

Performance evaluation of direction-finding techniques of an acoustic source with uniform linear array

Syed Farid Uddin, Ayan Alam Khan and Mohd Wajid
*Department of Electronics Engineering, ZHCET, Aligarh Muslim University,
Aligarh, India*

Mahima Singh
*Department of Electronics and Communication Engineering,
Indira Gandhi Delhi Technical University for Women, New Delhi, India and
Department of Electronics Engineering, ZHCET, Aligarh Muslim University,
Aligarh, India, and*

Faisal Alam
*Department of Computer Engineering, ZHCET, Aligarh Muslim University,
Aligarh, India*

Abstract

Purpose – The purpose of this paper is to show a comparative study of different direction-of-arrival (DOA) estimation techniques, namely, multiple signal classification (MUSIC) algorithm, delay-and-sum (DAS) beamforming, support vector regression (SVR), multivariate linear regression (MLR) and multivariate curvilinear regression (MCR).

Design/methodology/approach – The relative delay between the microphone signals is the key attribute for the implementation of any of these techniques. The machine-learning models SVR, MLR and MCR have been trained using correlation coefficient as the feature set. However, MUSIC uses noise subspace of the covariance-matrix of the signals recorded with the microphone, whereas DAS uses the constructive and destructive interference of the microphone signals.

Findings – Variations in root mean square angular error (RMSAE) values are plotted using different DOA estimation techniques at different signal-to-noise-ratio (SNR) values as 10, 14, 18, 22 and 26dB. The RMSAE curve for DAS seems to be smooth as compared to PR1, PR2 and RR but it shows a relatively higher RMSAE at higher SNR. As compared to (DAS, PR1, PR2 and RR), SVR has the lowest RMSAE such that the graph is more suppressed towards the bottom.

Originality/value – DAS has a smooth curve but has higher RMSAE at higher SNR values. All the techniques show a higher RMSAE at the end-fire, i.e. angles near 90°, but comparatively, MUSIC has the lowest RMSAE near the end-fire, supporting the claim that MUSIC outperforms all other algorithms considered.

Keywords Direction-of-arrival (DOA) Delay-and-sum (DAS) beamforming, Multiple signal classification (MUSIC) multivariate-linear-regression (MLR), Multivariate-curvilinear-regression (MCR), RMSAE, Support vector regression (SVR)

Paper type Research paper



1. Introduction

The determination of the direction-of-arrival (DOA) of an acoustic signal is a problem that is studied under the ambit of localization and tracking. It has applications in various domains, namely, robotics, where unmanned vehicles have to move in an unexplored/new environment, radar systems for aerial/underwater target tracking, sonar system, surveillance system, where a camera needs to align to a direction from where sound is coming (Johnson and Dudgeon, 1993; Godara, 1997; Asaei *et al.*, 2016; Bekkerman and Tabrikian, 2006; Zhao *et al.*, 2010, 2012; Clark and Tarasek, 2006; Bechler *et al.*, 2004; Argentieri and Danes, 2007; Xiao *et al.*, 2014; Delikaris-Manias *et al.*, 2016; Zhang *et al.*, 2008). The DOA estimation is challenging, as there is a certain distortion in the acquired signals, the possible reasons being sensor noise, ambient noise, non-uniformity in array elements, reverberation, interferences or a combination of these impairments. The presence of distortion causes inaccurate estimation of DOAs of an acoustic source. Techniques for DOA estimation with small errors in the presence of impairments use different acoustic vector sensors and microphone/sensor array configurations (Bogaert *et al.*, 2011; Wajid *et al.*, 2017a, b, 2019, 2020a, b; Yadav *et al.*, 2020; Alam *et al.*, 2021; Liu *et al.*, 2019). The DOA estimation techniques are classified as based on regression, beamforming and subspace method. delay-and-sum (DAS), minimum-variance-distortionless-response (MVDR) and fast-fourier transform-effective aperture distribution function (FFT-EADF) are present under the subcategory of the beamforming method. Polynomial-regression of order 1 (PR1), Polynomial-regression of order 2 (PR2), support vector regression (SVR), ridge-regression (RR) are under regression techniques. Multiple signal classification (MUSIC), root MUSIC, estimation of signal parameters via rotational invariance technique (ESPRIT), etc. algorithms are present under the subcategory of the subspace method (Shi, 2019; Cui *et al.*, 2019; Gupta *et al.*, 2020; Varma, 2002; Zhou *et al.*, 2017; Tang *et al.*, 2014). DOA estimation has paved its path from early methods where narrow beams are steered in a particular direction for knowing the incident angle. Digital signal processors have been used as an approach for finding the direction. Methods such as subspace decomposition, analysis of eigen values and compressed sensing-based methods are playing an important role in achieving better performance in terms of speed, accuracy and robustness (Ge *et al.*, 2021; Zhang *et al.*, 2021).

Understanding the wide range of applications for DOA estimation, the increased variety of sensor configurations, and the wide knowledge of constraints implied by the hardware, the research into DOA estimation strategies has continued uninterrupted. Recently, new approaches based on deep neural networks (DNNs) for many speech sources localization using an array of smaller dimensions. These techniques result into high-resolution DOA estimation. DNNs, amongst the data-driven methods, show the potency for high precision DOA estimation. The recent deep learning technology undergo heavy analysis, proving the importance of DOA estimation; various combinations of convolutional neural network and deep neural network are taken into account. The evaluation criteria include root mean-squared error, accuracy and mean absolute error. They evaluate this new deep and machine learning technology in DOA estimation, and various factors (signal-to-noise ratio, number of snapshots, number of antennas and number of signal sources) affecting DOA estimation are also processed. Based on findings, it is being believed that advanced technologies like deep learning has improved the direction-finding techniques to a greater extent. Such kind of study helps researchers to conduct detailed analysis (Ge *et al.*, 2021; Zhang *et al.*, 2021).

This paper is divided into the following sections. Section 2 presents the signal model for the uniformly linear array. Section 3 briefly describes the different direction-finding techniques. In Section 4, simulation parameters and results are presented and analysed, Section 5 concludes the paper.

2. Signal model for uniform linear array

A uniform linear array (ULA) of M number of microphones is used as a receiver, where two adjacent array microphones are separated by a distance, d . Let there are D number of far-field acoustic sources and transmitted signal, $s_1(t)$, a narrowband signal. Figure 1 depicts the far-field sound source and a ULA of microphones with $M = 4$ and $D = 1$. Assuming a single sound source in the far-field and the wavelength of the incoming signal is λ_1 , which arrives at an angle θ_1 with respect to the y -axis and in the clockwise direction. Thus, the equation of the signal received by the i th microphone is expressed as

$$r_i(t) = s_1(t)e^{-j[\beta_1(\theta_1) + \gamma(\theta_1)]} + n_i(t), \quad (i = 1, 2, \dots, M), \quad (1)$$

where $e^{-j\gamma(\theta_1)}$ is the phase component which is common to all microphones signal and introduced due to wave travel from the sound source (at an angle θ) to the first microphone of ULA, and $\beta_1(\theta) = 2\pi(i-1)\frac{d}{\lambda_1}\sin\theta_1$, is called additional phase difference caused by the path difference between the first microphone and the i th microphone. $n_i(t)$ represents the AWGN at the i th microphone.

Additive white Gaussian noise (AWGN) has been used to represent the inherent noise of microphone sensor and electronics system noise, ambient noise, etc. Now, (1) can also be rewritten in the matrix form and is given by Wajid et al. (2020c)

$$\begin{aligned} \mathbf{r}(t) &= \mathbf{A}(\theta)\mathbf{s}(t) + \mathbf{n}(t) \\ &= [r_1(t) \ r_2(t) \ \dots \ r_M(t)]^T, \end{aligned} \quad (2)$$

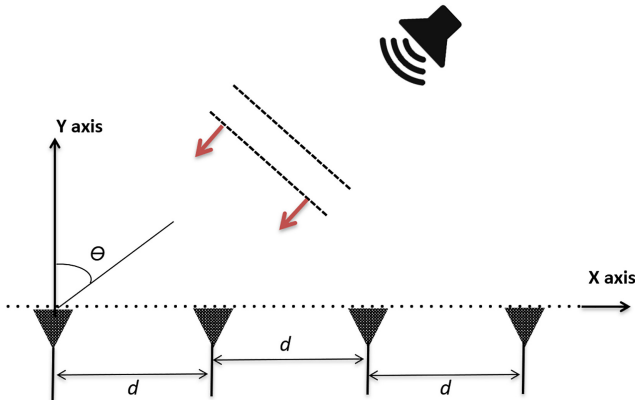
where

$$\mathbf{s}(t) = [s_1(t)e^{-j\gamma(\theta_1)} \ s_2(t)e^{-j\gamma(\theta_2)} \ \dots \ s_D(t)e^{-j\gamma(\theta_D)}]^T, \quad (3)$$

$$\mathbf{A}(\theta) = \begin{bmatrix} 1 & 1 & \dots & 1 \\ e^{-j\cdot\beta_2(\theta_1)} & e^{-j\cdot\beta_2(\theta_2)} & \dots & e^{-j\cdot\beta_2(\theta_D)} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-j\cdot\beta_M(\theta_1)} & e^{-j\cdot\beta_M(\theta_2)} & \dots & e^{-j\cdot\beta_M(\theta_D)} \end{bmatrix}, \quad (4)$$

$\mathbf{n}(t)$ is a vector representation of AWGN and $[\cdot]^T$ denotes the transpose, $\mathbf{A}(\theta)$ is the Vandermonde structure of the array steering matrix $\mathbf{A}(\theta)$ (matrix order $M \times D$). The

Figure 1. Uniform linear array of microphones with $M = 4$ and $D = 1$, where the filled triangles represent the microphones. Assume that the sound source is in the far-field



correlation matrix $\mathbf{R}_{rr}$ (having M rows and M columns) of microphone's signal vector, $\mathbf{r}(t)$, is expressed in the following equation (Wajid *et al.*, 2020c)

$$\begin{aligned}\mathbf{R}_{rr} &= E[\mathbf{r}(t)\mathbf{r}^H(t)] \\ &= \mathbf{A}(\theta)\mathbf{R}_s\mathbf{A}^H(\theta) + \mathbf{N}_n\end{aligned}\quad (5)$$

where, $[\cdot]^H$ denote the conjugate-transpose and $E[\cdot]$ denote the ensemble average. Similarly, (say) $\mathbf{R}_s$ and $\mathbf{N}_n$ represents signal and noise correlation matrices, respectively. Therefore, they can be expressed as follows:

$$\mathbf{R}_s = E[\mathbf{s}(t)\mathbf{s}^H(t)] \quad (6)$$

and

$$\mathbf{N}_n = E[\mathbf{n}(t)\mathbf{n}^H(t)]. \quad (7)$$

Since the noise realizations are mutually uncorrelated, so their cross-correlation is zero and all noise realizations will have the same variance. Thus, $\mathbf{N}_n$ is expressed as

$$\mathbf{N}_n = \sigma^2 \mathbf{I} \quad (8)$$

where σ^2 is the variance of zero-mean AWGN and $\mathbf{I}$ is the identity matrix. Substituting value of $\mathbf{N}_n$ from (8) into (5) results in the following equation (Wajid *et al.*, 2020c).

$$\mathbf{R}_{rr} = \mathbf{A}(\theta)\mathbf{R}_s\mathbf{A}^H(\theta) + \sigma^2\mathbf{I}. \quad (9)$$

3. Techniques of the direction of arrival estimation

There are many existing techniques of DOA estimation, which can be categorized based on three broad approaches, (1) regression modelling, (2) classical beamformer and (3) subspace methods, which are shown in Figure 2. This paper presents the extension work of

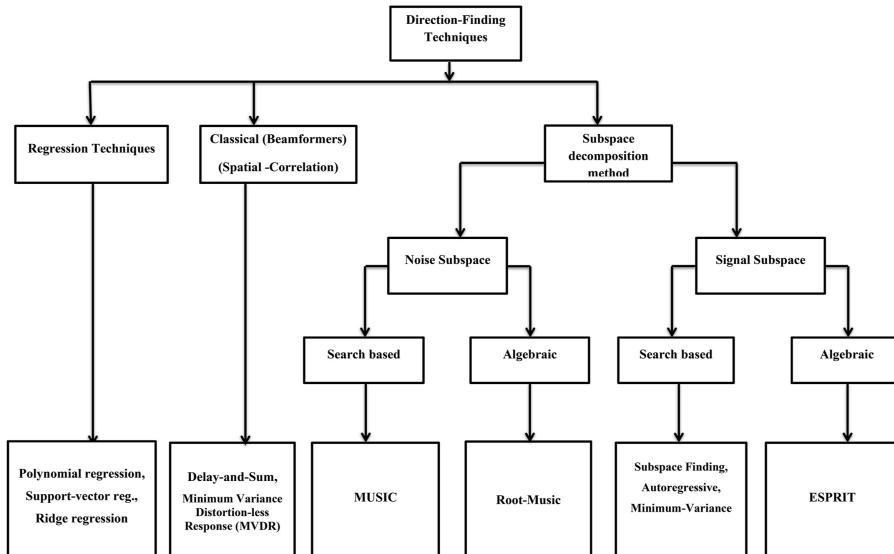


Figure 2.
Direction of arrival
estimation techniques

Wajid *et al.* (2020c). In this paper, we have compared polynomial regression, RR and SVR, DAS beamforming with the subspace technique, i.e. MUSIC algorithm for DOA estimation. The details of the direction-finding techniques among which comparison has been made are given in the subsequent subsections.

3.1 Regression technique

Regression is a statistical method that attempts to determine the nature and degree of relationship between a dependent variable based on many independent variables. The nature of the relationship is produced in the form of a mathematical model (equation) between the predictors and response. The coefficients in the mathematical model are found by undergoing training of parameters. The training process aims at reducing the error distance between the predicted and the actual values in the training by using the best-fit parameters to adapt to the training set. The error estimate between the predictors and response is assessed by the least-squares method in this work. This error is given by the following equation:

$$\text{Error} = \left(\frac{1}{n} \right) \sum_{i=1}^n \left(\mathbf{Y}_i - \widehat{\mathbf{Y}}_i \right)^2 \quad (10)$$

where $\mathbf{Y}$ is the vector of observed values and $\widehat{\mathbf{Y}}$ is the set of predicted values and n is the number of predictions. The regression techniques require the identification of the features derived from the independent variables. These features are then used as an input to identify the mathematical model that is to be determined.

In this work, we have used Pearson Product–Moment–Correlation–Coefficient (PPMCC) as the feature that is taken as the input in the training of the mathematical model. PPMCC is the degree of association or dissociation between two variables. If a variable increases with the increase in the other variable, then the correlation between the two variables is +1. If the variable decreases with the increase in another variable, then the coefficient is –1. The rest of the values lie between +1 and –1 commensurate with the degree of association between the variables. As a feature, PPMCC is calculated on each signal pair acquired at each microphone. The PPMCC thus calculated between each microphone pair is indicative of the phase difference between the sinusoidal waves received at each microphone. The phase difference occurs due to a certain time delay in the reception of signals in the ULA of the microphones. Different regression techniques used are discussed as follows:

3.2 Polynomial-regression of order 1 and order 2

Polynomial-regression or linear-regression is the simplest machine learning algorithm that can be used for estimating the DOA which is given in (11). If $k = 1$ and $k = 2$, then they are denoted with PR1 and PR2, respectively.

$$\mathbf{y} = b_0 + b_1 x_1 + b_2 x_2^2 + \dots + b_k x_k^k + c \quad (11)$$

where $\mathbf{x} = [x_1 x_2 x_3 \dots x_k]$ is the input vector, $\mathbf{b} = [b_1 b_2 b_3 \dots b_k]$ is a vector that consists of weights for different input vectors, c is a constant and $\mathbf{y}$ is the output vector that is dependent on $\mathbf{x}$.

3.3 Support-vector-regression (SVR)

SVR model uses a non-linear model for the estimation of DOA which is trained to relate the input correlation–coefficient features and the output DOA. It uses the Vapnik–Chervonenkis theory of support vectors to form a relationship between predictors and response. Assuming that the predictor variable is denoted by variable $\mathbf{x}$ and the variable of importance, the

dependent response variable is denoted by $G(\mathbf{x})$. The variable $\mathbf{x}$ encompasses all the individual variables that would determine $G(\mathbf{x})$ after training. $\mathbf{x}$ is ' a ' dimensional indicating that ' a ' independent variables are used for prediction. It is defined as follows:

$$\mathbf{x}^T = [x_1, x_2, \dots, x_a]. \quad (12)$$

A general-regression technique requires that the order of the relationship between the predictors and response be predetermined before the training process. The order of relationships could be linear or polynomial. This pre-ascertained relationship hinders the establishment of a mathematical model that is closer to the actual values, as the real relationship could be of scores of a different order than surmised on the proposed order of the polynomial. SVR has a different methodology for determining this mathematical model. To identify a closer model, it uses a kernel function that projects the input variable to an infinitely high dimensional space, with other dimensions as derived dimensions of the input space. To fine-tune the model, the training is performed on a set of known predictor and response values. At the end of the training process, a linear hyperplane is identified in this high dimension that helps minimize the prediction error. The hyperplane thus identified is linear in the high-dimensional derived input space but it is non-linear when projected back in the ' a ' dimensional input space. The process of projecting in the high dimension and then projecting it back in the input space relieves us from predetermining the order of the mathematical model before the training process, thus helping in the establishment of a closer relationship between predictors and response.

Radial basis function (RBF) is a popular function that is used for transforming the input space to a high-dimensional space. This RBH has been used in our experimental work. Its mathematical equation is given as follows:

$$K(\mathbf{x}, \mathbf{x}') = \exp\left(-\gamma|\mathbf{x} - \mathbf{x}'|^2\right), \quad \gamma > 0 \quad (13)$$

where $\mathbf{x}$ and $\mathbf{x}'$ are all vectors in feature space $\mathbb{R}^d$. The function in (13) on expanding reveals that it has an infinitely high number of dimensions. The final value that computes ranges from 0 to 1. The final value thus computed is commensurate with the distance $|\mathbf{x} - \mathbf{x}'|$. To establish a close relationship between predictors and response, multiple linear regression is performed with each variable being the derived dimension in the projected higher dimension. The established mathematical model forms a hyper tube of predicted values in the high-dimension space such that it rotates around the actual values. The linear regression is trained on a set of known values such the overall loss is minimized.

The estimation is measured using a loss-function given by the following equation:

$$L(y, f(\mathbf{x}, \mathbf{w})) = \begin{cases} 0 & \text{if } |y - f(\mathbf{x}, \mathbf{w})| \leq \varepsilon \\ |y - f(\mathbf{x}, \mathbf{w})| - \varepsilon, & \text{otherwise.} \end{cases} \quad (14)$$

This function is an ε -insensitive loss function that forms a tube of width ε such that if the predicted value is in its periphery, then the loss is 0 otherwise the loss is the measure of distance between the predicted value and the tube periphery. The training process performs a linear regression on this high-dimensional feature space and an initially haphazard and high-width tube. The training process then reduces the width of this tube. This is done by minimizing the loss between the predictors and the response using the above-mentioned ε -insensitive loss function. It minimizes the parameter, $\min \frac{1}{2} \|\mathbf{w}\|^2$, where w is the vector normal to the tube. The emphasis is on finding the most flattened tube such that most of the predictions lie within its boundaries (Awad and Khanna, 2015; Alam *et al.*, 2021; Drucker *et al.*, 1997).

3.4 Ridge-regression

RR is used to eradicate some of the drawbacks of the linear regression technique. This technique is for the analysis of multiple regression data in which data has the issue of multicollinearity, in which there is the existence of non-linear relationships among the independent variables. In the case of occurrence of multicollinearity, the linear regression estimates are unbiased but the value of variance for different inputs are so large that the estimated value would be far from the actual value. By adding a percentage of bias to this technique, it reduces the percentage of error and thus RR provides more suitable results.

3.5 Beamforming technique

DAS algorithm is a beamforming technique that estimates DOA using signal power, $P_{DAS}(\theta)$. The DOA is estimated by searching for the values of θ for which $P_{DAS}(\theta)$ show peaks (Awad and Khanna, 2015; Alam *et al.*, 2021; Drucker *et al.*, 1997). $P_{DAS}(\theta)$ is defined as follows:

$$P_{DAS}(\theta) = \underline{\mathbf{a}}^H(\theta) \mathbf{R}_{rr} \underline{\mathbf{a}}(\theta) \quad (15)$$

where $\underline{\mathbf{a}}(\theta)$ contains the look-angle vector of ULA. The look angle vector $\underline{\mathbf{a}}(\theta)$, scans for all possible values of DOA angles to evaluate the estimated values of DOA (Awad and Khanna, 2015).

3.6 Subspace algorithm-based DOA estimation

MUSIC algorithm is a subspace algorithm that uses data collected from ULA to estimate covariance matrix to form subspaces. The steering vector is imposed on the noise-only subspace which leads to the formation of the pseudo-spectrum, the number of peaks in the pseudo-spectrum represents the number of sources and the angular value at which peaks occur is the estimated DOA (Zhang *et al.*, 2021). The eigen-decomposition is used to separate noise subspace and signal subspace. In this algorithm, eigen-decomposition is performed for covariance matrix for any output data of the ULA. This decomposition results in the formation of signal-plus-noise and noise-only subspace. These resulting subspaces will be orthogonal to each other. Later in the algorithm, the orthogonality property is exploited using a steering vector which forms a spectrum function. In the pseudo-spectrum function, we search for peaks, and the corresponding angle at which this peak occurs becomes the estimated DOA (Ahmad and Zhang, 2016; Liao and Abouzaid, 2014). Implementation of MUSIC algorithm is as follows:

- (1) Estimation of covariance matrix from the signal vector acquired by ULA. In practice, $\mathbf{R}_{rr}$ is estimated by averaging over snapshots (N). These snapshots are output data of M -microphones of ULA captured N time instances

$$\hat{\mathbf{R}}_{rr} = \frac{1}{N} \sum_{n=1}^N \mathbf{r}_n \cdot \mathbf{r}_n^H \quad (16)$$

where $\mathbf{r}_n$ (order $M \times 1$) is the output of M -sensors at n th time instant.

- (2) The second step involves the eigen-decomposition of the estimated covariance matrix, $\hat{\mathbf{R}}_{rr}$ with the assumption that $\hat{\mathbf{R}}_{rr}$ is a non-singular matrix. $\hat{\mathbf{R}}_{rr}$ being an $M \times M$ matrix results in M eigenvalues and corresponds to M eigenvectors.
- (3) The third step is the formation of subspaces. The eigenvalues obtained from the second step are used. Among M eigenvalues, the first D number of larger eigenvalues forms eigenvectors which represent signal-plus-noise subspace. The rest of $M - D$ eigenvalues and associated eigenvectors represent noise-only subspace $\mathbf{Q}_n$. If eigenvalues are

$$\lambda_1 > \lambda_2 > \lambda_3 > \dots > \lambda_M,$$

and their corresponding eigenvectors (column vector of $\mathbf{Q}_n$) are

$$\mathbf{v}_1 > \mathbf{v}_2 > \mathbf{v}_3 > \dots > \mathbf{v}_M.$$

Thus, noise subspace becomes

$$\mathbf{Q}_n = [\mathbf{v}_{D+1} \ \mathbf{v}_{D+2} \ \dots \ \mathbf{v}_M] \quad (17)$$

237

- (4) Formation of the pseudo-spectrum is done by projecting look angle vector on the noise subspace (i.e. $\underline{a}^H(\theta) \mathbf{Q}_n$) and is given by

$$P(\theta) = \frac{1}{\underline{a}^H(\theta) \mathbf{Q}_n \left(\underline{a}^H(\theta) \mathbf{Q}_n \right)^H} = \frac{1}{\underline{a}^H(\theta) \mathbf{Q}_n \mathbf{Q}_n^H \underline{a}(\theta)}. \quad (18)$$

Scan the pseudo-spectrum by varying the value of θ , for peaks. For multiple sources, multiple peaks are observed, the corresponding number of values of θ are the estimated DOAs (Ahmad and Zhang, 2016; Liao and Abouzaid, 2014).

4. Simulation environment and results

The properties of the sound wave propagating in the air medium are assumed to be quiescent, isotropic and homogeneous. The microphones are placed along the x -axis in a uniform linear manner. Beam patterns of the microphone array are assumed to be omnidirectional. The separation “ d ” between each of them is 10 cm. A point-sized single sound source is placed at a far distance which is transmitting a sinusoidal signal of frequency 1 kHz and traveling at a speed of sound in air which happens to be 343 m/s. It is assumed that the source is transmitting signals from the far-field. The sampling rate of 48 kHz is chosen for the received signal and the signals are recorded for the duration of 25 ms. The attenuation of signals which are impinging on the microphone surface is not considered in this analysis. The measurements of the DOA are done in the clockwise direction w.r.t. the positive y -axis. A zero-mean white Gaussian noise is added in the received signal vectors with different values of SNR. For every DOA angle, a total of 2,000 independent noisy-signal vectors have been used out of which 1,400 are used for training of the regression model and 600 are used for the testing purpose of the model, for SNR values ranging from 26 to 10 dB, decrementing by 4 dB at each step (Awad and Khanna, 2015).

For training, data of the 46-ary system is used where DOA varies from 0° to 90° with steps of 2° . For testing of the trained models, 91-ary system has been used, where DOAs range from 0° to 90° in steps of 1° . Training has been performed on the signal acquired at the microphones of ULA with SNR = 26 dB with 1,400 independent realizations at each DOA; however, testing has been performed on the signal acquired with SNR = 10, 14, 18, 22 and 26 dB with 600 independent realizations at each DOA. PPMCC on a combination of any two of the microphones for each vector has been calculated and has been used as the feature in the training and testing of regression models.

DAS and MUSIC have also been applied on the ULA microphone signals corresponding to SNR = 10 dB, 14 dB, 18 dB, 22 dB and 26 dB with 600 independent realizations at each DOA 0° – 90° in steps of 1° (91-ary). The spatial scanning/searching of peaks w.r.t. θ is done with a step size of 0.1° in the range of DOA from 0° to 90° as per (15) and (18).

The metrics root mean square angular error ($RMSAE$) and average root mean square angular error ($\overline{RMSAE}$) have been used to evaluate the performance of direction-finding techniques. These evaluation metrics are expressed in (19) and (20)

$$RMSAE(\theta) = \sqrt{\frac{\sum_{i=1}^N (\hat{\theta}_i - \theta)^2}{N}} \quad (19)$$

where $\hat{\theta}_i$ is the estimated angle obtained using i th realization of the actual angle, $N(=600)$ is the total number of times a source was at θ_i . The formula of $\overline{RMSAE}$ can be written as follows:

$$\overline{RMSAE} = \frac{1}{N_T} \sum_{\theta=0^\circ}^{90^\circ} RMSAE(\theta) \quad (20)$$

where N_T is the total number of possible actual-DOAs in a given array (for 91-ary, $N_T = 91$).

In Figures 3–8, graphs represent the result of DOA estimation using different DOA estimation techniques that were tested with 600 independent realizations at different SNR values as described above. Each of Figures 3–7 shows variations in RMSAE values with

Figure 3. RMSAE versus actual-DOA for the (a) SVR, (b) RR, (c) PR1, (d) PR2, (e) DAS and (f) MUSIC algorithm. Training of regression model is done at 26dB SNR and testing is done at 26dB SNR

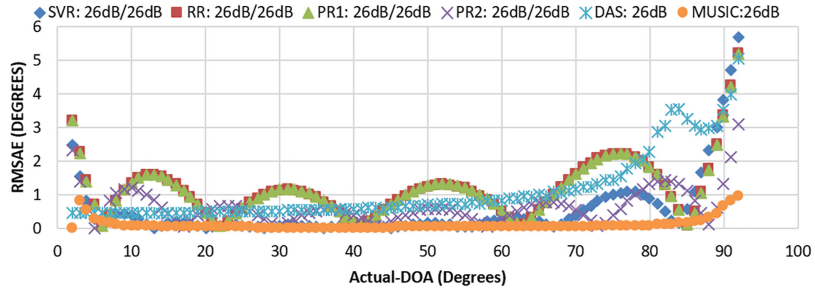


Figure 4. RMSAE vs actual-DOA for the (a) SVR, (b) RR, (c) PR1, (d) PR2, (e) DAS and (f) MUSIC algorithm. Training of regression model is done at SNR = 26dB and testing is performed at SNR = 22dB

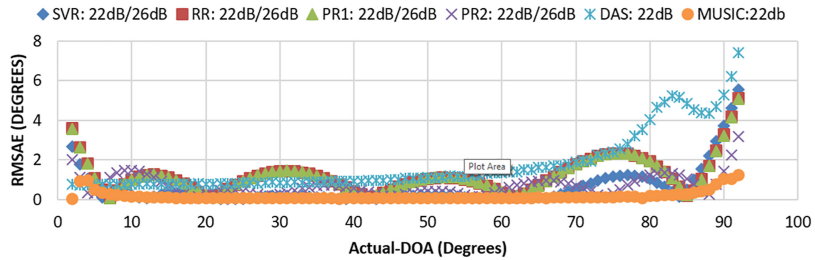
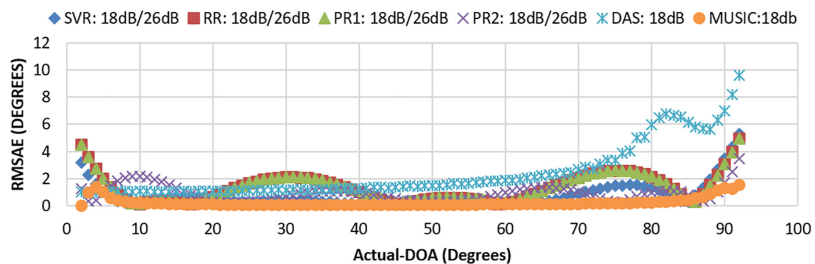


Figure 5. RMSAE vs actual-DOA for the (a) SVR, (b) RR, (c) PR1, (d) PR2, (e) DAS, and (f) MUSIC algorithm. Training of regression model is done at SNR = 26dB and testing is performed at SNR = 18dB



different SNR values as 10, 14, 18, 22 and 26dB, for 600 independent realizations for every value of SNR. Figure 8 represents the comparison of the mentioned DOA estimation techniques in terms of *RMSAE* for signals acquired at each of the SNR. It can be observed from Figures 3 to 7 that the *RMSAE* curve for RR and PR1 seems to be overlapping, with considerably more lobes than other techniques. They also reveal to have higher SNR on average as compared with other techniques. PR2 follows a similar pattern of a higher number of lobes but shows lesser *RMSAE* for all SNR values considered. The *RMSAE* curve for DAS seems to be smooth as compared to PR1, PR2 and RR but it shows a relatively higher *RMSAE* at higher SNR values (14, 18, 22 and 26dB). As compared to (DAS, PR1, PR2 and RR), SVR has the lowest *RMSAE* such that the graph is more suppressed towards the bottom. SVR also shows much fewer lobes than other regression techniques. The lobes in the *RMSAE* curve are not indicative of a good machine learning model as the predicted DOA may have a higher error for a randomly tested angle. A common observation among all the machine learning methods and DAS is that the *RMSAE* is considerably higher towards the end-fire, i.e. angles

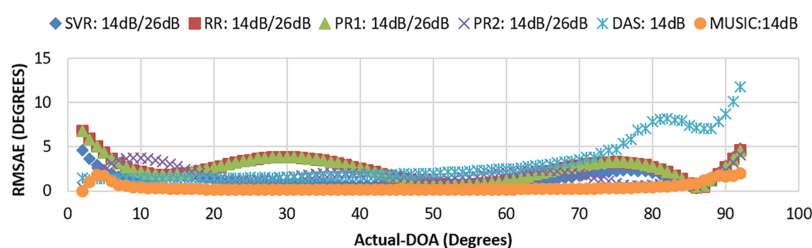


Figure 6. *RMSAE* vs actual-DOA for the (a) SVR, (b) RR, (c) PR1, (d) PR2, (e) DAS and (f) MUSIC algorithm. Training of regression model is done at SNR = 26dB and testing is performed at SNR = 14dB

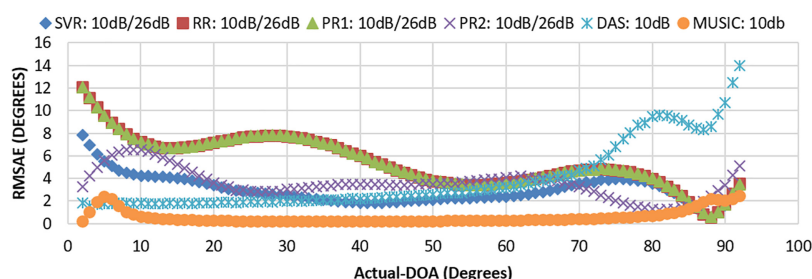


Figure 7. *RMSAE* vs actual-DOA for the (a) SVR, (b) RR, (c) PR1, (d) PR2, (e) DAS and (f) MUSIC algorithm. Training of regression model is done at SNR = 26dB and testing is performed at SNR = 10dB

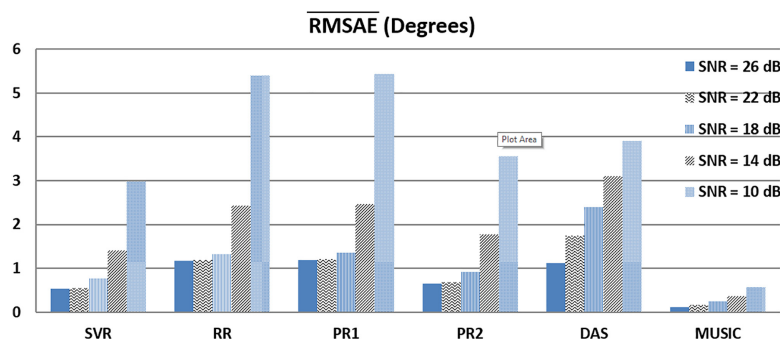


Figure 8. *RMSAE* for the (a) SVR, (b) RR, (c) PR1, (d) PR2, (e) DAS and (f) MUSIC algorithm. The training of regression models is done at SNR = 26dB, and the testing is performed at SNR values ranging between 10 and 26dB with an increment of 4dB

near 90°. MUSIC algorithms prove to be the best, having the lowest RMSAE among all the algorithms considered. It also has the least number of lobes as compared to other techniques. Even close to the end-fire where all other techniques have large RMSAE, the MUSIC algorithm shows a small RMSAE value.

It can be inferred from Figure 8 that the cumulative/average of RMSAE for MUSIC is the least when compared with other techniques. In fact, $\overline{RMSAE}$ for MUSIC is much lower when compared with other techniques for all SNR values. Other than MUSIC techniques, SVR performs better with lower RMSAE at all SNR values as compared with other methods. PR2 performs better than PR1 and RR.

5. Conclusion

A comparative analysis of multiple techniques of DOA estimation, namely, SVR, RR, PR1, PR2, DAS and MUSIC, have been performed in this work. It has been revealed from the experiment that the MUSIC algorithm outperforms all other techniques in terms of RMSAE. Amongst the machine learning techniques, SVR performs better in terms of RMSAE. Techniques such as PR1, PR2 and RR have higher RMSAE and have lobes in the RMSAE curve for DOA estimation. These lobes in the RMSAE curve indicate that the predicted DOA may have a higher error for a randomly tested angle. DAS has a smooth curve but has higher RMSAE compared to other techniques at higher SNR values. All the techniques show a higher RMSAE at the end-fire, i.e. angles near 90°, but comparatively, MUSIC has the lowest RMSAE near the end-fire, supporting the claim that MUSIC outperforms all other algorithms considered.

In the future, this work can be extended by implementing a root-MUSIC algorithm that avoids searching for peaks in the spectrum and angle corresponding to it, rather, it finds roots by defining a variable-based steering vector and uses it to estimate DOA.

References

- Ahmad, M. and Zhang, X. (2016), "Performance of MUSIC algorithm for DOA estimation", *1st International Conference in Aerospace for Young Scientists*.
- Alam, F., Usman, M., Alkhamash, H.I. and Wajid, M. (2021), "Improved direction-of-arrival estimation of an acoustic source using support vector regression and signals correlation", *Sensors*, Vol. 21, p. 2692, doi: [10.3390/s21082692](https://doi.org/10.3390/s21082692).
- Argentieri, S. and Danes, P. (2007), "Broadband variations of the MUSIC high-resolution method for sound source localization in robotics", *Intelligent Robots and Systems, 2007, IROS 2007, IEEE/RSJ International Conference on*, pp. 2009-2014.
- Asaei, A., Taghizadeh, M., Haghighatshoar, S., Raj, B., Boulard, H. and Cevher, V. (2016), "Binary sparse coding of convolutive mixtures for sound localization and separation via spatialization", *IEEE Transactions on Signal Processing*, Vol. 64 No. 3, pp. 567-579.
- Awad, M. and Khanna, R. (2015), "Support vector regression", *Efficient Learning Machines*, Apress, Berkeley, CA, pp. 67-80.
- Bechler, D., Schlosser, M. and Kroschel, K. (2004), "System for robust 3D speaker tracking using microphone array measurements", *Intelligent Robots and Systems, 2004 (IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, pp. 2117-2122.
- Bekkerman, I. and Tabrikian, J. (2006), "Target detection and localization using MIMO radars and sonars", *IEEE Transactions on Signal Processing*, Vol. 54 No. 10, pp. 3873-3883.
- Bogaert, T., Carette, E. and Wouters, J. (2011), "Sound source localization using hearing aids with microphones placed behind-the-ear, in-the-canal, and in-the-pinna", *International Journal of Audiology*, Vol. 50 No. 3, pp. 164-176.

- Clark, J. and Tarasek, G. (2006), "Localization of radiating sources along the hull of a submarine using a vector sensor array", *OCEANS 2006*, IEEE (OCEANS), pp. 1-3.
- Cui, X., Yu, K., Zhang, S. and Wang, H. (2019), "Azimuth-only estimation for TDOA-based direction finding with three-dimensional acoustic array", *IEEE Transactions on Instrumentation and Measurement*.
- Delikaris-Manias, S., Vilkamo, J. and Pulkki, V. (2016), "Signal-dependent spatial filtering based on weighted-orthogonal beamformers in the spherical harmonic domain", *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Vol. 24 No. 9, pp. 1511-1523.
- Drucker, H., Burges, C.J., Kaufman, L., Smola, A.J. and Vapnik, V. (1997), "Support vector regression machines", *Advances in Neural Information Processing Systems*, pp. 155-161.
- Ge, S., Li, K. and Rum, S.N.B.M (2021), "Deep learning approach in DOA estimation: a systematic literature review", *Mobile Information Systems*, Vol. 2021, p. 14, 6392875, doi: [10.1155/2021/6392875](https://doi.org/10.1155/2021/6392875).
- Godara, L.C. (1997), "Application of antenna arrays to mobile communications, part II: beam-forming and direction-of-arrival considerations", *Proceedings of the IEEE*, Vol. 85 No. 8, pp. 1195-1245.
- Gupta, O., Kumar, M., Mushtaq, A. and Goyal, N. (2020), "Localization schemes and its challenges in underwater wireless sensor networks", *Journal of Computational and Theoretical Nanoscience*, Vol. 17 No. 6, pp. 2750-2754.
- Johnson, D.H. and Dudgeon, D.E. (1993), *Array Signal Processing: Concepts and Techniques*, Prentice-Hall signal processing series, PTR Prentice Hall, Englewood Cliffs, NJ, p. 533, ISBN 0130485136.
- Liao, Y. and Abouzaid, A. (2014), "Resolution improvement for MUSIC and ROOT MUSIC algorithms", *Journal of Information Hiding and Multimedia Signal Processing*, Vol. 69 No. 4, pp. 985-994.
- Liu, A., Yang, D., Shi, S., Zhu, Z. and Li, Y. (2019), "Augmented subspace MUSIC method for DOA estimation using acoustic vector sensor array", *IET Radar, Sonar and Navigation*, Vol. 13 No. 6, pp. 969-975.
- Shi, F. (2019), "Two-dimensional direction-of-arrival estimation using compressive measurements", *IEEE Access*, Vol. 7, pp. 20863-20868.
- Tang, H., Nordebo, S. and Cijvat, P. (2014), *DOA Estimation Based on MUSIC Algorithm*.
- Varma, K. (2002), "Time delay estimate based direction of arrival estimation for speech in reverberant environments", Doctoral Dissertation, Virginia Tech.
- Wajid, M., Kumar, A. and Bahl, R. (2017), "Direction-finding accuracy of an air acoustic vector sensor in correlated noise field", *2017 4th International Conference on Signal Processing, Computing and Control (ISPCC)*, pp. 21-25.
- Wajid, M., Kumar, A. and Bahl, R. (2017), "Direction-of-arrival estimation algorithms using single acoustic vector-sensor", *2017 International Conference on Multimedia, Signal Processing and Communication Technologies (IMPACT)*, pp. 84-88.
- Wajid, M., Kumar, B., Goel, A., Kumar, A. and Bahl, R. (2019), "Direction of arrival estimation with uniform linear array based on recurrent neural network", *5th International Conference on Signal Processing, Computing and Control (ISPCC)*.
- Wajid, M., Kumar, A. and Bahl, R. (2020), "Direction estimation and tracking of coherent sources using a single acoustic vector sensor", *Archives of Acoustics*, Vol. 45.
- Wajid, M., Yadav, S. and Usman, M. (2020), "Multivariate quadratic regression based direction estimation of an acoustic source", *Journal of Acoustical Society of India*, Vol. 47 Nos 2-3, pp. 102-111.
- Wajid, M., Alam, F., Yadav, S., Khan, M.A. and Usman, M. (2020), "Support vector regression based direction of arrival estimation of an acoustic source", *2020 International Conference on Innovation and Intelligence for Informatics, Computing and Technologies (3ICT)*, IEEE, pp. 1-6.
- Xiao, X., Zhao, S., Nguyen, D., Zhong, X., Jones, D., Chng, E.S. and Li, H. (2014), "The NTU-ADSC systems for reverberation challenge 2014", *Proc. REVERB Challenge Workshop*.

- Yadav, S., Wajid, M. and Usman, M. (2020), *Support Vector Machine-Based Direction of Arrival Estimation with Uniform Linear Array*, Springer, Singapore.
- Zhang, C., Florêncio, D., Ba, D. and Zhang, Z. (2008), "Maximum likelihood sound source localization and beamforming for directional microphone arrays in distributed meetings", *IEEE Transactions on Multimedia*, Vol. 10 No. 3, pp. 538-548.
- Zhang, M., Pan, X., Shen, Y. and Qiu, J. (2021), "Deep learning-based direction-of-arrival estimation for multiple speech sources using a small scale array", *The Journal of the Acoustical Society of America*, Vol. 149, pp. 3841-3850, doi: [10.1121/10.0005127](https://doi.org/10.1121/10.0005127).
- Zhao, S., Chng, E., Hieu, N. and Li, H. (2010), "A robust real-time sound source localization system for olivia robot", *2010 APSIPA Annual Summit and Conference*.
- Zhao, S., Ahmed, S., Liang, Y., Rupnow, K., Chen, D. and Jones, D. (2012), "A real-time 3D sound localization system with miniature microphone array for virtual reality", *Industrial Electronics and Applications (ICIEA), 2012 7th IEEE Conference on*, pp. 1853-1857.
- Zhou, C., Gu, Y., Zhang, Y., Shi, Z., Jin, T. and Wu, X. (2017), "Compressive sensing-based coprime array direction-of-arrival estimation", *IET Communications*, Vol. 11 No. 11, pp. 1719-1724.

Corresponding author

Mahima Singh can be contacted at: mahidiwakar085@gmail.com

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgroupublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com