

2025-09-17

The self-archived postprint version of this journal article is available at Linköping University Institutional Repository (DiVA):

<https://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-217636>

Directional Sensitivity-Based DOA Estimation Using a Fourier Series Model

Gustav Zetterqvist, Fredrik Gustafsson and Gustaf Hendeby

IEEE Sensors Journal, (2025), 25(20), pp. 1-1.

Publisher: Institute of Electrical and Electronics Engineers

<https://doi.org/10.1109/jsen.2025.3604893>

N.B.: When citing this work, cite the original publication.

© 2025 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

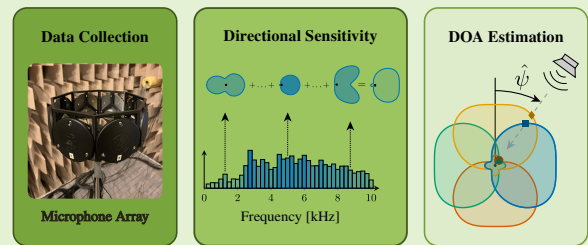
Directional Sensitivity-Based DOA Estimation Using a Fourier Series Model

Gustav Zetterqvist , Fredrik Gustafsson , *Fellow, IEEE*,
and Gustaf Hendeby , *Senior Member, IEEE*

Abstract— *Direction of arrival* (DOA) estimation is a fundamental problem in signal processing and has applications in various fields such as radar, sonar, and acoustic. In this paper, we propose a method for DOA estimation using the received power at each sensor. The method is based on the directional sensitivity of the sensor elements at various frequencies. We model the directional sensitivity using a *Fourier series* (FS) model, where the parametric model enables *Cramér-Rao lower bound* (CRLB) computations. The FS model is estimated from measurements of a wideband noise signal. To estimate the DOA, the measured power profile is compared to the FS model using the least squares method.

The proposed power-based method offers several advantages over classical time-delay methods, particularly in allowing arbitrarily small arrays and still handle broadband signals. Additionally, it enables low-rate sampling, which simplifies hardware requirements and significantly reduces processor load. In numerical evaluations with a microphone array and natural sound sources, we still benchmark our method against state-of-the-art time-delay methods. Real-world experiments show promising results, performing on par with the best of the other evaluated methods for all natural signals, despite relying on significantly less information. A key benefit is robustness against array size limitations. By utilizing the received signal power instead of time delays or phase information, the method enables small arrays with great DOA resolution. Furthermore, outdoor data collected a year after calibration confirms its robustness, demonstrating consistent performance over time.

Index Terms— DOA Estimation, Directional Sensitivity, Microphone Array, Fourier Series, CRLB, YALMIP



I. INTRODUCTION

The *direction of arrival* (DOA) of a sound source is an essential parameter in various applications, including surveillance, tracking, and localization. The estimation of DOA has been extensively studied for the past decades [1]–[3]. For a more recent study, refer to [4]. Traditional DOA estimators rely on sensor arrays and are based on time delays of the received signal at each microphone, known as *time-difference of arrival* (TDOA) measurements. Array processing techniques, such as *minimum variance distortionless response* (MVDR) beamforming [5], *multiple signal classification* (MUSIC) [6], *estimation of signal parameters via rotational invariance techniques* (ESPRIT) [7], and *multi channel cross-correlation coefficient* (MCCC) [8], [9], are commonly used for DOA estimation. These principles can be applied to essentially any

array geometry and signal waveform. The rule of thumb for classical methods is to separate the sensors with half the wavelength. With a fixed amount of sensors, placing them farther away will give aliasing (ambiguity in DOA) and too close will decrease frequency resolution and accuracy. As an example, for a broadband signal with frequency content in the interval 340–3400 Hz (corresponding to human speech), the sensor separation should be 5–50 cm. That is, it is inherently difficult to design one sensor array for broadband signals when using time-delay based methods [10]–[12].

Alternative approaches, such as sum-difference beamforming in radar systems [13], [14] and tensor-based techniques that exploit the multidimensional structure of signals [15], [16], have been introduced to enhance DOA resolution. Amplitude-comparison methods like sum-difference beamforming rely on coherent array processing and phase-preserving sum and difference beams, typically under narrowband radar assumptions with carefully structured antenna patterns. These methods do not extend naturally to compact broadband microphone arrays. In contrast, our method requires only received power measurements and introduces a frequency-dependent *Fourier series* (FS) model of the directional sensitivity, which enables *Cramér-Rao lower bound* (CRLB) analysis and practical DOA estimation with arbitrarily small arrays. Tensor-based ap-

G. Zetterqvist has received funding from ELLIIT. This work was partially funded by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation. The authors would like to thank Jonas Nordlöf at the Swedish Defence Research Agency (FOI) for the help and assistance during the data collection.

G. Zetterqvist, F. Gustafsson and G. Hendeby are with the Department of Electrical Engineering, Linköping University, 581 83 Linköping, Sweden (e-mail: gustav.zetterqvist@liu.se, fredrik.gustafsson@liu.se, gustaf.hendeby@liu.se).

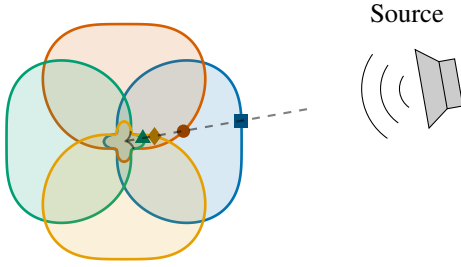


Fig. 1: Visualization of the concept with 4 microphones facing away from each others. The colorful lines represent the directional sensitivity of each microphone. In the figure the sound is arriving with a DOA of 10° with respect to the blue microphone, the dashed line represents the direction of the sound. Note that each microphone receives a different sound level, marked with dark-colored markers (■, ●, ▲, ◆) in the figure. The data for the directional sensitivity is from a wideband noise signal.

proaches, on the other hand, are powerful for multidimensional signal processing and coherent sources, but they generally require large arrays and phase information, leading to high computational cost. Our method complements tensor techniques by targeting scenarios where array size and sampling rate are heavily constrained.

Recently, there has been growing interest in DOA estimation using compact arrays with a limited number of microphones [17]–[19]. A key challenge with small arrays lies in the restricted sensor spacing, which significantly reduces the resolution and accuracy of classical time-delay-based methods. Moreover, most existing techniques inherently rely on spatial separation between sensors — a requirement that becomes impractical in small-form-factor applications.

To address this limitation, we previously proposed a method to estimate the DOA using only the received power at each microphone [12]. This approach enables simple hardware, low sampling frequency, and smaller arrays. The main limitation of the method is that the accuracy quickly deteriorates for signals with a different frequency content than the training signal. In this paper, we extend the method to estimate the DOA of signals with different frequency content. Additionally, the proposed method is benchmarked to state-of-the-art methods in both an anechoic chamber and an outdoor scene.

We propose a training phase where the directional sensitivity of each microphone element is estimated for different frequencies. The directional sensitivity is the power attenuation of the microphone as a function of the DOA angle. The directional sensitivity is estimated by emitting a wideband noise signal from different DOA angles and using the received power at each microphone to estimate a FS model of the directional sensitivity. The FS model is then used as a fingerprint of the observed power vector in a real-time estimator. The parametric model enables the use of the CRLB to evaluate the performance of the estimator. A visualization of the concept is shown in Fig. 1.

The contributions in this paper extend the work in [12] in the following ways:

- A frequency dependency is introduced to the signal model to estimate the DOA of signals with different frequency content.
- A logarithmic signal model is proposed and analyzed to

enhance the performance of the method.

- Estimation data is collected from an outdoor scene to show the robustness of the method.
- The proposed method is benchmarked to state-of-the-art methods in both an anechoic chamber and an outdoor scene.

To clarify the setup throughout the paper, we list the following assumptions made in the signal model and the proposed method:

- A1 The microphones are assumed to have directional sensitivity, either by design or due to the construction of the array.
- A2 The array is small enough to be approximated as a single point, *i.e.*, the time delays between microphones are negligible, unlike in traditional DOA methods.
- A3 The microphones measure the signal power directly, or compute it from the received signal.
- A4 There is only one sound source present in the environment, which is assumed to be a stationary point source.
- A5 Environmental reverberation and reflections are assumed to be none or minimal.
- A6 For ease of presentation, the DOA angle is assumed to lie in the plane of the array.

The paper is organized as follows: Section II introduces the signal model and establishes the assumed noise distribution. The proposed method is described in Section III, followed by the derivation of the CRLB in Section IV. Subsequently, Section V delves into the array design and experimental setup. The results of the proposed method are discussed in Section VI, along with a comparison to state-of-the-art methods. Finally, Section VII presents the conclusions from the study.

In order to ease reading and comprehension, a list of notations used throughout the paper is provided in Table I.

TABLE I

LIST OF NOTATIONS USED IN THE PAPER.

Notation	Description
y_n	Measured signal at microphone n
s_n	Signal at microphone n
v_n	Measurement noise in the signal y_n
σ_n	Noise standard deviation in the signal y_n
p_n	Measured power at microphone n
e_n	Measurement noise of the power measurement p_n
λ_n	Noise variance of the power measurement p_n
L	Number of samples in the measurement window of p_n
α	Absolute level of received power
g_n	Microphone gain at microphone n
$h_n(\psi)$	Directional sensitivity of microphone n at angle ψ
ψ	DOA angle of the sound source
θ_n	Vector of parameters for microphone n
$\Phi(\psi)$	FS vector for the angle ψ
P_n	Measurement model of the power at microphone n
\mathbf{x}	Vector of optimization variables
N	Number of microphones in the array
D	Order of the FS
K	Number of observed directions
M	Number of band-pass filters
$H_{BP}(f_m)$	Band-pass filter with frequency f_m
w_m	Frequency weight for band-pass filter m
$\bar{\cdot}$	Variables in logarithmic scale
\vdots	Estimated variables
\mathcal{I}_n	Fisher information matrix (FIM) for microphone n

II. SIGNAL MODEL

The signal model, based on [12], denotes the measured signal from microphone n at discrete time l with the scalar

$$y_n(l) = s_n(l) + v_n(l), \quad (1)$$

where $s_n(l)$ is the received signal and $v_n(l)$ is the measurement noise, which is assumed to be uncorrelated with $s_n(l)$, and normally distributed, *i.e.*, $v_n(l) \sim \mathcal{N}(0, \sigma_n^2)$. The power of the measured signal from microphone n can then be calculated as

$$p_n = \frac{1}{L} \sum_{l=1}^L y_n^2(l), \quad (2)$$

where L is the number of samples in the signal of interest. By inserting (1) into (2), the signal power can thus be expressed in three terms

$$p_n = \underbrace{\frac{1}{L} \sum_{l=1}^L s_n^2(l)}_{p_n^s} + \underbrace{\frac{1}{L} \sum_{l=1}^L 2s_n(l)v_n(l)}_{p_n^{sw}} + \underbrace{\frac{1}{L} \sum_{l=1}^L v_n^2(l)}_{e_n}, \quad (3)$$

where p_n^s is the power of the received signal, p_n^{sw} is a cross-term from the signal and the noise, and e_n is the power of the measurement noise. Here, the number of samples L is assumed to be large, hence $p_n^{sw} \rightarrow 0$. Further, since $v_n(t)$ is normally distributed, e_n will be chi-squared distributed with L degrees of freedom, *i.e.*, $\frac{L}{\sigma_n^2} e_n \sim \chi_L^2$. However, since the degrees of freedom is assumed to be large, the chi-squared distribution can be approximated with a normal distribution

$$\frac{L}{\sigma_n^2} e_n \sim \chi_L^2 \xrightarrow{\text{Approx}} e_n \sim \mathcal{N}\left(\sigma_n^2, \frac{2\sigma_n^4}{L}\right) = \mathcal{N}(\sigma_n^2, \lambda_n). \quad (4)$$

We assume that each microphone has a directional sensitivity in the power attenuation by design or from the construction of the array. Further, the absolute level of the received power, assumed to be the same at all sensors, is denoted α and is considered to be a nuisance parameter from a DOA estimation perspective. Thus, the power p_n measured by microphone n can be expressed as

$$p_n = \alpha g_n h_n(\psi) + e_n, \quad (5)$$

where g_n is the microphone gain, $h_n(\psi)$ is the amplification of the signal from direction ψ , hereinafter referred to as the directional sensitivity of microphone n . The noise is denoted e_n , and assumed to be normally distributed as described in (4). The microphone gain g_n is partly caused by the microphone's sensitivity and the gain of the pre-amplifier, and is assumed to be unknown.

During data collection, it was evident that the noise mean and variance depend on the sound level of the recorded sound, *i.e.*, a higher α yields a larger noise mean and variance. To account for this dependency, we transform the measurement model (5) into the logarithmic scale

$$\bar{p}_n = \bar{\alpha} + \bar{g}_n + \bar{h}_n(\psi) + \bar{e}_n, \quad (6)$$

where $\bar{p}_n = 10 \log_{10}(p_n)$, and the notation \bar{x} is introduced to refer to the parameter x in logarithmic scale. We assume that the noise is distributed according to

$$\bar{e}_n \sim \mathcal{N}(\bar{\sigma}_n^2, \bar{\lambda}_n). \quad (7)$$

In the logarithmic scale, the *signal-to-noise ratio* (SNR) is assumed to be sufficiently high such that the approximation

$$\log(x+y) = \log x + \log\left(1 + \frac{y}{x}\right) = \log x + \frac{y}{x} + \mathcal{O}\left(\left(\frac{y}{x}\right)^2\right) \quad (8)$$

holds, as the higher-order terms become negligible. This approximation is valid at sufficiently high SNR, and its use at low SNR remains a limitation that may reduce accuracy.

III. PROPOSED METHOD

The approach we propose is based on two phases, a training phase to estimate the parameters of the directional sensitivity for each microphone, and an estimation phase to estimate the DOA of the signal.

A. Fourier Series Model

Since the directional sensitivity is 2π periodic, $h_n(\psi) = h_n(\psi + 2\pi)$, a *Fourier series* (FS) is a natural parametric model structure for approximating this function based on calibration measurements. This approach eliminates the need for table lookups and interpolation. The FS model is defined as

$$\begin{aligned} h_n(\psi) &= h\left(\psi + \frac{n2\pi}{N}, \boldsymbol{\theta}_n\right) \\ &= \theta_0^n + \sum_{d=1}^D \theta_{d,c}^n \cos\left(d\left(\psi + \frac{n2\pi}{N}\right)\right) \\ &\quad + \theta_{d,s}^n \sin\left(d\left(\psi + \frac{n2\pi}{N}\right)\right) \\ &= \boldsymbol{\Phi}\left(\psi + \frac{n2\pi}{N}\right) \boldsymbol{\theta}_n, \end{aligned} \quad (9)$$

where $\boldsymbol{\theta}_n$ is a parameter vector of length $2D + 1$, D is the order of the FS, and

$$\begin{aligned} \boldsymbol{\Phi}(\psi) &= [1 \quad \cos(\psi) \quad \sin(\psi) \quad \dots \quad \cos(D\psi) \quad \sin(D\psi)] \\ \boldsymbol{\theta}_n &= [\theta_0^n \quad \theta_{1,c}^n \quad \theta_{1,s}^n \quad \dots \quad \theta_{D,c}^n \quad \theta_{D,s}^n]^T. \end{aligned}$$

The choice of D is a trade-off between model complexity and model fit. A higher order FS model can capture more complex directional sensitivity patterns, but also increases the risk of overfitting. However, a low order FS model can be too simple to capture the directional sensitivity and thus yield a poor model fit.

Besides being a compact representation of $h_n(\psi)$, a parametric model enables CRLB calculations for the DOA estimation.

B. Training Phase: Sensor Parameter Estimation

In the parameter estimation phase, the array is exposed to wideband noise from different directions ψ , and the measured signal y_n is observed in a controlled environment. Then, the power of the measured signal, p_n , is computed according to (2). This yields the measurement model

$$P_n(\psi; \alpha, g_n, \boldsymbol{\theta}_n) = \alpha g_n h\left(\psi + \frac{n2\pi}{N}, \boldsymbol{\theta}_n\right) + \sigma_n^2, \quad (10)$$

where α , g_n , and $\boldsymbol{\theta}_n$ are the parameters to be estimated, and σ_n^2 is the noise bias.

The parameters of the directional sensitivity, $\boldsymbol{\theta}_n$, as well as the microphone gain g_n are thereafter estimated using

YALMIP with the FMINCON solver [20]. The optimization problem is formulated as

$$\begin{aligned} & \underset{\mathbf{x}=\{\alpha, g_1, \dots, g_N, \theta_1, \dots, \theta_N\}}{\text{minimize}} && V(\mathbf{x}) \\ & \text{subject to} && \alpha > 0 \\ & && g_1 = 1 \\ & && h(0, \theta_n) = 1 \quad \forall n = 1, 2, \dots, N \end{aligned} \quad (11)$$

where N is the number of microphones and $V(\mathbf{x})$ is the loss function

$$V(\mathbf{x}) = \sum_{n=1}^N \frac{1}{\lambda_n} \sum_{k=1}^K (p_{n,k} - P_n(\psi_k; \alpha, g_n, \theta_n))^2, \quad (12)$$

where \mathbf{x} contains the optimization variables α , $\{g_1, \dots, g_N\}$ and $\{\theta_1, \dots, \theta_N\}$, and K is the number of observed directions. Measurements $p_{n,k}$ capture the power of the signal at microphone n when the sound source is located at direction ψ_k . The term σ_n^2 appears since the noise does not have zero mean, and is estimated from measurements of the background noise. To ensure the optimization problem has a unique solution, the constraints $g_1 = 1$ and $h(0, \theta_n) = 1$ are imposed, as the gain and directional sensitivity are otherwise defined only up to a scaling factor. The number of optimization variables in (11), *i.e.*, the dimensionality of \mathbf{x} , is $1 + N + N(2D + 1)$, where 1 is for α , N is for the microphone gains g_n , and $N(2D + 1)$ is for the parameters θ_n of the directional sensitivity. Thus, to ensure a unique solution, the number of measurements KN must be larger than the number of optimization variables, *i.e.*, $KN > 1 + N + N(2D + 1)$. However, in practice, the more measurements the better, as this will yield a more accurate estimate of the directional sensitivity.

Similarly, the logarithmic scale model can be derived by expressing the power of the measured signal in decibels as, $\bar{p}_n = 10 \log_{10}(p_n)$, resulting in the measurement model

$$\bar{P}_n(\psi; \bar{\alpha}, \bar{g}_n, \bar{\theta}_n) = \bar{\alpha} + \bar{g}_n + h(\psi + \frac{n2\pi}{N}, \bar{\theta}_n) + \bar{\sigma}_n^2. \quad (13)$$

The corresponding optimization problem is then formulated as

$$\begin{aligned} & \underset{\bar{\mathbf{x}}=\{\bar{\alpha}, \bar{g}_1, \dots, \bar{g}_N, \bar{\theta}_1, \dots, \bar{\theta}_N\}}{\text{minimize}} && V(\bar{\mathbf{x}}) \\ & \text{subject to} && \bar{g}_1 = 0 \\ & && h(0, \bar{\theta}_n) = 0 \quad \forall n = 1, 2, \dots, N \end{aligned} \quad (14)$$

with the loss function

$$V(\bar{\mathbf{x}}) = \sum_{n=1}^N \frac{1}{\bar{\lambda}_n} \sum_{k=1}^K (\bar{p}_{n,k} - \bar{P}_n(\psi; \bar{\alpha}, \bar{g}_n, \bar{\theta}_n))^2, \quad (15)$$

where $\bar{\mathbf{x}}$ contains the optimization variables $\bar{\alpha}$, $\{\bar{g}_1, \dots, \bar{g}_N\}$ and $\{\bar{\theta}_1, \dots, \bar{\theta}_N\}$.

C. Online DOA Estimation

In the estimation phase, the DOA is estimated using the directional sensitivity model and the measured power of the signal. We formulate the problem as a *nonlinear least squares*

(NLS) problem without constraints to enhance the computational efficiency. The estimation method is applicable both in the absolute and logarithmic scale, and is solved using the *separable least squares* (SLS) method [21], [22].

1) *Absolute scale*: From the training, the model for microphone n can be expressed as

$$P_n(\psi; \alpha, \hat{g}_n, \hat{\theta}_n) = \alpha \underbrace{\hat{g}_n \Phi(\psi + \frac{n2\pi}{N}) \hat{\theta}_n}_{\hat{h}_n(\psi)} + \sigma_n^2, \quad (16)$$

where \hat{g}_n and $\hat{\theta}_n$ are the estimated parameters from the training phase from (11). This model can be seen as linear in α and nonlinear in ψ , and thus the SLS approach can be applied to estimate these parameters.

The DOA estimator can then be computed using SLS as

$$\hat{\psi} = \arg \min_{\psi} V(\psi, \hat{\alpha}(\psi)) \quad (17a)$$

$$\hat{\alpha}(\psi) = \arg \min_{\alpha} V(\psi, \alpha), \quad (17b)$$

where the cost function is defined as

$$V(\psi, \alpha) = \sum_{n=1}^N \left(p_n - P_n(\psi; \alpha, \hat{g}_n \hat{\theta}_n) \right)^2. \quad (18)$$

This can be rephrased as

$$\hat{\psi} = \arg \max_{\psi} \frac{\left(\sum_{n=1}^N \hat{h}_n(\psi) (p_n - \sigma_n^2) \right)^2}{\sum_{n=1}^N \hat{h}_n^2(\psi)} \quad (19a)$$

$$\hat{\alpha} = \frac{\sum_{n=1}^N \hat{h}_n(\psi) (p_n - \sigma_n^2)}{\sum_{n=1}^N \hat{h}_n^2(\psi)}, \quad (19b)$$

where $\hat{h}_n(\psi) = \hat{g}_n \Phi(\psi + \frac{n2\pi}{N}) \hat{\theta}_n$ denotes the nonlinear part of the model, as shown in [21, p. 257].

2) *Logarithmic scale*: Similarly, the logarithmic model for microphone n can be expressed as

$$\bar{P}_n(\psi; \bar{\alpha}, \hat{g}_n, \hat{\theta}_n) = \bar{\alpha} + \underbrace{\hat{g}_n + \Phi(\psi + \frac{n2\pi}{N}) \hat{\theta}_n}_{\hat{h}_n(\psi)} + \bar{\sigma}_n^2, \quad (20)$$

where \hat{g}_n is the estimated microphone gain and $\hat{\theta}_n$ is the estimated parameters from the training phase from (14). This model can be seen as linear in $\bar{\alpha}$ and nonlinear in ψ .

The DOA estimator can then be computed using SLS as

$$\hat{\psi} = \arg \min_{\psi} V(\psi, \hat{\bar{\alpha}}(\psi)) \quad (21a)$$

$$\hat{\bar{\alpha}}(\psi) = \arg \min_{\bar{\alpha}} V(\psi, \bar{\alpha}), \quad (21b)$$

where the cost function is defined as

$$V(\psi, \bar{\alpha}) = \sum_{n=1}^N \left(\bar{p}_n - \bar{P}_n(\psi; \bar{\alpha}, \hat{g}_n, \hat{\theta}_n) \right)^2. \quad (22)$$

This can be rephrased as

$$\hat{\psi} = \arg \min_{\psi} \sum_{n=1}^N \left(\bar{p}_n - \bar{P}_n(\psi; \hat{\bar{\alpha}}(\psi), \hat{g}_n, \hat{\theta}_n) \right)^2 \quad (23a)$$

$$\hat{\bar{\alpha}}(\psi) = \frac{1}{N} \sum_{n=1}^N \left(\bar{p}_n - \left(\hat{h}_n(\psi) + \bar{\sigma}_n^2 \right) \right), \quad (23b)$$

where $\hat{h}_n(\psi) = \hat{g}_n + \Phi(\psi + \frac{n2\pi}{N}) \hat{\theta}_n$ is the nonlinear part of the model, as shown in [22, p. 66].

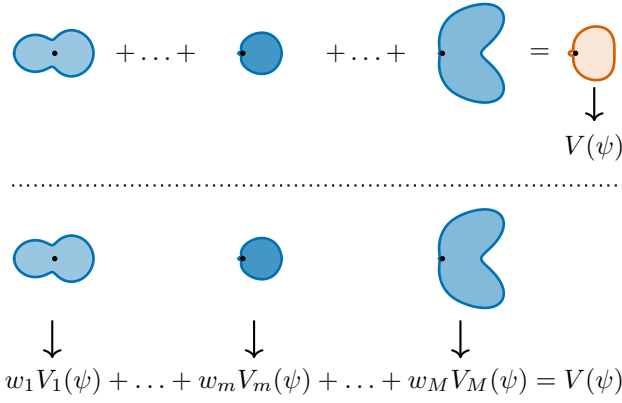


Fig. 2: Visualization of the different methods. In the upper one, we sum all the band-pass filtered models with weights w_m to create a model of the incoming sound based on its frequency content. In the lower visualization, we calculate the cost of each band-pass separately and sum these terms with the weights w_m . The opacity in the figure represent the weights w_m .

D. Frequency dependency

During data collection it was observed that the directional sensitivity highly varies with the frequency content of the signal. Therefore, a dependency of the frequency is introduced in the parameters of the directional sensitivity, *i.e.*, $\theta_n(f)$, where f is the frequency content of the signal.

1) *Training*: In the training phase, when learning the directional sensitivity, we thus need to include the frequency dependency. One way to do this is by band-pass filtering a wideband noise signal through M number of adjacent band-pass filters in order to span the whole spectrum

$$y_n(l, f_m) = H_{BP}(f_m) \star y_n(l), \quad \forall f_m = \{f_1, \dots, f_M\} \quad (24a)$$

$$p_{m,n} = \frac{1}{L} \sum_{l=1}^L y_n^2(l, f_m), \quad (24b)$$

where $H_{BP}(f_m)$ denotes a band-pass filter with frequency content f_m and \star denotes the convolution operation. Thereafter, the optimization problems (11) and (14) can be solved for the signal power in each band-pass filtered signal, separately. Thus, for each band-pass frequency and microphone, a model for the directional sensitivity is learned.

2) *Estimation*: For the frequency dependency the estimation can be addressed in various ways, and one way is to create a model of the directional sensitivity by computing the frequency content of the received sound. Another way is to compare a band-pass filtered version of the received sound with the directional sensitivity for the corresponding band-pass filter, for all band-pass filters. Both these methods are visualized in Fig. 2.

a) *Combine into one directional sensitivity*: First, we compute the frequency content of the incoming sound in the bins of the filter bank, denoted w_m

$$w_m = \sum_{n=1}^N \frac{1}{N} \left(\sum_{f \in f_m} \left| \sum_{l=1}^L y_n[l] e^{-j2\pi f \frac{l}{L}} \right|^2 \right). \quad (25)$$

These weights must be calculated for each sound signal input to the system. To model the directional sensitivity of the incoming sound, we sum the models for the different frequencies with the weight of the frequency content, that is

$$\begin{aligned} P_n(\psi; \alpha, \hat{g}_{m,n}, \hat{\theta}_n(f)) \\ = \alpha \underbrace{\sum_{m=1}^F w_m \hat{g}_{m,n} \Phi\left(\psi + \frac{n2\pi}{N}\right) \hat{\theta}_n(f_m) + \sigma_n^2}_{\hat{h}_n(\psi)}. \end{aligned} \quad (26)$$

This can be seen as a SLS problem, which is linear in α and nonlinear in ψ . Therefore, (19a) can be applied.

The frequency weights are introduced as an estimate of the SNR of the incoming sound in the different frequency bins. It can be shown that the weights are proportional to the inverse of the noise variance in the frequency bins, and thus the weights are used to enhance the signal in the frequency bins with low noise variance.

b) *Each band-pass separately*: The second approach is to treat each band-pass filter separately, and then sum them in the cost function. The signal model for microphone n and band-pass filter m is expressed as

$$\begin{aligned} P_{m,n}(\psi; \alpha, \hat{g}_{m,n}, \hat{\theta}_n(f_m)) \\ = \alpha_m \underbrace{\hat{g}_{m,n} \Phi\left(\psi + \frac{n2\pi}{N}\right) \hat{\theta}_n(f_m) + \sigma_{m,n}^2}_{\hat{h}_{m,n}(\psi)}. \end{aligned} \quad (27)$$

The DOA estimate is then be calculated according to

$$\hat{\psi} = \arg \max_{\psi} \sum_{m=1}^F w_m V_m(\psi), \quad (28a)$$

$$V_m(\psi) = \frac{\left(\sum_{n=1}^N \hat{h}_{m,n}(\psi) (p_{m,n} - \sigma_{m,n}^2) \right)^2}{\sum_{n=1}^N \hat{h}_{m,n}^2(\psi)}, \quad (28b)$$

$$\hat{\alpha}_m = \frac{\sum_{n=1}^N \hat{h}_{m,n}(\psi) (p_{m,n} - \sigma_{m,n}^2)}{\sum_{n=1}^N \hat{h}_{m,n}^2(\psi)}. \quad (28c)$$

This method can also be performed in the logarithmic scale using the signal model

$$\begin{aligned} \bar{P}_{m,n}(\psi; \bar{\alpha}_m, \hat{g}_{m,n}, \hat{\theta}_n(f_m)) \\ = \bar{\alpha}_m + \underbrace{\hat{g}_{m,n} + \Phi\left(\psi + \frac{n2\pi}{N}\right) \hat{\theta}_n(f_m)}_{\hat{h}_{m,n}(\psi)} + \bar{\sigma}_{m,n}^2. \end{aligned} \quad (29)$$

The DOA estimate is then be calculated as

$$\hat{\psi} = \arg \min_{\psi} \sum_{m=1}^F w_m V_m(\psi), \quad (30a)$$

$$V_m(\psi) = \sum_{n=1}^N \left(\bar{p}_{m,n} - \bar{P}_{m,n}(\psi; \hat{\alpha}_m, \hat{g}_{m,n}, \hat{\theta}_n(f_m)) \right)^2, \quad (30b)$$

$$\hat{\alpha}_m(\psi) = \frac{1}{N} \sum_{n=1}^N \left(\bar{p}_{m,n} - \left(\hat{h}_{m,n}(\psi) + \bar{\sigma}_{m,n}^2 \right) \right). \quad (30c)$$

As the logarithmic scale assumes a high SNR, only the frequency bins with a SNR above 20 dB are used in the DOA estimation.

IV. CRAMÉR-RAO LOWER BOUND

With the signal model defined in Section II, the CRLB can be derived for the DOA angle ψ .

A. Absolute Scale

To calculate the CRLB, a *probability density function* (PDF) of the measurement is required. Since the noise e_n is assumed to be normally distributed as $e_n \sim \mathcal{N}(\sigma_n^2, \lambda)$, the PDF of the measurement is denoted as

$$\mathcal{P}(p_n|\psi, \boldsymbol{\theta}_n, \alpha, \lambda) = \mathcal{N}(\alpha g_n h(\psi, \boldsymbol{\theta}_n) + \sigma_n^2, \lambda), \quad (31)$$

where $\lambda = 2\sigma_n^4/L$ is the noise variance and assumed to be the same for all microphones.

This yields the FIM

$$\begin{aligned} \mathcal{I}_n(\psi, \alpha) &= -\mathbb{E} \left[\begin{pmatrix} \frac{\partial^2}{\partial \psi^2} & \frac{\partial^2}{\partial \psi \partial \alpha} \\ \frac{\partial^2}{\partial \alpha \partial \psi} & \frac{\partial^2}{\partial \alpha^2} \end{pmatrix} \log \mathcal{P}(p_n|\psi, \boldsymbol{\theta}_n, \alpha, \lambda) \right] \\ &= \frac{1}{\lambda} \begin{bmatrix} (\alpha g_n h'_{\psi,n})^2 & \alpha g_n^2 h_n h'_{\psi,n} \\ \alpha g_n^2 h_n h'_{\psi,n} & g_n^2 h_n^2 \end{bmatrix}, \end{aligned} \quad (32)$$

where h_n denotes $h(\psi, \boldsymbol{\theta}_n)$ and $h'_{\psi,n} = \frac{\partial h(\psi, \boldsymbol{\theta}_n)}{\partial \psi}$.

Assuming that the estimate in (19a) is unbiased and that the PDF $\mathcal{P}(\mathbf{p}|\psi)$ satisfies the regularity conditions, the CRLB can be used to evaluate the performance of the estimation method [21]. Thus, the CRLB of ψ is computed as

$$\begin{aligned} \text{var}(\hat{\psi}) \geq \text{CRLB}(\psi) &= \left[\left(\sum_{n=1}^N \mathcal{I}_n(\psi, \alpha) \right)^{-1} \right]_{1,1} \\ &= \frac{\lambda}{\alpha^2} \frac{\|\mathbf{h}\|^2}{\left(\|\mathbf{h}\|^2 \|\mathbf{h}'_{\psi}\|^2 - \langle \mathbf{h}, \mathbf{h}'_{\psi} \rangle^2 \right)}, \end{aligned} \quad (33)$$

where $\|\cdot\|$ denotes the Euclidean norm of a vector and $\langle \cdot, \cdot \rangle$ denotes the inner product of two vectors, and

$$\mathbf{h} = [g_1 h_1 \quad g_2 h_2 \quad \dots \quad g_N h_N]^T, \quad (34)$$

$$\mathbf{h}'_{\psi} = [g_1 h'_{\psi,1} \quad g_2 h'_{\psi,2} \quad \dots \quad g_N h'_{\psi,N}]^T. \quad (35)$$

Note that the denominator is strictly positive as long as \mathbf{h} and \mathbf{h}'_{ψ} are not parallel according to Cauchy-Schwarz inequality. The term α^2/λ can be interpreted as the SNR of the signal.

B. Logarithmic scale

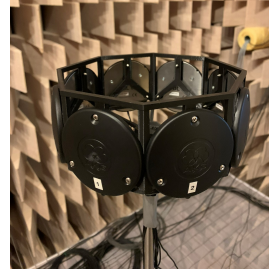
Using the logarithmic scale, the PDF of the measurement is described by

$$\mathcal{P}(\bar{p}_n|\psi, \bar{\boldsymbol{\theta}}_n, \bar{\alpha}, \bar{\lambda}) = \mathcal{N}(\bar{\alpha} + \bar{g}_n + h(\psi, \bar{\boldsymbol{\theta}}_n) + \bar{\sigma}_n^2, \bar{\lambda}), \quad (36)$$

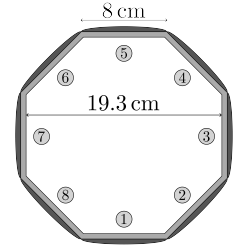
where $\bar{\lambda} = 2\bar{\sigma}_n^4/L$ is the noise variance assumed to be the same for all microphones.

This yields the FIM

$$\begin{aligned} \mathcal{I}_n(\psi, \bar{\alpha}) &= -\mathbb{E} \left[\begin{pmatrix} \frac{\partial^2}{\partial \psi^2} & \frac{\partial^2}{\partial \psi \partial \bar{\alpha}} \\ \frac{\partial^2}{\partial \bar{\alpha} \partial \psi} & \frac{\partial^2}{\partial \bar{\alpha}^2} \end{pmatrix} \log \mathcal{P}(\bar{p}_n|\psi, \bar{\boldsymbol{\theta}}_n, \bar{\alpha}, \bar{\lambda}) \right] \\ &= \frac{1}{\bar{\lambda}} \begin{bmatrix} (\bar{h}'_{\psi,n})^2 & \bar{h}'_{\psi,n} \\ \bar{h}'_{\psi,n} & 1 \end{bmatrix}, \end{aligned} \quad (37)$$



(a) The sensor array in the anechoic chamber.



(b) Top view of the dimension and microphone placement of the sensor array.

Fig. 3: An overview of the sensor array used in this paper.

where $\bar{h}'_{\psi,n} = \frac{\partial h(\psi, \bar{\boldsymbol{\theta}}_n)}{\partial \psi}$.

Thus, the CRLB of ψ is computed as

$$\begin{aligned} \text{var}(\hat{\psi}) \geq \text{CRLB}(\psi) &= \left[\left(\sum_{n=1}^N \mathcal{I}_n(\psi, \bar{\alpha}) \right)^{-1} \right]_{1,1} \\ &= \bar{\lambda} \frac{1}{\left(\sum_{n=1}^N (\bar{h}'_{\psi,n})^2 - \frac{1}{N} \left(\sum_{n=1}^N \bar{h}'_{\psi,n} \right)^2 \right)}, \end{aligned} \quad (38)$$

where $\bar{\mathbf{h}}'_{\psi}$ denotes a vector of the derivatives $\bar{h}'_{\psi,n}$. Note that the CRLB does not depend on the SNR of the signal in this case, only the noise variance and the derivatives of $h(\psi, \bar{\boldsymbol{\theta}}_n)$. This is due to the fact that we have different assumptions on the noise distribution in the two cases.

We can also calculate the CRLB for the wideband noise signal as

$$\text{CRLB}_{wn}(\psi) = \left[\left(\sum_{m=1}^M \bar{w}_m \mathcal{I}_m(\psi, \alpha_m) \right)^{-1} \right]_{1,1}, \quad (39)$$

where \bar{w}_m is a normalized version of the weights w_m , such that $\bar{w}_m = w_m / \max w_m$, i.e., $\max \bar{w}_m = 1$. This weighing is similar to weighing the variance of the noise in the CRLB calculation, $\lambda_m = \lambda/w_m$, i.e., the noise variance is higher for the frequencies with lower weights.

V. EXPERIMENTAL ARRAY DESIGN

To evaluate the method, a hardware prototype consisting of eight microphones has been designed and used to collect data in an anechoic chamber and outdoors. We use the data collected in the anechoic chamber from [12], and the outdoor data were collected in December 2023.

A. Hardware Prototype

The sensor array used in this paper is a *uniform circular array* (UCA) in the form of an octagon with eight microphones. The microphones used are the omnidirectional CBL99 from AKG [23] connected to a Behringer UMC1820 preamplifier [24]. The preamplifier is connected to a computer which allows for a sampling frequency of up to 96 kHz. The array and its dimensions are shown in Fig. 3.

B. Anechoic Measurements

Data were collected in an anechoic chamber, at the Swedish Defence Research Agency (FOI) in Linköping in May 2022, ISO rated from 100 Hz to 10 kHz. The sensor array was put on a turntable, and a Genelec 1029A speaker [25] was placed at a distance of 5 meters from the center of the array. In the setup, microphone 1 was facing the speaker at $\psi = 0^\circ$, microphone 2 at $\psi = -45^\circ$ and so on. Thereafter, data were recorded with a sampling frequency of 48 kHz using 19 different types of signals.

- 1) *Wideband noise* with a bandwidth of 100 Hz–10 kHz that lasted for 10 seconds.
- 2–11) *Sinusoids* with frequencies of 500–9500 Hz separated with 1 kHz for 1 second each.
- 12) *Hovering drone* sound during 5 seconds.
- 13) *Elephant trumpet* of an Indian elephant lasting 2.39 seconds [26].
- 14) *Police siren* lasting 10.48 seconds [27].
- 15) *Female scream* lasting 2.2 seconds [28].
- 16) *Attenuated wideband noise*, *i.e.*, the wideband noise signal with half the amplitude.
- 17) *Amplified wideband noise*, *i.e.*, the wideband noise signal with twice the amplitude.
- 18) *Gunshot* of a Smith & Wesson Chief's Special 9mm revolver shot with blanks inside the anechoic chamber.
- 19) *Background noise* collected during 26.41 seconds.

C. Outdoor Measurements

Data were also collected in an outdoor environment at Linköping University in December 2023. The weather conditions were calm with a light breeze and the temperature was around -4°C , with approximately 10 centimeters of snow on the ground. The microphone array was put in the middle, and then a circle was drawn around it with a radius of 5 m. Thereafter, points separated by 3.09 meters were then measured and marked, leading to 10 different points separated by 36° . The speaker, a Genelec 8030C [29], was then moved to these points. In total, data from 10 different angles was collected using the same type of signals as in the anechoic chamber, with the exception of the gunshot.

D. Simulated Data

In order to validate the noise assumption, as well as the method, simulated data has been generated. First, the signal is generated using the signal model in (1) to validate the derivation of the noise distribution in (4). Thereafter, data is simulated using the signal model in (29) and the noise distribution in (7). This is done to evaluate the method and compare it to the CRLB.

VI. ESTIMATION RESULTS

This section presents the results of the proposed method for DOA estimation. The method is evaluated using both anechoic and outdoor data, and the results are compared to state-of-the-art methods.

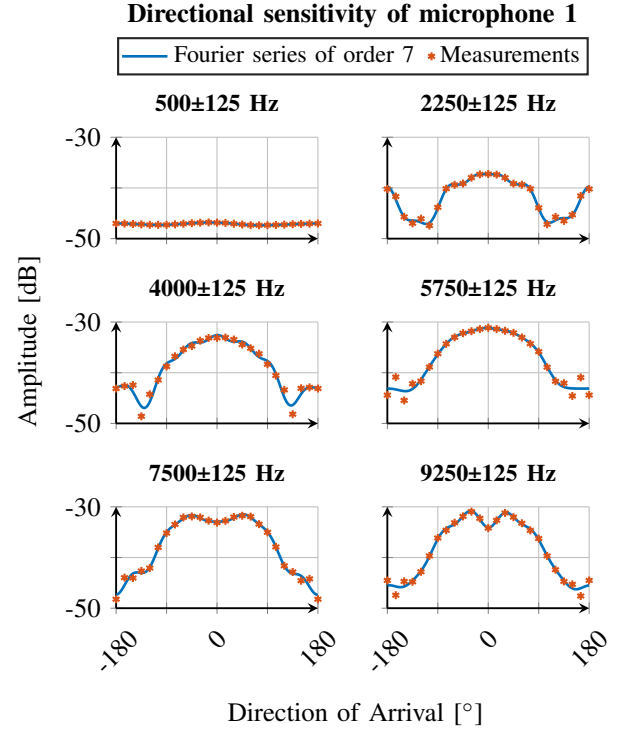


Fig. 4: The measured power of the band-pass filtered wideband noise signal, and the corresponding FS of order 7 as a function of the DOA for microphone 1, both in the logarithmic scale.

A. Training

For the training, the wideband noise signal is measured from 24 angles uniformly distributed around the array in the anechoic chamber. The signal is band-pass filtered into 40 different frequency bands with a bandwidth of 250 Hz, centered around the frequencies 250 Hz, 500 Hz, ..., 10 kHz. Then, the signal power of the band-pass filtered signals is computed according to (2). Thereafter, the noise variances, $\sigma_n^2(f_m)$ and $\bar{\sigma}_n^2(f_m)$, are computed from measurements of the band-pass filtered background noise. Finally, the optimization problems (11) and (14) are solved to estimate the model parameters.

B. Fourier Series Fit

In the optimization problem, different orders of the FS are examined to find the best one. As mentioned, the choice of model order is a trade-off between model fit and model complexity. Higher model orders also leads to overfitting, which is undesirable. By experimental evaluation, a model order of 7 is a good trade-off for both the absolute and the logarithmic model. The measured power of the band-pass filtered wideband noise signal as well as the corresponding FS of order 7 in logarithmic scale is illustrated in Fig. 4. It is worth noting that below 1000 Hz the directional sensitivity is uniform for all angles. Hence, in this interval there is not much information to estimate the DOA from.

C. DOA Estimation

For the DOA estimation, all signals are used except for the background noise. The validation data from the anechoic

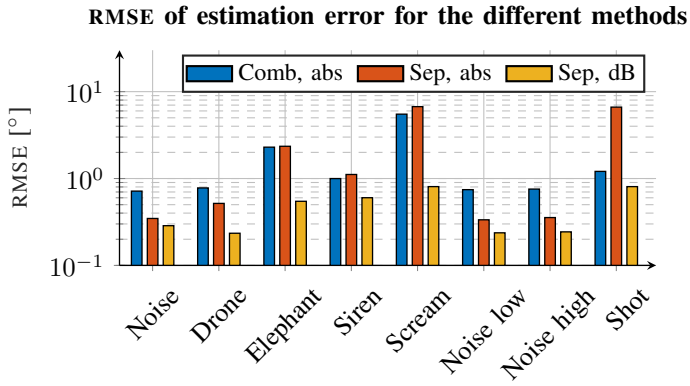


Fig. 5: Bar graph of the performance for the different methods for all signals except the sinusoids, all models have an order of 7. Comb refers to the combined method in Section III-D.2.a, and Sep refers to the separate method in Section III-D.2.b. The y -axis is in logarithmic scale.

chamber, was collected at 24 angles uniformly distributed around the microphone array, but at different angles than for the training data. For the gunshots, the data was collected at 8 angles distributed around the microphone array. The data from the outdoor scene was collected at 10 angles uniformly distributed around the microphone array.

Then, the power of the signals is calculated and compared to the estimated FS model using the *least squares* (LS) formulation from (19a). For the frequency dependent model, 40 band-pass filters were used with a bandwidth of 250 Hz in order to get the results. The performance for the different methods are presented in Fig. 5.

Here, it is seen that all signals are estimated with good precision. It is clearly noted that the method where we sum the band-pass version of the incoming signal in logarithmic scale performed best for all signals. With this method, all signals have a *root mean square error* (RMSE) value below 1° . This method will thus be referred to as the proposed method in the following sections.

We also compare the performance from the proposed method to state-of-the-art methods, where the results for the anechoic data are presented in Fig. 6a, while the data from the outdoor experiments are presented in Fig. 6b. The violin plots show the distribution of the estimation error for the different methods, where the width of the violin corresponds to the density of the data, and the white dot represents the median value. Each colored dot represents an individual estimation error for a specific signal, where the color corresponds to the signal number. For the narrowband methods, *i.e.*, MUSIC and MVDR, the same band-pass filtering steps as our method has been performed

$$y_n^m(l) = H_{BP}(f_m) \star y_n(l), \quad (40)$$

$$V_{\text{MUSIC}}^m(\psi) = \text{MUSIC}(y_{1:N}^m), \quad (41)$$

$$V_{\text{MUSIC}}(\psi) = \sum_m w_m V_{\text{MUSIC}}^m(\psi). \quad (42)$$

where $H_{BP}(f_m)$ is a narrowband band-pass filter around frequency f_m , and w_m is the same frequency weights as in our method [30]. Since the state-of-the-art methods rely on

the phase information, the signal is passed through the Hilbert transform to obtain the complex-valued signal [31].

For the anechoic data, the proposed method performs similar to the MCCC for the natural signals, *i.e.*, Signal 1 and 12-18. While for the pure sinusoids, MCCC performs better. Compared to MUSIC and MVDR beamformer, the proposed method outperforms these methods for all signals, except for the low frequency sinusoid. This is especially evident for the 500 Hz sinusoid, where the proposed method fails to provide accurate estimates. By looking at the direction sensitivity at 500 Hz in Fig. 4, it is seen that the directional sensitivity is almost uniform for all angles, which explains the poor performance of the proposed method. Further, it is clearly noted that the narrowband methods perform worse for higher frequencies, especially with frequencies that do not fulfil the spatial Shannon sampling theorem, *i.e.*, $\Delta_d \leq \frac{\lambda}{2} \rightarrow f \lesssim 2100$ Hz [10].

In the results for the outdoor data, it is seen that the proposed method performs similar to the MCCC for most of the signals. Our method outperforms MCCC for some sinusoids, especially the high frequency sinusoids. For the outdoor data, both MUSIC and MVDR beamformer perform worse than the proposed method for all signals, except for the low frequency sinusoid. The resulting RMSE values for the different methods are presented in Table II.

The reason why the estimation error is higher for the pure sinusoids is probably caused by the choice of band-pass filters. With a bandwidth of 250 Hz, the band-pass filters are too wide for the pure sinusoids, and thus the directional sensitivity is not well estimated. However, for the natural signals, the band-pass filters are well suited, and the directional sensitivity is well estimated. It is worth noting that all natural signals have a majority of the frequency content above 1000 Hz, which includes more information in the directional sensitivity.

It is also noted that for the outdoor data the estimation error of all methods has a bias of around 2.5° . This is most likely due to the fact that the setup is not perfectly calibrated, which is hard to achieve in an outdoor environment. This bias could be removed by retraining the model with additional outdoor data, but this is not the focus of this paper.

The SNR for the anechoic data ranged between 22.2 dB and 68.0 dB, where the gunshot had the highest SNR, otherwise the highest SNR was 59.7 dB. For the outdoor data the SNR ranged between -14.1 dB and 32.8 dB, and thus the outdoor data had approximately 30 dB lower SNR than the anechoic data.

D. CRLB vs Simulated RMSE

The CRLB is evaluated at each filter-bank separately and is then compared to the RMSE of simulated signal power, the results are presented in Fig. 7. To calculate the RMSE of the sinusoids, we simulate the signal power for the sinusoids using the model in (36) with $\bar{\lambda} = 0.1^2$ using 100 *Monte Carlo* (MC) simulations. The number of MC simulations is chosen such that the variance of the RMSE is reasonably low, thus it can be concluded that the RMSE follows the CRLB.

The CRLB for the wideband noise signal can also be calculated as in (39), the resulting CRLB along with the RMSE

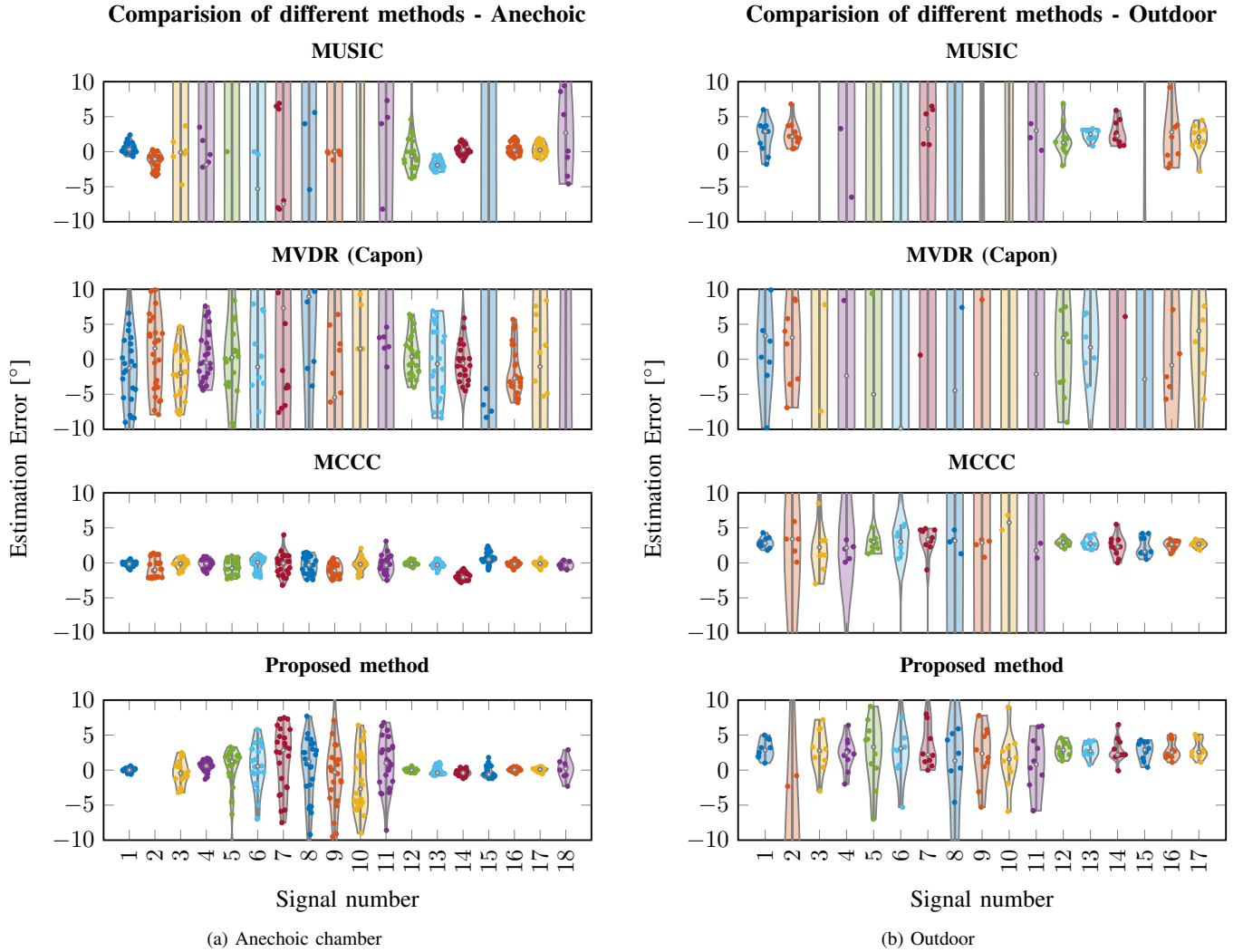


Fig. 6: Violin plot of the estimation error for the different signals using various methods for the data collected in the different settings. The signal number corresponds to the signals described in Section V-B, where 1, 12–18 are natural signals.

TABLE II

MEAN RMSE VALUES IN DEGREES FOR THE DIFFERENT METHODS ACROSS THE SIGNALS, BOTH FOR THE ANECHOIC AND OUTDOOR DATA. THE SIGNAL NUMBER CORRESPONDS TO THE SIGNALS DESCRIBED IN SECTION V-B, WHERE 1, 12–18 ARE NATURAL SIGNALS.

	Signal No.	Proposed	MUSIC	MVDR	MCCC
Anechoic	1–17	11.64	60.04	27.07	0.99
	2–11	19.48	92.77	31.95	1.17
	1, 12–17	0.46	15.33	22.51	0.79
	18	1.46	9.68	92.43	0.58
Outdoor	1–17	5.18	74.25	76.08	42.03
	2–11	6.60	108.43	92.57	69.49
	1, 12–17	3.14	29.15	59.23	2.78

from simulated data are presented in Fig. 8. To calculate the RMSE of the wideband noise signal, we simulate the signal power for the sinusoids using the model in (36) with $\bar{\lambda} = 0.1^2$ using 150 MC simulations. The frequency weights w_m are the same as for the measured wideband signal, using the anechoic data.

From the comparison between the CRLB and the RMSE it

is seen that the RMSE seem to follow the CRLB using the simulated data, both for the sinusoids and the wideband noise signal. This is expected since the simulated data is generated using the same model as the CRLB is calculated for.

Further, it is noted that the CRLB is lower for the wideband noise signal than for the sinusoids, which is expected since the wideband noise signal has more information in the directional sensitivity than the sinusoids.

E. CRLB and RMSE as a function of $\bar{\lambda}$

Usually, the performance of the method is evaluated using the RMSE of the estimated DOA as a function of the SNR. However, as mentioned in Section IV, the CRLB is not dependent on the SNR but rather on the noise variance $\bar{\lambda}$. Since $\bar{\lambda}$ and the SNR are inversely related ($\text{SNR} \propto 1/\bar{\lambda}$), we can equivalently evaluate the performance of the method as a function of $\bar{\lambda}$.

The CRLB is calculated for different values of $\bar{\lambda}$, and the RMSE is calculated from simulated data using the same model as in (29). For the RMSE calculation, we generated 10 MC simulations for each value of $\bar{\lambda}$, and the DOA angles are

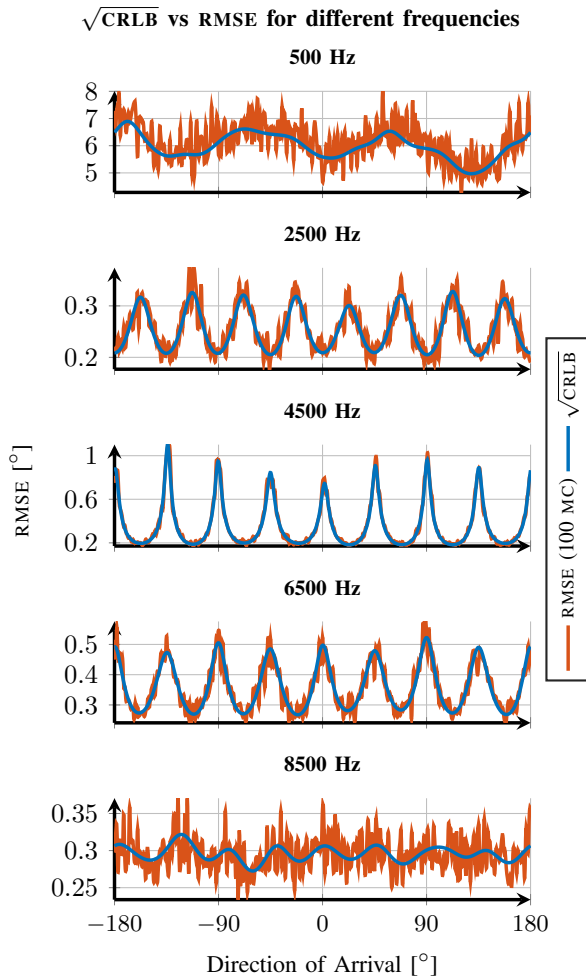


Fig. 7: $\sqrt{\text{CRLB}}$ compared to the RMSE of the sinusoidal signals for simulated data. The blue line represent the CRLB and the red line the RMSE. The signals power were simulated using $\bar{\lambda} = 0.1^2$.

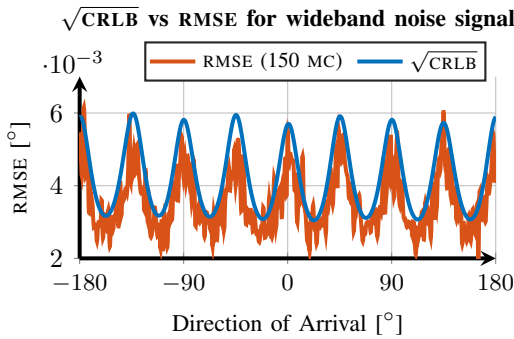


Fig. 8: $\sqrt{\text{CRLB}}$ compared to the RMSE of the wideband noise signal for simulated data. The blue line represent the CRLB and the red line the RMSE. The signal power were simulated using $\bar{\lambda} = 0.1^2$.

uniformly distributed with a step size of 5° from -180° to 180° , thus resulting in 72 different DOA angles. The resulting CRLB and RMSE are presented in Fig. 9.

As seen in Fig. 9, the RMSE follows the CRLB for the different values of $\bar{\lambda}$, as expected. Also, the RMSE is lower for higher values of $1/\bar{\lambda}$, which is also expected since the SNR is higher for lower values of $\bar{\lambda}$. In the figure, it is also

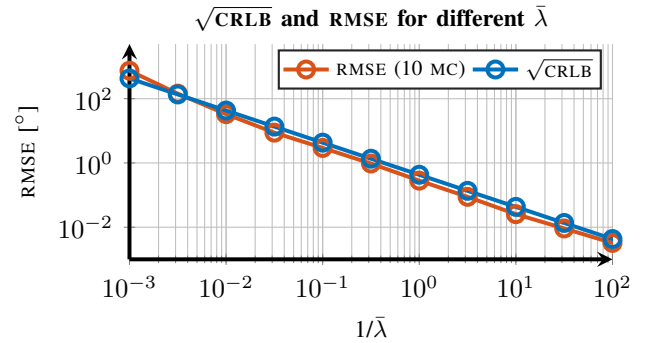


Fig. 9: $\sqrt{\text{CRLB}}$ compared to the RMSE of the wideband noise signal for simulated data, using different values of $\bar{\lambda}$. The blue line represent the CRLB and the red line the RMSE.

noted that the RMSE is marginally lower than the CRLB for all values of $\bar{\lambda}$, this might be due to wrong assumptions in the CRLB calculation for the wideband noise signal in (39), *i.e.*, $\max \bar{w}_m = 1$ may be slightly too pessimistic.

VII. CONCLUSIONS

In this paper, we have proposed an extension to the method in [12] to estimate the DOA of signals with arbitrary frequency content. To show the potential, a sensor array with eight microphones has been used to estimate the DOA of 18 different types of signals. First, a FS model of the directivity pattern of each microphone was derived for several band-pass filtered versions of a wideband noise signal emitted from different DOA angles. We use measurements in logarithmic scale to get a more robust model, as the noise mean and variance was shown to depend on the sound level. Then, the received signal power for the band-pass filtered signal at each microphone was compared to the model.

In an anechoic setting, the method performs well compared to state-of-the-art methods for the natural signals evaluated, all with the majority of the frequency content above 1000 Hz. Less good results are obtained for lower frequency content, as the directivity pattern becomes uniform for low frequencies. In theory, the proposed method avoids degradation of the performance with the size of the array, enabling small arrays with great DOA resolution. However, this needs to be validated experimentally using a smaller array. The outdoor data shows that the model of the directional sensitivity is fairly robust over time and temperature changes, as the performance is great even after a year.

Future work includes to investigate how reverberations affect the algorithm, and whether they can be used to map the environment or to localize the sound source. Moreover, the model error should be investigated to understand the limitations of the method. It would also be interesting to investigate if the method can be used to estimate the 2D-DOA of a sound source, *i.e.*, the azimuth and elevation angles. Thus, relying on the directional sensitivity of the microphones. Lastly, the method should be tested on more complex signals, such as speech, to evaluate the performance in real-life scenarios.

REFERENCES

- [1] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 4, pp. 320–327, 1976.
- [2] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 67–94, 1996.
- [3] P. Stoica, P. Babu, and J. Li, "SPICE: A Sparse Covariance-Based Estimation Method for Array Processing," *IEEE Transactions on Signal Processing*, vol. 59, no. 2, pp. 629–638, 2011.
- [4] C. Rascon and I. Meza, "Localization of sound sources in robotics: A review," *Robotics and Autonomous Systems*, vol. 96, pp. 184–210, 2017.
- [5] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.
- [6] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.
- [7] R. Roy and T. Kailath, "Esprit-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 984–995, 1989.
- [8] J. Benesty, J. Chen, and Y. Huang, Eds., *Microphone Array Signal Processing*. Berlin, Heidelberg: Springer-Verlag, 2008.
- [9] J. Benesty, J. Chen, and Y. Huang, "Time-delay estimation via linear interpolation and cross correlation," *IEEE Transactions on Speech and Audio Processing*, vol. 12, no. 5, pp. 509–519, 2004.
- [10] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Pearson Prentice Hall, 2005.
- [11] C. Shannon, "Communication in the Presence of Noise," *Proceedings of the IRE*, vol. 37, no. 1, pp. 10–21, 1949.
- [12] G. Zetterqvist, F. Gustafsson, and G. Hendeby, "Using received power in microphone arrays to estimate direction of arrival," in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1–5.
- [13] X. Wan, Q. Xiao, Y. Z. Zhang, Y. Li, J. Eisenbeis, J. W. Wang, Z. A. Huang, H. X. Liu, T. Zwick, and T. J. Cui, "Reconfigurable sum and difference beams based on a binary programmable metasurface," *IEEE Antennas and Wireless Propagation Letters*, vol. 20, no. 3, pp. 381–385, 2021.
- [14] A. F. Morabito and P. Rocca, "Optimal synthesis of sum and difference patterns with arbitrary sidelobes subject to common excitations constraints," *IEEE Antennas and Wireless Propagation Letters*, vol. 9, pp. 623–626, 2010.
- [15] X. Qu, Y. Lou, Y. Zhao, Y. Lu, and G. Qiao, "Augmented tensor music for doa estimation using nested acoustic vector-sensor array," *IEEE Signal Processing Letters*, vol. 29, pp. 1624–1628, 2022.
- [16] M. Haardt, F. Roemer, and G. Del Galdo, "Higher-order SVD-based subspace estimation to improve the parameter estimation accuracy in multidimensional harmonic retrieval problems," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3198–3213, 2008.
- [17] A. Rahamanand and B. Kim, "Sound source localization in 2d using a pair of bio-inspired mems directional microphones," *IEEE Sensors Journal*, vol. 21, no. 2, pp. 1369–1377, Jan. 2021.
- [18] A. Waqar, A. Khan, B. Kim, and D. Park, "Ormia ochracea inspired single-microphone approach for 3-d sound localization," *IEEE Sensors Letters*, vol. 8, no. 8, pp. 1–4, Aug. 2024.
- [19] A. Rahaman and B. Kim, "Microscale devices for biomimetic sound source localization: A review," *Journal of Microelectromechanical Systems*, vol. 31, no. 1, pp. 9–18, Feb. 2022.
- [20] J. Löfberg, "YALMIP : A Toolbox for Modeling and Optimization in MATLAB," in *In Proceedings of the CACSD Conference*, Taipei, Taiwan, 2004.
- [21] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice-Hall, Inc., 1993.
- [22] F. Gustafsson, *Statistical Sensor Fusion*. Studentlitteratur AB, 2018.
- [23] AKG, "CBL99 Hemispherical boundary layer microphone," accessed: 2022-10-05. [Online]. Available: <https://www.akg.com/support/CBL99.html>
- [24] Behringer, "U-PHORIA UMC1820," accessed: 2022-10-05. [Online]. Available: <https://www.behringer.com/product.html?modelCode=P0B2J>
- [25] Genelec, "Genelec 1029A Data Sheet," 2003, accessed: 2022-10-05. [Online]. Available: https://assets.ctfassets.net/4zjzn055a4v/7547Y1zhZuQeaO68yMCYGe/98dd2e2995803eae08cddc2709797e38/1029A_datasheet.pdf
- [26] Quicksounds, "Indian Elephant 1," accessed: 2024-05-02. [Online]. Available: <https://quicksounds.com/sound/3446/indian-elephant-1>
- [27] Videvo, "Police Siren Variou TE2027601," accessed: 2024-05-02. [Online]. Available: <https://www.videvo.net/sound-effect/police-siren-variou-te2027601/253680/>
- [28] Soundjay, "Woman Scream 01," accessed: 2024-05-02. [Online]. Available: <https://www.soundjay.com/human/sounds/woman-scream-01.mp3>
- [29] Genelec, "Genelec 8030C Manual," 2017, accessed: 2024-01-08. [Online]. Available: https://assets.ctfassets.net/4zjzn055a4v/1my4dzvqUoe4oKYwS6KsUu/a8f964c3ce07eb802b88b8d861799a07/8030c_opman_gb.fi.pdf
- [30] S. Argentieri and P. Danes, "Broadband variations of the music high-resolution method for sound source localization in robotics," in *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, Oct. 2007, pp. 2009–2014.
- [31] L. Marple, "Computing the discrete-time "analytic" signal via FFT," *IEEE Transactions on Signal Processing*, vol. 47, no. 9, pp. 2600–2603, 1999.



Gustav Zetterqvist received the B.Sc. degree as well as the M.Sc. degree in applied physics and electrical engineering from Linköping University, Linköping, Sweden, in 2018 and 2020, respectively.

He is currently pursuing the Ph.D. degree in automatic control with the Department of Electrical Engineering, Linköping University. His research interests include signal processing, sensor fusion, and array processing with applications to acoustic and seismic signals.



Fredrik Gustafsson (Fellow, IEEE) received the M.Sc. degree in electrical engineering and the Ph.D. degree in automatic control from Linköping University, Linköping, Sweden, in 1988 and 1992, respectively.

He has been a Professor in Sensor Informatics with the Department of Electrical Engineering, Linköping University, since 2005. His research interests are in stochastic signal processing, adaptive filtering and change detection, with applications to communication, vehicular,

airborne, and audio systems. He is a Co-Founder of the companies NIRA Dynamics (automotive safety systems), Softube (audio effects), and Senion (indoor positioning systems).

Prof. Gustafsson was an Associate Editor for IEEE Transactions of Signal Processing in 2000–2006, IEEE Transactions on Aerospace and Electronic Systems in 2010–2012, and EURASIP Journal on Applied Signal Processing in 2007–2012. He was the recipient of the Arnberg prize by the Royal Swedish Academy of Science (KVA) 2004, elected member of the Royal Academy of Engineering Sciences (IVA) 2007, and elevated to IEEE Fellow 2011. He was also the recipient of the Harry Rowe Mimno Award 2011 for the tutorial “Particle Filter Theory and Practice with Positioning Applications,” which was published in the AESS Magazine in July 2010, and was coauthor of “Smoothed state estimates under abrupt changes using sum-of-norms regularization” that received the Automatica paper prize in 2014.



Gustaf Hendeby (Senior Member, IEEE) received the M.Sc. degree in applied physics and electrical engineering and the Ph.D. degree in automatic control from Linköping University, Linköping, Sweden, in 2002 and 2008, respectively.

He worked as a Senior Researcher at German Research Center for Artificial Intelligence (DFKI) (2009–2011), and a Senior Scientist at the Swedish Defense Research Agency (FOI), Kista, Sweden, and held an Adjunct Associate

Professor position at Linköping University (2011–2015). He is an Associate Professor and a Docent with the Division of Automatic Control, Department of Electrical Engineering, Linköping University. He is the author of several published articles and conference papers in the area. He has experience of both theoretical analysis and implementation aspects. His main research interests include stochastic signal processing and sensor fusion with applications to nonlinear problems, target tracking, and simultaneous localization and mapping (SLAM).

Dr. Hendeby has been an Associate Editor of IEEE Transactions on Aerospace and Electronic Systems in the area of target tracking and multisensor systems since 2018. In 2022, he served as the General Chair for the 25th IEEE International Conference on Information Fusion (FUSION), Linköping.