCEDURE

The data collected using 7 tags were used in our experiments. There were a total of 2000 positions where the tags were placed. Theoretically, a total of $2000 \times 7 = 14000$ samples are available for training and evaluation. As described in 5.2, for certain phase shifts the tags were not read. And for corresponding phase shifts along x and y directions we filled the matrix with zeros. However, there were certain cases where the tags were not read for any combination of phase shifts. This results in a complete zero matrix.

Table 3. 3D localization errors given in (cms) for 6 different environments cross validated with each other using only RSSI as training features with train/test ratio split 50%/50%. With left entries indicated by bold character corresponding to localization achieved based on proposed approach and right entries based on DNN based localization [14]. Each column represents a environment on which a model was trained and the corresponding localization errors for environment given in the first column.

Localization Error (cms)		**Room 1**				**Room 2**	**Room 3**
		Zone 1	**Zone 2**	**Zone 3**	**Zone 4**		
Room 1	**Zone 1**	**2.64**/4.23	**2.83**/4.32	**2.91**/3.97	**2.98**/4.16	**2.53**/4.32	**2.54**/4.30
	Zone 2	**2.16**/4.26	**2.22**/4.35	**1.41**/4.00	**1.98**/4.19	**2.16**/4.35	**2.12**/4.33
	Zone 3	**1.66**/4.27	**1.71**/4.37	**1.91**/4.02	**1.74**/4.21	**1.58**/4.36	**1.58**/4.34
	Zone 4	**4.02**/4.15	**2.82**/4.25	**2.81**/3.89	**2.43**/4.08	**3.83**/4.24	**3.84**/4.22
Room 2		**3.03**/3.99	**3.49**/4.09	**2.49**/3.74	**2.96**/3.93	**3.04**/4.08	**2.89**/4.06
Room 3		**1.77**/4.23	**1.76**/4.33	**1.38**/3.97	**1.64**/4.16	**1.82**/4.32	**1.77**/4.30
Avg=2.40/4.17		**2.54**/4.18	**2.47**/4.28	**2.15**/3.93	**2.28**/4.12	**2.49**/4.27	**2.45**/4.25

Such matrices with all zeros are removed during preprocessing stage prior to using them in our training and testing purposes. For they are essentially outliers and the model does not learn anything. After cleaning the data we are left with 9394 data samples. Given the 9394 samples collected for 7 tags in 6 different locations, the number of samples for each tag in each location on average are about $9394/(6 \times 7) \approx 224$. From these samples, we separate 50% of samples of each tag for training and the rest 50% for testing purposes. I.e. 112 samples for training and 112 for testing. Hence, it can be assumed that the tags are spaced over wide area id not very dense. DL based techniques usually depend on the amount of data for higher accuracy. This requires that a good amount of time is dedicated to collecting and processing data which maybe a few hours or maybe a full day. However, this needs to be done only once. Once the models are trained, they can then be used for inference and locating the tags in any other environment with no knowledge about the environment it

may be used in. The experimental setup is our research lab as seen in Figure. 5 & 6. The tags can be seen in Figure. 6 (a) mounted on metal plates and covered with green shields as shown in Figure. 8 (a).

Table 4. 3D localization errors given in (cms) for 6 different environments cross validated with each other using phase and RSSI as training features with train/test ratio split 50%/50%. With left entries indicated by bold character corresponding to localization achieved based on proposed approach and right entries based on DNN based localization [14]. Each column represents a environment on which a model was trained and the corresponding localization errors for environment given in the first column.

Localization Error (cms)		**Room 1**				**Room 2**	**Room 3**
		Zone 1	**Zone 2**	**Zone 3**	**Zone 4**		
Room 1	**Zone 1**	**1.38**/2.03	**1.38**/2.11	**1.33**/1.81	**1.31**/2.00	**1.38**/2.11	**1.38**/2.10
	Zone 2	**1.33**/1.67	**1.38**/1.76	**1.38**/1.42	**1.3**/1.62	**1.32**/1.75	**1.31**/1.73
	Zone 3	**1.39**/1.38	**1.38**/1.47	**1.35**/1.15	**1.31**/1.31	**1.40**/1.46	**1.39**/1.43
	Zone 4	**1.34**/1.44	**1.35**/1.54	**1.31**/1.19	**1.29**/1.38	**1.34**/1.53	**1.33**/1.51
Room 2		**1.27**/2.03	**1.28**/2.11	**1.32**/1.86	**1.27**/2.02	**1.26**/2.12	**1.25**/2.12
Room 3		**1.38**/2.10	**1.37**/2.18	**1.34**/1.88	**1.29**/2.06	**1.37**/2.19	**1.36**/2.18
Avg=1.33/1.77		**1.34**/1.77	**1.35**/1.86	**1.33**/1.55	**1.29**/1.73	**1.34**/1.86	**1.33**/1.84

5.5. LOSS FUNCTION AND TRAINING PARAMETERS

We consider an mean squared error (MSE) as the loss function given by:

$$\mathcal{E} = \frac{1}{N}\sum_{n=1}^{N}(\hat{r} - r)^2 + (\hat{\theta} - \theta)^2 + (\hat{\phi} - \phi)^2 \tag{3}$$

where N is the batch size, (r, ϕ, θ) represent the ground truth azimuth and elevation angles and $(\hat{r}, \hat{\phi}, \hat{\theta})$ are the predicted azimuth and elevation angles.

The hyper parameters used for training is described below. The CNN is trained for 100 epochs using adam optimizer. The learning rate was set to 10^{-4}. The batch size fed to the network each time during training is 4.

We use the distance between the predicted $\hat{\mathcal{L}} = (r, \hat{\phi}, \hat{\theta})$ and the ground truth $\mathcal{L} = (r, \phi, \theta)$ as the localization error metric. As seen in Figure. 8 (C) the localization error is measured by calculating the distance between two vectors in spherical coordinates given by:

$$\begin{aligned} \mathcal{L}_e &= ||\mathcal{L} - \hat{\mathcal{L}}|| \\ &= \sqrt{r^2 + \hat{r}^2 - 2r\hat{r}[\sin(\theta)\sin(\hat{\theta})\cos(\phi - \hat{\phi}) + \cos(\theta)\cos(\hat{\theta})]} \end{aligned} \tag{4}$$

The metric as described in ((4)), is computed for every testing tag and the average σ_e is computed. In addition to this we also decrease the resolution of RSSI values captured by increasing consecutive phase differences while steering beams as discussed in Section 3.3.

6. RESULTS

The CNN model was trained on the training set as described in 5.4. The CNN model complexity is based on the number of convolution and dense layers. For a given network of internal dimensions as described in Table. 2, weights and biases across all layers are optimized such that the output loss in (3) is minimized. Multiple models are trained to learn the location of tag in each room or zone and cross validated to examine the generalization of the proposed approach. We compare the performance of proposed approach with slightly modified DNN based localization method [14]. The output of the DNN now has the radial distance of the tag along with the azimuth and elevation angles.

6.1. LOCALIZATION ERROR WITH ONLY RSSI FEATURES

Firstly, we evaluate the localization performance using only RSSI as features for training. The input feature matrix given in Table 2 is modified with $3 \times 7 \times 7$ matrix consisting of only maximum, minimum and mean RSSI values. The rest of the layers in Table 2 remain unchanged. The localization error using RSSI only features are tabulated

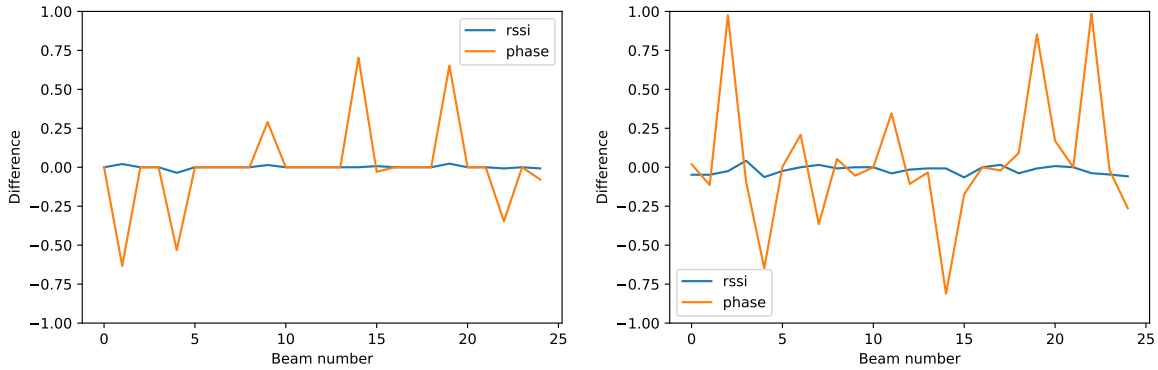
in Table 6.1. The results tabulated in Table 6.1 have 6 different environments in which the data was collected. The models were trained for each environment and tested on itself and other environment. For example, in first column, the results represent the performance of the model when trained on Room 1-Zone 1 and tested on all other six environments.

The average for each environment is given in the bottom row. The average error given in the lower right most cell is 2.40 cms and 4.17 cms for the proposed and DNN based approach. The proposed approach performs better than the baseline method [14] by about 1.77 cms. It can be seen that the zone 3 and 4 have comparatively lower localization errors of about 2.15 & 2.28 cms respectively compared to the other environments. This might be due to weaker multipath reflections for Zone 3 and 4 are quite small in size compared to other environments in which the data was collected where multipath were more prevalent. Hence, the performance of the proposed approach works well with just RSSI as training features.

6.2. LOCALIZATION ERROR WITH RSSI AND PHASE FEATURES

Next, we consider using both phase and RSSI as training features for estimating the tag position. The CNN model is trained as given in Table 2 with input feature dimensions $6 \times 7 \times 7$. The localization error using both phase and RSSI features are tabulated in Table 4.

Clearly, the phase information along with RSSI improves the tag localization. The average error for proposed approach and baseline method is 1.33 & 1.77 cms respectively. As described in Section 6.1, the average localization error for zone 3 and 4 given by **1.33**/1.55 & **1.29**/1.73 for proposed approach and baseline method respectively is lower compared to the other areas. Although the average improvement of the proposed approach over the baseline method is higher by 0.44 cms, the CNN based approach proposed in the paper as described in Section 4 has lower training parameters. This leads to faster estimation of tag locations when implementing in real time and faster training of the neural network.



(a) (b)

Figure 9. RSSI and phase difference between closely placed tags for about 25 beam positions. (a) 2cm (b) 5cm.

Compared to recent state-of-the-art 3DLRA [18] with 4.25 cm error, the proposed approach achieves 2.92 cms reduction in localization error. While the improvement may not be an order of magnitude, we only need to train the model once and results include in all other locations without retraining.

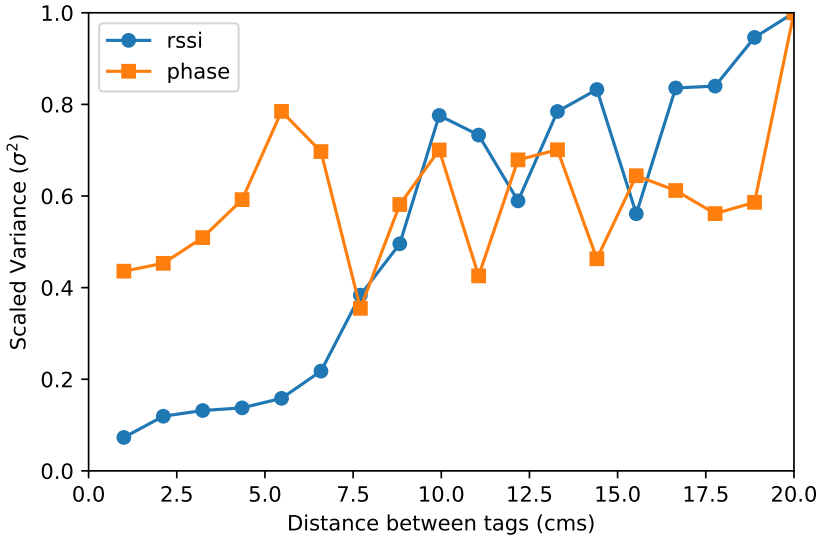


Figure 10. Scaled variance plots for RSSI and phase differences between tags placed at 1cm - 20 cm distance.

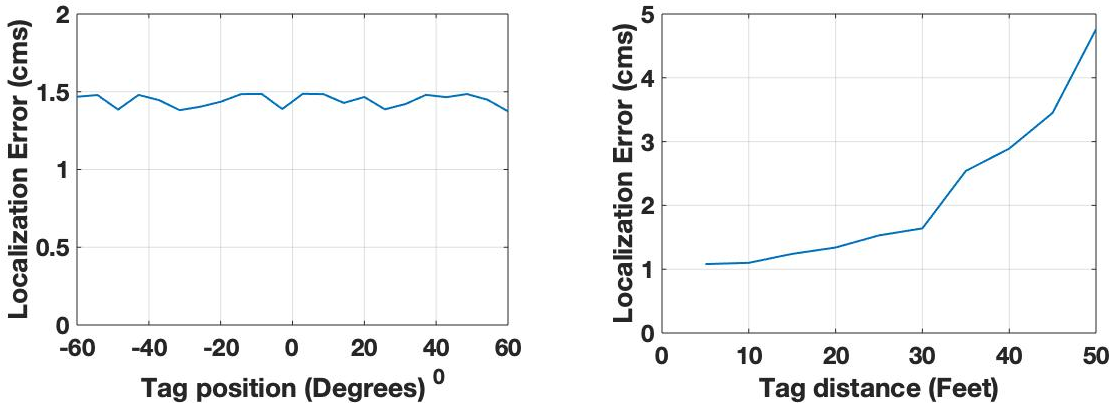
One possible reason for this could be that phase is varying much more faster and on larger scale as compared to RSSI. For example, we plot the difference between RSSI and phase among closely placed tags as seen in Figure. 9. It can be seen that when the tags are separated by about 2cm, the RSSI change is almost negligible. While the phase has large difference for certain beam positions as seen in Figure. 9 (a). Next, the difference in RSSI is a quite larger as seen in Figure. 9 (b) while phase is still varying largely. This change in RSSI keeps getting greater as the distance between tag increases.

To analyze this RSSI and phase difference closely between tags, we plot the variance for RSSI and phase difference for tags when separated by 1cm - 20cm as shown in Figure. 10. It can be seen that when tags are closely placed, the phase information is informative in terms of providing higher localization accuracy. And the RSSI is slowly varying thus leading to lower localization accuracy. This could potentially be the reason for higher accuracy when using phase along with RSSI as features for training the neural network.

6.3. LOCALIZATION ERROR IN LARGE SPACES

This section discusses the localization results for large spaces as described in Figure. 7. The localization experiment is performed on tags data collected for two different cases as discussed below. We retrained the models for this specific case using 60% of the training samples from the original dataset. Since we used 50% of the 9394 samples for training i.e. 4697 samples, we use 60% of these samples i.e. $\approx$ 2818 for training models for testing in large rooms.

In first case, the tag's radial distance $'r'$ from the array is fixed and the azimuth angle is varied from extreme left to extreme right i.e. $\phi = (60^0 - 120^0)$. The distance $'r'$ is fixed to $20feet$ from the array and the elevation is set to $\theta = 90^0$. The results for this case are plotted in Figure. 11(a). The localization error is quite consistent for all the given azimuth angles with a mean of $1.43cms$. For second case, we fix the azimuth angle ϕ of the tag from the array and vary the radial distance $r = (5 - 50)cms$. Although, the data is trained on tags



(a) (b)

Figure 11. Localization error for tags placed in large auditorium as described in Figure. 7. (a) 2cm (b) 5cm.

placed upto $30feet$ from the array, we also experiment what happens when the tags are placed upto $50feet$ from the array. The localization error is plotted in Figure. 11(b). It can be observed that for radial distance starting from $5feet$ and upto $30feet$, the localization error varies from $1.08cms$ to about $1.64cms$ respectively. However, the error drastically increases there after for $30feet$ through $50feet$ from $1.64cms$ to $\approx 5cms$ respectively.

Table 5. 3D localization errors given in (cms) for data collected on 5 different days in the Room 1-Zone 1 and tested on different environments using proposed approach. With left entries indicated by bold character corresponding to localization achieved based on phase and RSSI features and right entries based on RSSI features only.

$\mathcal{L}_e$(cms)	**Day 1**	**Day 2**	**Day 3**	**Day 4**	**Day 5**
Zone 1	**1.38**/2.03	**1.69**/1.96	**1.89**/2.45	**1.21**/2.22	**1.29**/2.12
Zone 2	**1.33**/1.67	**1.53**/1.58	**1.24**/1.78	**1.42**/1.63	**1.32**/1.76
Zone 3	**1.39**/1.38	**1.23**/1.35	**1.29**/1.45	**1.32**/1.47	**1.41**/1.35
Zone 4	**1.34**/1.44	**1.40**/1.39	**1.34**/1.32	**1.33**/2.32	**1.37**/1.57
Room 2	**1.27**/2.03	**1.73**/1.85	**1.31**/2.76	**1.38**/2.92	**1.27**/1.89
Room 3	**1.38**/2.10	**1.37**/2.24	**1.36**/1.52	**1.29**/1.98	**1.39**/1.73
Average	**1.34**/1.77	**1.49**/1.72	**1.40**/1.88	**1.32**/2.09	**1.34**/1.73

This could be because of firstly that the read rates for larger distance are much low even when placed at $30feet$ and in addition, the tags being tested were placed beyond $30feet$. Though beyond the scope of this paper, this increment in error can be mitigated by training the model on data collected by placing the tags at distances beyond 30 feet. Although, this approach might lead to lower localization error for larger distance, the drop in error might not be significant for the read-rates of tag at such larger distance might not be as good when compared to smaller distance from the array.

Table 6. 3D localization errors given in (cms) for RFID tag with same IC but different form factor as given in Figure. 8 (b). With left entries indicated by bold character corresponding to localization achieved based on phase and RSSI features and right entries based on RSSI features only.

$\mathcal{L}_e$ (cms)		**Dogbone Formfactor**
Room 1	**Zone 1**	**1.61**/2.22
	Zone 2	**1.71**/1.66
	Zone 3	**1.02**/1.29
	Zone 4	**1.4**/1.4
Room 2		**1.50**/1.69
Room 3		**1.53**/1.60
Average		**1.46**/1.64

6.4. LOCALIZATION ERROR FOR DIFFERENT DAYS

We analyze the localization when the tag location is to be determined on different days. We estimate the tag position using the pretrained models for data collected on previous days and test it on the data collected in Room 1-Zone 1. We collect the tag information for 5 different days over a period of month. This is done to incorporate the changes in room environment such as movement of people, changes in room properties such as reflections, different number of people in the room, change in dominant path etc.

For example few of such properties have been addressed in our previous work [14]. We show that the proposed approach can locate tags even when the dominant path is not available such as Non Line Of Sight (NLOS) and when the dominant path changes due to change in environment such different placement of furniture. For this paper, we test the performance on different days to test the sensitivity of the proposed approach to above mentioned real world problems.

Table 7. 3D localization errors given in (cms) for 6 different environments cross validated with each other using phase and RSSI as training features with train/test ratio split 40%/60%. With left entries indicated by bold character corresponding to localization achieved based on proposed approach and right entries based on DNN based localization [14]. Each column represents a environment on which a model was trained and the corresponding localization errors for environment given in the first column.

Localization Error (cms)		**Room 1**				**Room 2**	**Room 3**
		Zone 1	**Zone 2**	**Zone 3**	**Zone 4**		
Room 1	**Zone 1**	**1.73**/3.18	**1.52**/2.74	**2.83**/2.95	**1.63**/1.75	**1.94**/3.82	**2.01**/2.83
	Zone 2	**1.62**/2.67	**1.64**/3.82	**1.65**/2.84	**1.83**/3.79	**2.63**/2.56	**1.78**/3.87
	Zone 3	**2.62**/2.97	**1.53**/2.93	**1.96**/3.57	**1.53**/2.06	**1.52**/3.87	**1.69**/2.54
	Zone 4	**1.79**/2.12	**1.62**/2.04	**2.03**/3.84	**1.82**/3.84	**1.74**/3.12	**2.52**/2.93
Room 2		**1.54**/3.72	**1.68**/3.85	**1.86**/2.85	**1.73**/2.78	**1.41**/3.84	**2.91**/3.86
Room 3		**2.84**/2.83	**1.83**/2.17	**2.85**/3.52	**1.69**/3.57	**1.63**/4.67	**1.65**/3.93
Avg=1.91/3.17		**2.02**/2.91	**1.63**/2.92	**2.19**/3.26	**1.70**/2.96	**1.81**/3.64	**2.09**/3.32

The results are described in Table 5. We estimate the localization accuracy for all 6 different zones. The average localization error as given on the bottom row appear to be quite consistent across different days. The standard deviation for RSSI and phase features is about 0.06 while for that of RSSI only features is about 0.15. Hence the proposed approach is more robust to change in environments on different time and days.

Table 8. 3D localization errors given in (cms) for 6 different environments cross validated with each other using phase and RSSI as training features with train/test ratio split 70%/30%. With left entries indicated by bold character corresponding to localization achieved based on proposed approach and right entries based on DNN based localization [14]. Each column represents a environment on which a model was trained and the corresponding localization errors for environment given in the first column.

Localization Error (cms)		**Room 1**				**Room 2**	**Room 3**
		Zone 1	**Zone 2**	**Zone 3**	**Zone 4**		
Room 1	**Zone 1**	**1.31**/1.73	**1.35**/2.09	**1.27**/1.76	**1.29**/1.82	**1.37**/2.08	**1.31**/2.08
	Zone 2	**1.32**/1.46	**1.37**/1.65	**1.32**/1.29	**1.28**/1.61	**1.32**/1.73	**1.24**/1.59
	Zone 3	**1.25**/1.27	**1.37**/1.39	**1.34**/0.89	**1.30**/1.27	**1.39**/1.43	**1.32**/1.38
	Zone 4	**1.33**/1.42	**1.31**/1.33	**1.26**/1.04	**1.29**/1.26	**1.31**/1.36	**1.32**/1.39
Room 2		**1.31**/1.89	**1.32**/1.27	**1.27**/1.69	**1.25**/1.97	**1.25**/2.09	**1.27**/2.05
Room 3		**1.37**/2.04	**1.24**/1.34	**1.32**/1.79	**1.26**/1.98	**1.31**/2.04	**1.29**/2.11
Avg=1.30/1.62		**1.31**/1.63	**1.33**/1.51	**1.30**/1.41	**1.28**/1.65	**1.33**/1.79	**1.29**/1.77

6.5. LOCALIZATION ERROR FOR DOGBONE

The Monza R6 IC used so far in the custom tags was cut and embedded onto the tag as given in Figure. 8(a). We crossvalidate the performance of the custom made tags with the same tag but in a different form factor i.e. the Smartrac Dogbone RFID tag as given in Table 6. We use a total of three such tags for data collection. The average error for 6 different environments is about **1.46** cms, slightly higher than the tags used in the previous experiments.

6.6. LOCALIZATION ERROR FOR DIFFERENT TRAINING/TESTING DATA SPLIT

The results discussed in Table.6.1 and Table.4 are for train/test split for 50%/50%. We discuss the case where the train/test is split by 40%/60% and 70%/30%. By intuition, the localization error of the former split should be higher as there are lesser samples for

training than for testing and the model may not reach its full learning capability. And the latter should yield equivalent or better results for we have more training samples thus leading to higher quality model.

The results are tabulated in Table.7 for train/test ratio split 40%/60% and Table.8 for train/test ratio split 70%/30%. Clearly, the model performs worse when the available train data is lesser than testing data. The average localization error for this train/test split i.e. 40%/60% is $\approx 1.91cms$, thus increasing the error by 43.61%. Similarly, when the train/test split i.e. 70%/30%, the localization error is $1.30cms$ leading to 2.26% decrease in error.

It can be noted that although the increase in training data reduces the localization error, the difference in improvement is lower than the difference in performance drop when lesser training data is available.

7. CONCLUSION

The goal of this project was to accurately localize the RFID tags with minimal prior information about the position of antennas or tags. This was achieved by steering the beams upto 25 positions and capturing the RSSI values of the tags for training a CNN model to learn the location of tags. We achieved an average localization error of $\mathbf{2.40}cms$. This error represents the generalization of the proposed approach for various environments in which the data was collected i.e. the localization performance was evaluated for models trained on data collected from different workspaces. We saw a reduction of $\approx$ $\mathbf{1.77}cms$(**42**%) in localization error as compared to the baseline DNN model. The proposed methodology was experimented for lesser number of RSSI values ideal for real time tracking with latency as low as 0.25 seconds. When phase information is used along with RSSI the localization was further reduced by **44.5**% to about $\mathbf{1.33}cms$. In addition to that, the performance of the proposed system is evaluated on different days resulting in consistent performance.

REFERENCES

[1] Angrisani, L., Arpaia, P., & Gatti, D. (2017, September). Analysis of localization technologies for indoor environment. In 2017 IEEE International Workshop on Measurement and Networking (M&N) (pp. 1-5). IEEE.

[2] Ni, L. M., Liu, Y., Lau, Y. C., & Patil, A. P. (2003, March). LANDMARC: Indoor location sensing using active RFID. In Proceedings of the First IEEE International Conference on Pervasive Computing and Communications, 2003.(PerCom 2003). (pp. 407-415). IEEE.

[3] Hightower, J., Want, R., & Borriello, G. (2000). SpotON: An indoor 3D location sensing technology based on RF signal strength.

[4] Nikitin, P. V., Martinez, R., Ramamurthy, S., Leland, H., Spiess, G., & Rao, K. V. S. (2010, April). Phase based spatial identification of UHF RFID tags. In 2010 IEEE International Conference on RFID (IEEE RFID 2010) (pp. 102-109). IEEE.

[5] Li, X., Zhang, Y., & Amin, M. G. (2009, April). Multifrequency-based range estimation of RFID tags. In 2009 IEEE International Conference on RFID (pp. 147-154). IEEE.

[6] Zhang, Y., Ma, Y., Miao, X., Zhang, S., & Wang, B. (2018). Multipath mitigation algorithm for multifrequency-based ranging via convex relaxation in passive UHF RFID. IEEE Internet of Things Journal, 6(1), 744-752.

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

[8] Roy, R., & Kailath, T. (1989). ESPRIT-estimation of signal parameters via rotational invariance techniques. IEEE Transactions on acoustics, speech, and signal processing, 37(7), 984-995.

[9] Bencheikh, M. L., & Wang, Y. (2010). Joint DOD-DOA estimation using combined ESPRIT-MUSIC approach in MIMO radar. Electronics letters, 46(15), 1081-1083.

[10] Stoica, P., & Sharman, K. C. (1990). Maximum likelihood methods for direction-of-arrival estimation. IEEE Transactions on Acoustics, Speech, and Signal Processing, 38(7), 1132-1143.

[11] Gazzah, H., & Abed-Meraim, K. (2009). Optimum ambiguity-free directional and omnidirectional planar antenna arrays for DOA estimation. IEEE Transactions on signal processing, 57(10), 3942-3953.

[12] Shu, F., Qin, Y., Liu, T., Gui, L., Zhang, Y., Li, J., & Han, Z. (2018). Low-complexity and high-resolution DOA estimation for hybrid analog and digital massive MIMO receive array. IEEE Transactions on Communications, 66(6), 2487-2501.

[13] Zhou, J., Zhang, H., & Mo, L. (2011, May). Two-dimension localization of passive RFID tags using AOA estimation. In 2011 IEEE International Instrumentation and Measurement Technology Conference (pp. 1-5). IEEE.

[14] Patel, S. J., & Zawodniok, M. (2021, May). DNN-Based RFID Antenna Tags Localization. In 2021 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) (pp. 1-6). IEEE.

[15] Ni, L. M., Liu, Y., Lau, Y. C., & Patil, A. P. (2003, March). LANDMARC: Indoor location sensing using active RFID. In Proceedings of the First IEEE International Conference on Pervasive Computing and Communications, 2003.(PerCom 2003). (pp. 407-415). IEEE.

[16] Shen, L., Zhang, Q., Pang, J., Xu, H., & Li, P. (2019). PRDL: relative localization method of RFID tags via phase and RSSI based on deep learning. IEEE Access, 7, 20249-20261.

[17] Yang, L., Chen, Y., Li, X. Y., Xiao, C., Li, M., & Liu, Y. (2014, September). Tagoram: Real-time tracking of mobile RFID tags to high precision using COTS devices. In Proceedings of the 20th annual international conference on Mobile computing and networking (pp. 237-248).

[18] Cheng, S., Wang, S., Guan, W., Xu, H., & Li, P. (2020). 3DLRA: an RFID 3D indoor localization method based on deep learning. Sensors, 20(9), 2731.

[19] Xu, H., Wang, D., Zhao, R., & Zhang, Q. (2019, November). FaHo: deep learning enhanced holographic localization for RFID tags. In Proceedings of the 17th Conference on Embedded Networked Sensor Systems (pp. 351-363).

[20] Liu, J., Zhu, F., Wang, Y., Wang, X., Pan, Q., & Chen, L. (2017, May). RF-scanner: Shelf scanning with robot-assisted RFID systems. In IEEE INFOCOM 2017-IEEE Conference on Computer Communications (pp. 1-9). IEEE.

[21] Wang, J., Adib, F., Knepper, R., Katabi, D., & Rus, D. (2013, September). RF-compass: Robot object manipulation using RFIDs. In Proceedings of the 19th annual international conference on Mobile computing & networking (pp. 3-14).

[22] Pathak, P. H., Feng, X., Hu, P., & Mohapatra, P. (2015). Visible light communication, networking, and sensing: A survey, potential and challenges. IEEE communications surveys & tutorials, 17(4), 2047-2077.

[23] Miesen, R., Kirsch, F., & Vossiek, M. (2011, April). Holographic localization of passive UHF RFID transponders. In 2011 IEEE international conference on RFID (pp. 32-37). IEEE.

[24] Parr, A., Miesen, R., & Vossiek, M. (2013, April). Inverse SAR approach for localization of moving RFID tags. In 2013 IEEE International Conference on RFID (RFID) (pp. 104-109). IEEE.

[25] Shangguan, L., & Jamieson, K. (2016, June). The design and implementation of a mobile RFID tag sorting robot. In Proceedings of the 14th annual international conference on mobile systems, applications, and services (pp. 31-42).

[26] Wang, J., & Katabi, D. (2013, August). Dude, where's my card? RFID positioning that works with multipath and non-line of sight. In Proceedings of the ACM SIGCOMM 2013 conference on SIGCOMM (pp. 51-62).

[27] Qiu, L., Huang, Z., Wirström, N., & Voigt, T. (2016, May). 3DinSAR: Object 3D localization for indoor RFID applications. In 2016 IEEE International Conference on RFID (RFID) (pp. 1-8). IEEE.

[28] Bianchi, V., Ciampolini, P., & De Munari, I. (2018). RSSI-based indoor localization and identification for ZigBee wireless sensor networks in smart homes. IEEE Transactions on Instrumentation and Measurement, 68(2), 566-575.

[29] Zawodniok, M. J., Patel, S., Iyyer, N., Goel, K., & Phan, N. (2019). Rfid solution for tracking components life for rotor head assembly.

[30] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.

[31] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research, 15(1), 1929-1958.

[32] Patel, S. J., & Zawodniok, M. (2019, May). Passive rfid tags for metallic environments using phased array reader antennas. In 2019 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) (pp. 1-6). IEEE.

[33] Ferreira, J. B., de Salles, Á. A. A., Bulla, G., & Rhod, E. L. (2017, November). A new structure of UHF RFID tag antenna mountable on metallic surface using double slits. In 2017 IEEE 9th Latin-American Conference on Communications (LATINCOM) (pp. 1-3). IEEE.

[34] Politis, A., Adavanne, S., & Virtanen, T. (2020). A dataset of reverberant spatial sound scenes with moving sources for sound event localization and detection. arXiv preprint arXiv:2006.01919.

[35] Patel, S., Zawodniok, M., & Benesty, J. (2020). Dcase 2020 task 3: A single stage fully convolutional neural network for sound source localization and detection. DCASE2020 Challenge.

SECTION

2. A SINGLE STAGE FULLY CONVOLUTIONAL NEURAL NETWORK FOR SOUND SOURCE LOCALIZATION AND DETECTION

In this report, we present our approach for DCASE 2020 Challenge Task3: Sound event localization and detection. We use a single step training method using SELDNet like models but using fully convolutional architectures. We consider the joint optimization of both event detection and doa estimation. For the metrics that evaluate the performance of the model consider interdependence of both parameters performance unlike independent performance like DCASE 2019 challenge. We use all the sound event classes and corresponding cartesian co-ordinates for each class to create an image like label for reference and make this an image to image mapping problem. The best model could get DOA error of around 13.5^0 and error rate of 0.55.

Sound event Localization and detection has been an interesting topic for research for a long time. Previously it was a very challenging to achieve satisfactory performance mainly because of their implementation was based on pure signal processing algorithms. Recently with availability of comprehensive databases and computational resources several interesting performances have been observed. Researchers have proposed several deep learning algorithms that combined solve the localization and detection problem. However, most of the proposed approaches consider raw spectrograms as an input feature to neural networks. Such kind of features may be okay for application involving speech enhancement, dereverberation and few other problems where the networks learn some kind of patterns like formants, pitch etc.. Such features are however not applicable to sound events detection most of the time for they do not consider human speech as their only inputs.

A very recent implementation of SELD has been described in [2] where a Convolutional Recurrent Neural Network (CRNN) is trained using magnitude and phase spectrograms to predict active sound events and their location w.r.t the microphone array. In DCASE 2019 challenge [1], many research teams have proposed models with state of the art performances with reduced feature sizes. For example in [8] mel-scale spectrograms and phase transformed generalized cross correlation (GCC-PHAT) have been extracted from the spectrograms which contain the necessary and sufficient information for the model to learn patterns to predict active sound events class and their location.

Most of the previous approaches make use of ensemble models [19] and average them [8] in the end to further reduce the training error and avoid overfitting. This method has an advantage when running the evaluation dataset on these trained models to achieve reasonable performance. However, in our approach we do not train such ensemble models for they require lot of hyperparameters tuning.

2.1. FEATURE AND LABEL EXTRACTION

The TAU-NIGENS Spatial Sound Events 2020 dataset [29] consists of two recordings format: first order ambisonic (foa) and 4 channels from a microphone array (mic). We use both microphone array (MIC) and first order ambisonics (FOA) format in our experiments. The development dataset consists of a total of 600 recordings each one minute long sampled at 24000 Hz.

The labels represent the location of each sound source and their corresponding class for every 0.1s. Therefore, for each recording there are 600 labels. For hop length of 0.02 seconds and window length 0.04 seconds, a complex spectrogram of size 3000x512 can be derived. This means the labels represent ground truth for every 5 spectrogram frames. Processing the spectrogram this size can be time consuming and redundant.

For this kind of application it has been shown the instead of using raw spectrograms for training neural networks, we can extract useful features like mel spectrograms [8]. For both FOA and MIC dataset, the mel spectrograms are calculated. In addition to this the GCC-PHAT are also calculated for the mel bands. We use the 64 mel frequency bands as suggested in the baseline thus reducing the input feature mel-spectrogram to 3000x64 with a total of 17 channels (10 MIC and 7 FOA).

There are a total of 14 sound events with 2 or less sound events active at any given time. Hence the sound event detection (SED) Labels can be structured as a vector of length 14 for 14 classes with any particular sound event as active or not by a binary 0 or 1. Similarly, we have another 14 reference labels for each x, y and z co-ordinates where any particular active sound event is given its co-ordinates rest being zero. We use the cartesian co-ordinate system over spherical for all the reference labels lie between (-1,1). This way each corresponding label now has 56 labels of which first 14 are binary 0/1 SED labels and next 42 are x, y and z DOA labels.

since the hop length (0.02 s) is $1/5^{th}$ of the time for which labels have been provided (0.1 s), we upsample the labels by copy the labels 5 times to match the input feature size in time steps. For example, for every 300 input frames, there would be (300/5=60) reference labels. After upsampling, there are labels for each time frame. As a result, for input feature size of (300 x 64) the labels matrix size is (300 x 56). This can be considered as image to image mapping problem since all the reference labels are within (-1,1)

2.2. ARCHITECTURE

The baseline architecture consists of a convolutional recurrent neural network with 3 convolutional layers followed by bidirectional GRUs and Dense layers. The SED task is considered as classification problem and the DOA is considered regression problem. In our approach, we consider the entire problem as regression and train image to image network using Fully convolutional layers with GRUs.

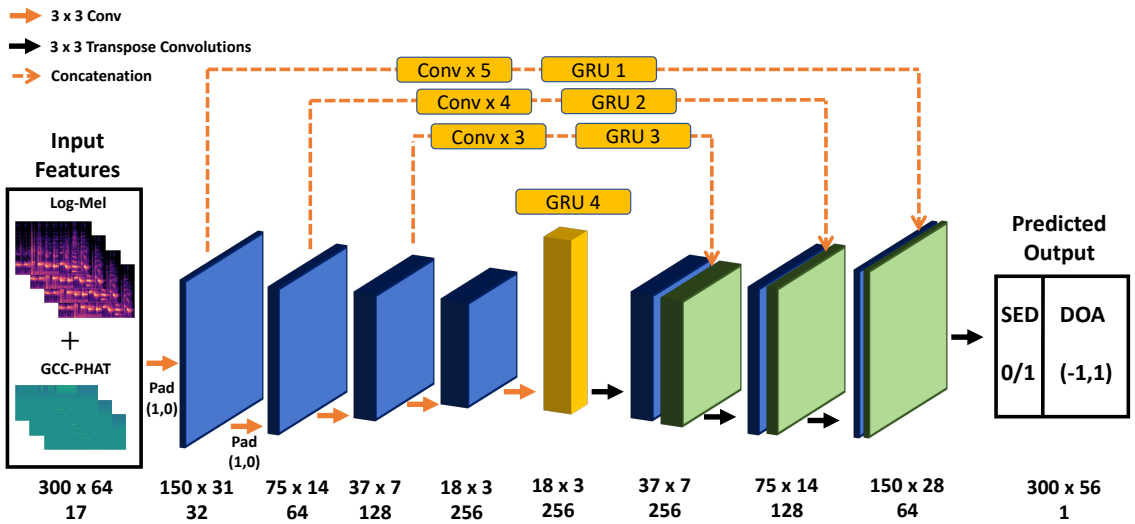


Figure 2.1. Fully convolutional architecture with skip conv-GRU layers for SELD.

The architecture we use is a modified version of U-NET [30] with skip connections replaced by convolutional and GRU layers as seen in Figure. 2.1. Each path towards the output with/without skip connections serves similar to SELDNet like CRNN model. The intuition behind this architecture is to avoid the use of ensemble models and stack all the models in the final stage. In total there are 4 paths serving as 4 SELDNet like models. The main path consists of the following sequence of layers. The squeezing path is given by: encoder1 $\rightarrow$ encoder2 $\rightarrow$ encoder3 $\rightarrow$ encoder4 $\rightarrow$ GRU4 $\rightarrow$ decoder1 $\rightarrow$ decoder2 $\rightarrow$ decoder3 $\rightarrow$ decoder4 $\rightarrow$ output. Throughout the training we use (3 x 3) convolutional filters with padding on both sides unless specified. We specifically use padding and max pooling (stride=(2 x 2)) to reduce feature maps instead of stride to change the shape of the output image to that of the reference label. The layers in the main path and their hyper-parameters with corresponding outputs are as described Table.2.1. Encoder1 and 2 blocks have not been zero padded along third dimension (frequency) while performing convolutions. As such after the convolution and max-pooling their size is reduced to (32 x 300 x 62) and (32

Table 2.1. Encoder Decoder Architecture

Stage	Output	Layers
input	17x300x64	input features
encoder1	32x150x31	conv,BN,pReLu,maxpool,pad(1,0)
encoder2	64x75x14	conv,BN,pReLu,maxpool,pad(1,0)
encoder3	128x37x7	conv,BN,pReLu,maxpool,pad(1,1)
encoder4	256x18x3	conv,BN,ReLu,maxpool,pad(1,1)
GRU 4	256x18x3	Bidirectional 2 layer,768 GRU units
decoder1	128x37x7	convTranspose,BN, ReLu
decoder2	64x75x14	convTranspose,BN, ReLu
decoder3	32x150x28	convTranspose,BN, ReLu
decoder4	1x300x56	convTranspose,BN, TanH

x 150 x 31) respectively instead of (32 x 300 x 64) and (32 x 150 x 32) if done with zero padding along third dimension. Similarly for Encoder2 the output is reduced to (64 x 75 x 14). Encoder3 and 4 follow usually with zero padding along both the dimensions.

The output of GRU4 is then upsampled by doing transposed convolutions to get (128 x 37 x 7). This output is concatenated with encoder3 output and again upsampled. However, the output of decoder3 has (150 x 28) feature maps as opposed to (150 x 31) of encoder 1. Hence in encoder1, we crop the first and last 2 features across last dimension and perform upsampling to achieve (150 x 28) features. The last layer does not need to any concatenation and is directly upsampled to achieve the predicted labels. Unlike U-NET that performs all symmetric reduction and upsampling of features, this approach has asymmetric skipped operations.

There are in total three skipped connections in this architecture. Each connection consists of repeated convolutions followed by Bidirectional GRUs. The (conv x 5), (conv x 4) and (conv x 3) represents performing 5, 4 and 3 repeated convolutions. The layers details can be found in Table.2.2. The convolutions are performed with same input and output features such that their dimensions remain same in order to concatenate with upsampled features.

Table 2.2. Skip connection/concatenation layers

Stage	Output	Layers
conv x 5	32x150x31	5 repeated convolutions conv,BN,ReLu
GRU 1	32x150x31	Bidirectional 2 layer,992 GRU units
conv x 4	64x75x14	4 repeated convolutions conv,BN,ReLu
GRU 2	64x75x14	Bidirectional 2 layer,896 GRU units
conv x 3	128x37x7	3 repeated convolutions conv,BN,ReLu
GRU 3	128x37x7	Bidirectional 2 layer,896 GRU units

2.3. RESULTS

The trained models are used to predict the outcomes of the validation and test dataset. All GRU layers are set with dropout of 0.2. The approach used in this study is very specific to the input feature dimensions. For other previous approaches like SELDNet models [2] and ensemble approaches, the model can accept variable inputs. But in our case, since we are performing asymmetric skip connections in the so called fully convolutional architecture and with image to image learning and the size of input and output image being different. The entire model needs to be changed in order to be compatible with different input feature shape.

We consider input feature to have 300 frames or 300 time steps with 64 mel bands. The image is then passed through the network to output a (300 x 56) image with predicted SED+DOA (14+42) values. From this output, the SED corresponding values are rounded to their nearest integer. Since the reference labels have been provided for every 5 frames we squeeze out one set of labels for every 5 time steps to calculate the SELD metrics.

Table 2.3. Results on development dataset

Submission	ER_{20^0}	F_{20^0}	LE_{CD}	LR_{CD}	SELD
1	0.55	54.2	13.6^0	63.6	0.35
2	0.56	53.7	14^0	62.6	0.37
3	0.55	55.4	14.9^0	66.5	0.35
4	0.54	55.6	15.2^0	67.2	0.35
baseline	0.72	37.4	22.8^0	60.7	0.49

2.4. SUBMISSIONS

Of the four submissions first two are FCn based and the other two are CRNN architecture based similar to SELDNet. Although the performance of FCn based model has better performance but it is only slight improvement in terms of DOA error as compared to submission 3 and 4. The layers for latter model are given in Table.2.4. The SED labels are considered as classification problem with and DOA are considered as regression while calculating losses. Unlike the FCn model where entire model output considered as regression problem. For all submissions the learning rate was fixed throughout training and was optimized using Adam optimizer. For submissions (1,3) and (2,4), the learning rate was 0.0001 and 0.00005 respectively.

2.5. SUMMARY

We conclude that the performance of our approach outperform baseline methods. But there definitely is further room for improvement. More complex and deeper architectures need to be experimented with like ResNET, Inception models among others. Also ensemble methods might lead to further reduction in training losses and better generalization for evaluation dataset.

Table 2.4. Alternate SELDNet like CRNN architecture

Stage	Output	Layers
input	17x300x64	input features
conv1	32x300x64	conv,BN,ReLu
conv2	64x300x32	conv,BN,ReLu,maxpool(1,2)
conv3	128x300x16	conv,BN,ReLu,maxpool(1,2)
conv4	256x300x8	conv,BN,ReLu,maxpool(1,2)
conv5	512x300x4	conv,BN,ReLu,maxpool(1,2)
conv6	512x60x2	conv,BN,ReLu,maxpool(5,2)
GRU	60x1024	2 layer,1024 GRU units
dense1	60x512	Fully connected layer,dropout
dense SED	60x14	Fully connected layer
dense2	60x512	Fully connected layer,dropout
dense DOA	60x42	Fully connected layer

3. CONCLUSIONS

The contributions are majorly for microphone array geometry optimization and RFID localization. As for the localization, both DOA and 3D estimation techniques are proposed. We conclude the following contributions for the dissertation.

1. Proposed simultaneous optimization of microphone array geometry and regularization parameter using analytical approach PSO.

2. Overall aperture reduction for linear arrays of 40% in comparison to nested arrays in case higher directivity is desired.

3. 15% smaller aperture compared to stochastic and analytic optimized arrays for frequency invariant response

4. Optimized planar array geometry for a given maximum aperture and the desired DF or WNG for multiple look directions.

5. The approach presented for planar arrays can achieve robust super directive frequency invariant response.

6. Reduction of the aperture by 10% and 20% along x and y dimension respectively when compared with UCA, uniform grid and the recently proposed GA based array optimization at 1kHz.

7. RFID tag Direction Of Arrival (DOA) estimation is estimated with minimal assumptions with lowest latency of 0.25 seconds and average error of $\approx 6.06^0$.

8. The proposed approach is setup using commonly available off the shelf equipment and does not require multiple antennas and refrence tag locations to be known in advance.

9. The proposed DOA approach was able to model the multi path, spatial aliasing and NLOS effects with minimal assumptions.

10. Proposed 3D estimation of RFID tags with average localization error of $\mathbf{2.40}cms$.

11. The localization performance was evaluated for models trained on data collected from different workspaces.

12. Reduction of $\approx \mathbf{1.77}cms(\mathbf{42}\%)$ in localization error as compared to the baseline DNN model with latency as low as 0.25 seconds.

13. Localization error was further reduced by **44.5**% to about $\mathbf{1.33}cms$ with use of phase information.

14. The performance of the localization is evaluated on different days resulting in consistent performance.

15. We also evaluate the performance for different form factor for the given IC with localization error increased by $\approx \mathbf{10}\%$ to $\mathbf{1.46}cms$.

16. The proposed approach was trained on 50%-50% for training and testing model while still achieving accurate results.

REFERENCES

[1] http://dcase.community/challenge2019/.

[2] Sharath Adavanne, Archontis Politis, Joonas Nikunen, and Tuomas Virtanen. Sound event localization and detection of overlapping sources using convolutional recurrent neural networks. *IEEE Journal of Selected Topics in Signal Processing*, 13(1):34–48, March 2018. ISSN 1932-4553. doi: 10.1109/JSTSP.2018.2885636. URL https://ieeexplore.ieee.org/abstract/document/8567942.

[3] Noor Bahaaldin and Ergun Ercelebi. Bvire improved algorithm for indoor localization based on rfid and a linear regression model. *Turkish Journal of Electrical Engineering & Computer Sciences*, 26(6):2943–2957, 2018.

[4] R. Berkun, I. Cohen, and J. Benesty. Combined beamformers for robust broadband regularized superdirective beamforming. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 23(5):877–886, May 2015. ISSN 2329-9290. doi: 10.1109/TASLP.2015.2410139.

[5] R. Berkun, I. Cohen, and J. Benesty. A tunable beamformer for robust superdirective beamforming. In *2016 IEEE International Workshop on Acoustic Signal Enhancement (IWAENC)*, pages 1–5, Sept 2016. doi: 10.1109/IWAENC.2016.7602952.

[6] M. Bjelić, M. Stanojević, D. Pavlović, and M. Mijić. Microphone array geometry optimization for traffic noise analysis. *The Journal of the Acoustical Society of America*, 141(5):3101–3104, 2017. doi: 10.1121/1.4982694. URL http://dx.doi.org/10.1121/1.4982694.

[7] M. Brandstein and D. Ward. *Microphone arrays: signal processing techniques and applications*. Springer Science & Business Media, 2013.

[8] Yin Cao, Qiuqiang Kong, Turab Iqbal, Fengyan An, Wenwu Wang, and Mark Plumbley. Polyphonic sound event detection and localization using a two-stage strategy. In *Proceedings of the Detection and Classification of Acoustic Scenes and Events 2019 Workshop (DCASE2019)*, pages 30–34, New York University, NY, USA, October 2019.

[9] Shuyan Cheng, Shujun Wang, Wenbai Guan, He Xu, and Peng Li. 3dlra: An rfid 3d indoor localization method based on deep learning. *Sensors*, 20(9), 2020. ISSN 1424-8220. doi: 10.3390/s20092731. URL https://www.mdpi.com/1424-8220/20/9/2731.

[10] H. Cox, R Zeskind, and T. Kooij. Practical supergain. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 34(3):393–398, 1986.

[11] H. Cox, R. Zeskind, and M. Owen. Robust adaptive beamforming. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 35(10):1365–1376, Oct 1987. ISSN 0096-3518. doi: 10.1109/TASSP.1987.1165054.

[12] M. Crocco and A. Trucco. Stochastic and analytic optimization of sparse aperiodic arrays and broadband beamformers with robust superdirective patterns. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(9):2433–2447, Nov 2012. ISSN 1558-7916. doi: 10.1109/TASL.2012.2203808.

[13] Javier Diaz, Rodrigo Maues, Rodrigo Soares, Eduardo Nakamura, and Carlos Figueiredo. Bluepass: An indoor bluetooth-based localization system for mobile applications. pages 778–783, 06 2010. doi: 10.1109/ISCC.2010.5546506.

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

[15] Juliana Borges Ferreira, Álvaro Augusto Almeida de Salles, Giovani Bulla, and Eduardo Luis Rhod. A new structure of uhf rfid tag antenna mountable on metallic surface using double slits. In *2017 IEEE 9th Latin-American Conference on Communications (LATINCOM)*, pages 1–3. IEEE, 2017.

[16] Sinan Gezici, Zhi Tian, G.B. Giannakis, Hisashi Kobayashi, A.F. Molisch, H. Vincent Poor, and Zafer Sahinoglu. Localization via ultra-wideband radios: A look at positioning aspects for future sensor networks. *IEEE Signal Processing Magazine*, 22: 70–84, 08 2005. doi: 10.1109/MSP.2005.1458289.

[17] Jeffrey Hightower, Roy Want, and Gaetano Borriello. Spoton: An indoor 3d location sensing technology based on rf signal strength. 2000.

[18] B. Hu, H. Peng, and Z. Sun. Landmarc localization algorithm based on weight optimization. *Chinese Journal of Electronics*, 27(6):1291–1296, 2018. doi: 10.1049/cje.2017.08.011.

[19] Sławomir Kapka and Mateusz Lewandowski. Sound source detection, localization and classification using consecutive ensemble of crnn models. *arXiv preprint arXiv:1908.00766*, 2019.

[20] J. Kennedy. Particle swarm optimization. In *Encyclopedia of machine learning*, pages 760–766. Springer, 2011.

[21] Chenyang Li, Lingfei Mo, and Dongkai Zhang. Review on uhf rfid localization methods. *IEEE Journal of Radio Frequency Identification*, 3(4):205–215, 2019.

[22] J. Liu, F. Zhu, Y. Wang, X. Wang, Q. Pan, and L. Chen. Rf-scanner: Shelf scanning with robot-assisted rfid systems. In *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, pages 1–9, 2017. doi: 10.1109/INFOCOM.2017.8056985.

[23] I. Mccowan. Microphone arrays : A tutorial. 2001.

[24] Robert Miesen, Fabian Kirsch, and Martin Vossiek. Holographic localization of passive uhf rfid transponders. In *2011 IEEE international conference on RFID*, pages 32–37. IEEE, 2011.

[25] Lionel M Ni, Yunhao Liu, Yiu Cho Lau, and Abhishek P Patil. Landmarc: indoor location sensing using active rfid. In *Proceedings of the First IEEE International Conference on Pervasive Computing and Communications, 2003.(PerCom 2003).*, pages 407–415. IEEE, 2003.

[26] P. Pal and P. P. Vaidyanathan. Nested arrays: A novel approach to array processing with enhanced degrees of freedom. *IEEE Transactions on Signal Processing*, 58(8): 4167–4181, Aug 2010. ISSN 1053-587X. doi: 10.1109/TSP.2010.2049264.

[27] Andreas Parr, Robert Miesen, and Martin Vossiek. Inverse sar approach for localization of moving rfid tags. In *2013 IEEE International Conference on RFID (RFID)*, pages 104–109. IEEE, 2013.

[28] Parth H Pathak, Xiaotao Feng, Pengfei Hu, and Prasant Mohapatra. Visible light communication, networking, and sensing: A survey, potential and challenges. *IEEE communications surveys & tutorials*, 17(4):2047–2077, 2015.

[29] Archontis Politis, Sharath Adavanne, and Tuomas Virtanen. A dataset of reverberant spatial sound scenes with moving sources for sound event localization and detection. *arXiv e-prints: 2006.01919*, 2020. URL `https://arxiv.org/abs/2006.01919`.

[30] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[31] H. Schjaer-Jacobsen and K. Madsen. Synthesis of nonuniformly spaced arrays using a general nonlinear minimax optimisation method. *IEEE Transactions on Antennas and Propagation*, 24(4):501–506, Jul 1976. ISSN 0018-926X. doi: 10.1109/TAP.1976.1141369.

[32] S. S. Shanan and C. A. Pomalaza-Raez. The use of nonuniform element spacing in array processing algorithms. In *12th Canadian Symposium on Remote Sensing Geoscience and Remote Sensing Symposium*, volume 5, pages 2977–2979, Jul 1989. doi: 10.1109/IGARSS.1989.575979.

[33] Longfei Shangguan and Kyle Jamieson. The design and implementation of a mobile rfid tag sorting robot. In *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys '16, page 31–42, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450342698. doi: 10.1145/2906388.2906417. URL `https://doi.org/10.1145/2906388.2906417`.

[34] L. Shen, Q. Zhang, J. Pang, H. Xu, and P. Li. Prdl: Relative localization method of rfid tags via phase and rssi based on deep learning. *IEEE Access*, 7:20249–20261, 2019. doi: 10.1109/ACCESS.2019.2895129.

[35] P. Townsend. Enhancements to the generalized sidelobe canceller for audio beamforming in an immersive environment. 2009.

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

[37] Jue Wang, Fadel Adib, Ross Knepper, Dina Katabi, and Daniela Rus. Rf-compass: Robot object manipulation using rfids. In *Proceedings of the 19th annual international conference on Mobile computing & networking*, pages 3–14, 2013.

[38] He Xu, Ye Ding, Peng Li, Ruchuan Wang, and Yizhu Li. An rfid indoor positioning algorithm based on bayesian probability and k-nearest neighbor. *Sensors*, 17(8):1806, 2017.

[39] Huatao Xu, Dong Wang, Run Zhao, and Qian Zhang. Faho: Deep learning enhanced holographic localization for rfid tags. In *Proceedings of the 17th Conference on Embedded Networked Sensor Systems*, SenSys '19, page 351–363, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450369503. doi: 10.1145/3356250.3360035. URL `https://doi.org/10.1145/3356250.3360035`.

[40] Lei Yang, Yekui Chen, Xiang-Yang Li, Chaowei Xiao, Mo Li, and Yunhao Liu. Tagoram: Real-time tracking of mobile rfid tags to high precision using cots devices. In *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking*, MobiCom '14, page 237–248, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 9781450327831. doi: 10.1145/2639108.2639111. URL `https://doi.org/10.1145/2639108.2639111`.

[41] J. Yu and K. D. Donohue. Performance for randomly described arrays. In *2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 269–272, Oct 2011. doi: 10.1109/ASPAA.2011.6082273.

[42] J. Yu and K. D. Donohue. Geometry descriptors of irregular microphone arrays related to beamforming performance. *EURASIP Journal on Advances in Signal Processing*, 2012(1):249, Nov 2012. ISSN 1687-6180. doi: 10.1186/1687-6180-2012-249. URL `https://doi.org/10.1186/1687-6180-2012-249`.

[43] J. Yu and K. D. Donohue. Optimal irregular microphone distributions with enhanced beamforming performance in immersive environments. *The Journal of the Acoustical Society of America*, 134(3):2066–2077, 2013. doi: 10.1121/1.4816540. URL `http://dx.doi.org/10.1121/1.4816540`.

[44] Y. Zhao, Y. Liu, and L. M. Ni. Vire: Active rfid-based localization using virtual reference elimination. In *2007 International Conference on Parallel Processing (ICPP 2007)*, pages 56–56, 2007. doi: 10.1109/ICPP.2007.84.

[45] Y. Zhao, K. Liu, Y. Ma, Z. Gao, Y. Zang, and J. Teng. Similarity analysis-based indoor localization algorithm with backscatter information of passive uhf rfid tags. *IEEE Sensors Journal*, 17(1):185–193, 2017. doi: 10.1109/JSEN.2016.2624314.

VITA

Sohel Jayesh Patel received a B. Tech from Jawaharlal Nehru Technical University, Hyderabad in 2012. And an M.S. in electrical Engineering from Missouri University of Science and Technology in 2016. He received a PhD in Electrical Engineering from Missouri University of Science and Technology in May 2022. His research is focused on array processing, RFID-based systems, digital signal processing for audio and communications and sensor array geometry optimization.