MIRACLE - Microphone Array Impulse Response Dataset for Acoustic Learning

Adam Kujawski[*,1,2] Art J. R. Pelling[†,1,2] Ennes Sarradj[‡,1]

[*]Email: adam.kujawksi@tu-berlin.de, ORCID: 0000-0003-4579-8813
[†]Email: a.pelling@tu-berlin.de, ORCID: 0000-0003-3228-6069
[‡]Email: ennes.sarradj@tu-berlin.de
[1] *Department of Engineering Acoustics, TU Berlin, Einsteinufer 25, 10587, Berlin, Germany*
[2] *These authors contributed equally to this work.*

Abstract: This work presents a large dataset of measured impulse responses of spatially distributed sources in a plane parallel to a planar microphone array in an anechoic environment. It can be used for various acoustic machine learning tasks and as a benchmark for data-driven modelling and interpolation methods. The dataset, which we call MIRACLE, contains a total of $856,128$ single-channel impulse responses across four different measurement scenarios. A regular grid of 64×64 source locations was sampled for two different source plane to array distances. The dataset also contains measurements on a densely sampled 33×33 grid for the short distance, as well as measurements with the presence of a reflective panel for the longer distance. We assess the quality of the provided source location labels and provide directivity measurements of the loudspeaker used for excitation.

Keywords: impulse response, dataset, microphone array, acoustics

Novelty statement: We provide a large dataset of spatially distributed multichannel impulse response measurements together with a thorough assessment of the source location accuracy.

1. Introduction

The availability of high-quality datasets is paramount in the evolving fields of machine learning for acoustic source localization and quantification, data-driven model order reduction, as well as sound field reconstruction. However, without adaptation to or training with realistic data, the performance of data-driven methods can be significantly impaired [1, 3] and the availability of openly available datasets containing rich and realistic experimental data is limited. This work fills this gap by introducing a large measured spatial room impulse response (SRIR) dataset which we call "**M**icrophone Array **I**mpulse **R**esponse Dataset for **Ac**oustical **Le**arning" (MIRACLE).

The key highlights are:

1. MIRACLE is the first SRIR dataset explicitly designed for acoustic testing applications using a planar microphone array focused on a rectangular observation area.
2. With a total of $856,128$ captured spatial room impulse responses and dense spatial sampling of the observation area, the dataset is well suited for machine learning and deep learning applications because numerous acoustic mulit-source scenarios with possibly closing neighboring sources can be created by superimposing signals that have been convolved with the provided SRIRs.
3. In contrast to most of the previously published datasets, the accuracy of the positional labels is statistically assessed. The assessment reveals outstandingly precise source locations with positional uncertainties in the order of a few millimeters. Hence the dataset can also serve as a validation benchmark for (reduced-order) modelling methods.

Data availability
The dataset presented in this paper can be obtained from

doi:10.14279/depositonce-18318

2. Materials and Methods

2.1. Experimental Setup

The experimental setup is illustrated in Figure 1. Details on the utilized hardware are given in Table 4.

Environment: The measurements were performed in the anechoic chamber of TU Berlin (room volume $V = 830\,\mathrm{m}^3$, lower cut-off frequency $f_\mathrm{c} = 63\,\mathrm{Hz}$). Neither heating nor air conditioning was active during the measurements. The temperature was monitored throughout at the microphone array center throughout the experiment.

Microphone Array: The phased microphone setup features a planar microphone array comprising $n_\mathrm{o} = 64$ channels mounted in a $1.5\,\mathrm{m}\times 1.5\,\mathrm{m}$ aluminum plate. The microphone arrangement follows Vogel's spiral [10]. The maximum pairwise distance between the array microphones is referred to as the aperture size $d_\mathrm{a} = 1.47\,\mathrm{m}$.

Sound Source: A dynamic 2" cone loudspeaker in a cylindrical 3D-printed enclosure was employed as the sound source. An exponential sweep with a duration of 3 s was used as the excitation signal. Due to the limited frequency range of the loudspeaker, the sweep was limited to $f \in [100, 16000]\,\mathrm{Hz}$.

Positioning: A high-precision motor-driven 2D positioning system was employed for loudspeaker positioning. The positioning system and the microphone array were manually aligned by with a laser distance meter and a cross-line laser level, achieving only minor alignment errors of a few millimeters at worst. During data post-processing, a spatial offset correction was applied based on a statistical evaluation of the spatial source strength maxima obtained by conventional microphone-array beamforming methods. Two different spatial configurations of source plane and microphone array were considered in the measurement campaign.

Acoustical Treatment: The supporting grid platform and the positioning system were cladded with absorptive foam in order to minimize reflections.

Data Acquisition: The microphone array data was acquired with a multichannel acquisition system (sampling rate: 51.2 kHz). In addition to the sensor signals, a loopback of the excitation signal was recorded as a reference signal for the post-processing.

2.2. Main Experiment

A customized and fully automated data acquisition procedure was implemented. Before each experiment, the loudspeaker was repeatedly excited with the excitation signal for a duration of 20 minutes (the duration was determined in a dedicated experiment). This warm-up phase accounts for the non-stationary dynamics of the transfer function related to internal temperature changes. Subsequently, the actual measurement routine was started by positioning the loudspeaker at the desired source location and measuring the room temperature simultaneously. After positioning, two repetitions of background noise measurement and loudspeaker excitation measurements were performed using all n_o microphones at once. Subsequently, the cross-correlation between all n_i recorded channels was evaluated according to the *the rule of two* [7]. Based on the measured sweep signals and the noise signal, the rule of two defines a cross-correlation threshold at which a pair of measured sweeps can be regarded free of corruption. In case of any violations, the measurement was repeated automatically. Finally, the measured data was stored redundantly on local and remote databases.

2.3. Directivty Measurement

Following the measurement campaign, an additional measurement was conducted in the anechoic chamber to obtain the angle dependent transfer function of the loudspeaker at discrete azimuth angles at a resolution of $\Delta\theta = 2.5°$. A microphone was placed at a distance of 0.5 m from the loudspeaker center. The latter was mounted on a motor-driven dispersion measurement turntable. A photograph of the measurement setup can be found in Figure 2. The same excitation signal and processing parameters as in the previous measurement campaign were used to determine the loudspeaker impulse response. Due to the cylindrical enclosure enclosing the loudspeaker, rotational symmetry around the z-axis can be assumed.

2.4. Sweep Synthesis

An exponential sine sweep was used as excitation signal because of its favourable properties in regards to crest factor and rejection of non-linearities [8]. It was designed according to [2] in the frequency range of the loudspeaker, namely from 100 Hz to 16 kHz. Because the anechoic chamber is nearly free of reflections and has very now noise levels, relatively short sweep times can be used to for measurement. However, short excitation signals also reduce the frequency resolution, especially towards the lower end of the spectrum. Since the sweep time is the main contributor to the total time of the experiment a sweep time of 3 s was chosen which resulted in a total time of about 20 hours for the large experiment scenarios. In order to ensure that the entire system response after excitation is captured, a safety window of 250 ms was added to the recording duration. At the chosen sampling time of $f_s = 51.2\,\mathrm{kHz}$, this resulted in $n_s = 166,400$ samples per measurement.

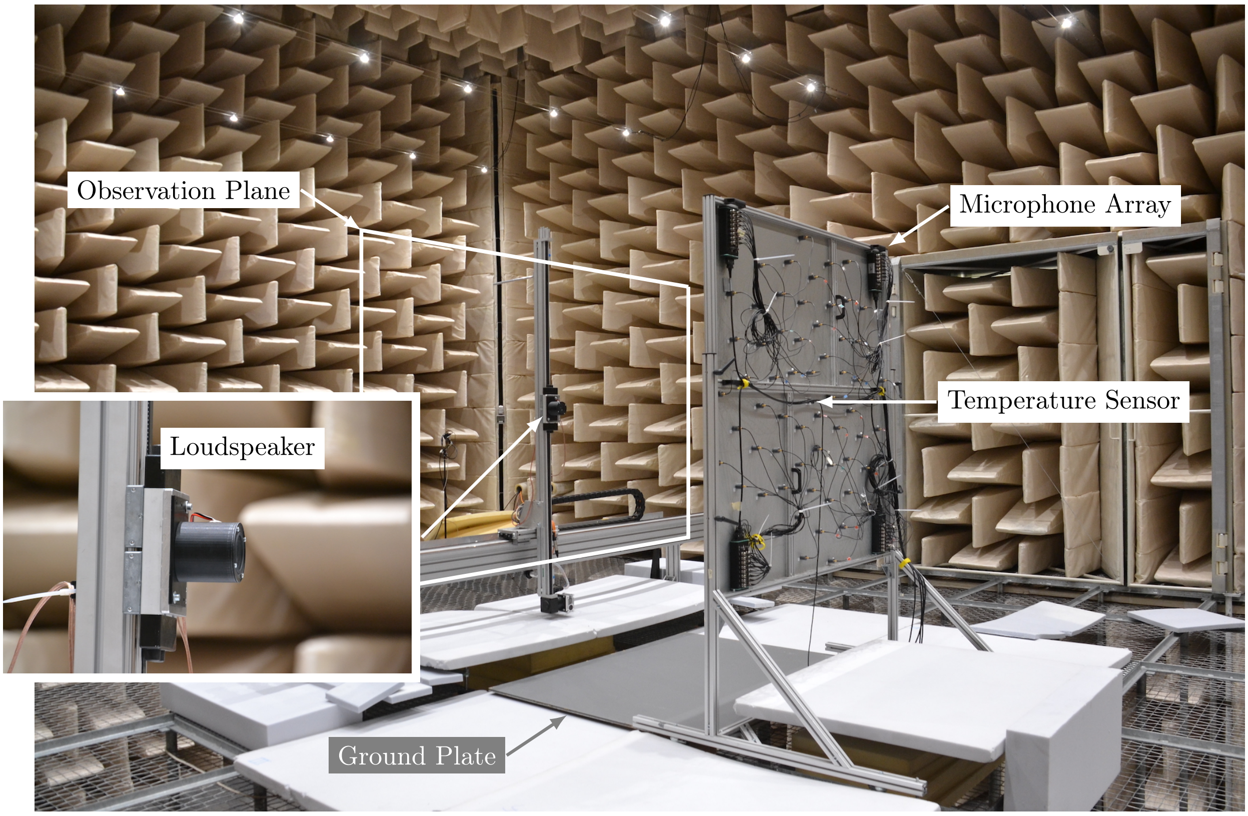

Figure 1: Experimental setup for the main experiment (R2) with reflective ground plate.

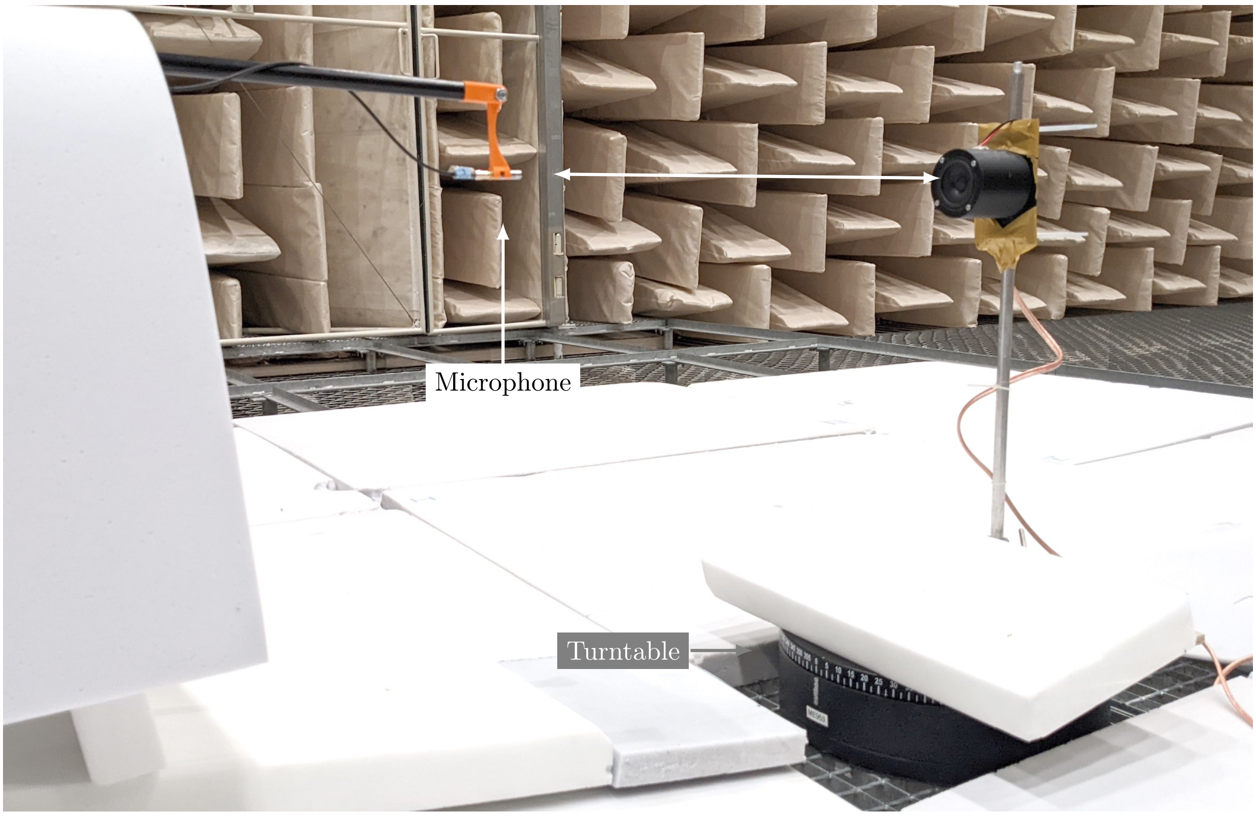

Figure 2: Experimental setup for the directivity measurement.

3. Results

A total of four different experimental scenarios were realized. They are are summarized in Table 1 and the corresponding mean and standard deviation of temperature and the speed of sound is given in Table 2.

3.1. Post Processing

Several post-processing steps were performed to obtain a good estimate for the system's impulse response from the measurements. Firstly, the two measurement repetitions were averaged in order to obtain a single (loopback) excitation signal $\tilde{u}_{i,j} \in \mathbb{R}^{n_s}$ and microphone signal $\tilde{y}_{i,j} \in \mathbb{R}^{n_s}$ at the i-th source to the j-th receiver location, respectively. According to that, all signals were downsampled to a sampling rate of $f_\mathrm{d} = 32\,\mathrm{kHz}$ since the loudspeaker transmission capability and excitation sweep have an upper frequency limit of 16 kHz.

Deconvolution

In the following, let $n_\mathrm{d} = 104,000$ denote the number of samples after resampling. An estimate of the transfer function was obtained by dividing the spectra of the averaged and downsampled measurement signals $\mathrm{DFT}(y_{i,j}) \in \mathbb{C}^{n_\mathrm{d}}$ by the corresponding spectrum of the averaged and downsampled loopback excitation signals $U_{i,j} = \mathrm{DFT}(u_{i,j}) \in \mathbb{C}^{n_\mathrm{d}}$, i.e.

$$H_{i,j}\left(e^{-\imath\omega_k}\right) = Y_{i,j}\left(e^{-\imath\omega_k}\right) U^{-1}\left(e^{-\imath\omega_k}\right) \in \mathbb{C},$$

for $\omega_k = \pi(k - n_\mathrm{d})/n_\mathrm{d}$. The inverse spectra $U_{i,j}^{-1} \in \mathbb{C}^{n_\mathrm{d}}$ were obtained by regularized inversion [4]

$$U_{i,j}^{-1}\left(e^{-\imath\omega_k}\right) = \frac{U_{i,j}^{*}\left(e^{-\imath\omega_k}\right)}{U_{i,j}^{*}\left(e^{-\imath\omega_k}\right) U_{i,j}\left(e^{-\imath\omega_k}\right) + M\lambda\left(e^{-\imath\omega_k}\right)},$$

where $M = \max_{k\in\{1\,\ldots,\,n_\mathrm{d}\}}\{|U_{i,j}\left(e^{-\imath\omega_k}\right)|^2\} = 1$ and the regularization term was chosen as

$$\lambda\left(e^{-\imath\omega_k}\right) = \begin{cases} 1 & \text{for} \quad |\omega_k| \in [0,\, \omega_\mathrm{fade}] \\ \frac{1+\cos\left(\frac{\omega_\mathrm{fade}-|\omega_k|}{\omega_\mathrm{fade}-\omega_\mathrm{cut}}\right)}{2} & \text{for} \quad |\omega_k| \in [\omega_\mathrm{fade},\, \omega_\mathrm{cut}] \\ 0 & \text{for} \quad |\omega_k| \in [\omega_\mathrm{cut},\, \pi] \end{cases}$$

such that the regularization term $\lambda\left(e^{-\imath\omega_k}\right)$ is equal to 0 above the cutoff frequency

$$\omega_\mathrm{cut} = 2\pi\frac{100\,\mathrm{Hz}}{f_\mathrm{d}}$$

and equal to 1 above $\omega_\mathrm{fade} = \frac{\omega_\mathrm{cut}}{\sqrt{2}}$. A crossfade based on a Hann window (raised-cosine) is used to smoothly transition in between. The estimate of the transfer function $H_{i,j}$ was then transformed back to the time domain to finally obtain the impulse response

$$h_{i,j} = \mathrm{DFT}^{-1}\left(H_{i,j}\right).$$

Truncation

The calculated impulse responses were subsequently truncated in order to contain the size of the final dataset. For user convenience, the impulse responses of all measurement scenarios were truncated identically. For this, the minimum cumulative energy $e \in \mathbb{R}^{n_\mathrm{d}}$ given by

$$e(t) = \min_{i\in n_\mathrm{i}, j\in n_\mathrm{o}} \sum_{\tau=1}^{t} |h_{i,j}(\tau)|^2, \quad t \in \{1,\,\ldots,\,n_\mathrm{d}\},$$

was calculated for each scenario. The truncation index n_t was chosen to be the smallest power of two that is larger than the time index for which 0.1% of the energy is truncated at worst, namely

$$n_t = 1024 \geq \tilde{t} = \underset{t\in\{1,\,\ldots,\,n_\mathrm{d}\}}{\arg\max} \left\{e(t) \leq 0.999\,\|e\|_\infty\right\}.$$

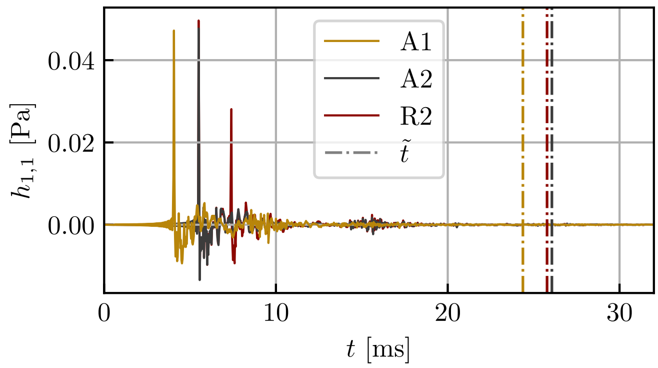

Figure 3: Measured impulse responses of the first source to first receiver location for the experiments A1, A2, and R2. The dash-dotted vertical lines indicate the respective truncation index.

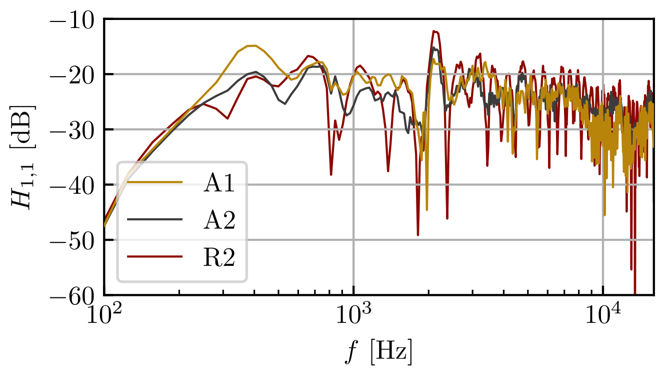

Figure 4: Measured transfer functions of the first source to first receiver location for the experiments A1, A2, and R2.

Table 1: MIRACLE experimental scenarios.

Scenario	Anechoic	n_i	n_o	$d_x = d_y$	$\Delta d_x = \Delta d_y$	d_z
A1	✓	$64^2 = 4096$	64	146.7 cm	23.3 mm	73.4 cm
D1	✓	$33^2 = 1089$	64	16.0 cm	5.0 mm	73.4 cm
A2	✓	$64^2 = 4096$	64	146.7 cm	23.3 mm	146.7 cm
R2	✗	$64^2 = 4096$	64	146.7 cm	23.3 mm	146.7 cm

Table 2: Mean μ and standard deviation σ of the temperature and speed of sound for each experiment.

Scenario	Temperature	[°C]	Speed of Sound	[m/s]
A1	$\mu = 22.0$	$\sigma = 0.1$	$\mu = 344.7$	$\sigma = 0.1$
D1	$\mu = 22.3$	$\sigma = 0.2$	$\mu = 344.8,$	$\sigma = 0.1$
A2	$\mu = 22.7$	$\sigma = 0.05$	$\mu = 345.0$	$\sigma = 0.03$
R2	$\mu = 22.9$	$\sigma = 0.02$	$\mu = 345.2$	$\sigma = 0.01$

3.2. Directivity of the Loudspeaker

The same post-processing steps as described in Section 3.1 were applied to the measurement data of the loudspeaker. Figure 5 shows the directivity D and the directivity index DI of the loudspeaker measured with a dispersion measurement turntable in the azimuthal plane. In this work, the directivity is defined as the ratio between the measured squared sound pressure $p_{\mathrm{RMS}}(\theta, f)$ at an angle θ and the maximum among all angles, i.e.

$$\mathrm{D}(\theta, f) = 10 \log_{10} \left(\frac{p_{\mathrm{RMS}}(\theta, f)}{\max_{\phi \in [0, 2\pi]} p_{\mathrm{RMS}}(\phi, f)} \right),$$

The directivity index under the assumption of rotational symmetry is expressed as

$$\mathrm{DI}(f) = 10 \log_{10} \left(\frac{4\pi p^2_{\mathrm{RMS}}(0, f)}{2\pi \int_0^\pi p^2_{\mathrm{RMS}}(\phi, f) \sin(\phi)\, \mathrm{d}\phi} \right),$$

whereby $p^2_{\mathrm{RMS}}(0, f)$ represents the squared sound pressure in front of the speaker.

3.3. Positional Validation and Offset Correction

Several uncertainty factors affected the spatial alignment precision regarding the microphone array center and the center of the observation area. These factors include measurement uncertainties with regard to the utilized cross-line laser and distance meter, and mechanical backlash, which occurred primarily with changes in direction on the horizontal axis of the positioning system. Therefore, a systematic spatial offset within the range of a few millimeters can be assumed.

In order to assess and correct the source positions with regard to the spatial offset, the latter needs to be determined accurately. Due to the anechoic environment and the use of a large-scale microphone array enabling an excellent spatial resolution, Conventional Frequency Domain Beamforming [5] serves as an appropriate method to obtain an estimate of the true source location. The large number of acoustic cases also permits a statistical approach to determine the spatial offset and the uncertainty regarding the source positions.

Beamforming

Let $\omega_k = \pi(k - n_\mathrm{d})/n_\mathrm{d}$ and let

$$H(e^{-\imath\omega_k}) = \begin{bmatrix} H_{i,1}(e^{-\imath\omega_k}) & \dots & H_{i,n_\mathrm{o}}(e^{-\imath\omega_k}) \end{bmatrix} \in \mathbb{C}^{n_\mathrm{o}}$$

denote the transfer function measurement from the i-th source at location x_s for $i \in \{1, \dots, n_\mathrm{i}\}$ to each of the n_o microphones. The cross-spectral matrix (CSM) induced by a sound source with unit strength is then defined as

$$C(\omega_k) = H(e^{-\imath\omega_k}) H(e^{-\imath\omega_k})^* \in \mathbb{C}^{n_\mathrm{o} \times n_\mathrm{o}}.$$

The beamforming result for an assumed source location $x_\mathrm{s} \in \mathbb{R}^3$ is then given by the square of the C-weighted norm of the steering vector $a(x_\mathrm{s}, \omega_k) \in \mathbb{C}^{n_\mathrm{o}}$, i.e.

$$b(x_\mathrm{s}, \omega_k) = \|a(x_\mathrm{s}, \omega_k)\|^2_{C(\omega_k)} = a(x_\mathrm{s}, \omega_k)^* C(\omega_k) a(x_\mathrm{s}, \omega_k).$$

Many different formulations of the steering vector can be found in the literature. The formulations I and IV in [9] result in a coincidence of the beamformer's steered response power maximum and the actual source location for a single monopole source radiating under free-field conditions. In this work, formulation IV was used which defines the entries of a via

$$\{a(x_\mathrm{s}, \omega)\}_j = \frac{e^{-\imath\omega(r_j - r_0)/c}}{r_j \sqrt{n_\mathrm{o} \sum_{k=1}^{n_\mathrm{o}} r_k^{-2}}},$$

where $r_j = \|x_\mathrm{s} - x_j\|_2$ is the distance between the assumed source location x_s and the j-th microphone location x_j, and $r_0 = \|x_\mathrm{s} - x_0\|_2$ is the distance between x_s and the reference position, in our case the origin of the coordinate system.

Validation of each measured source position commenced with the spatial discretization of a neighbourhood around the assumed source position. A 201×201 equidistantly spaced focus-grid with a resolution of $\Delta x = 0.5$ mm

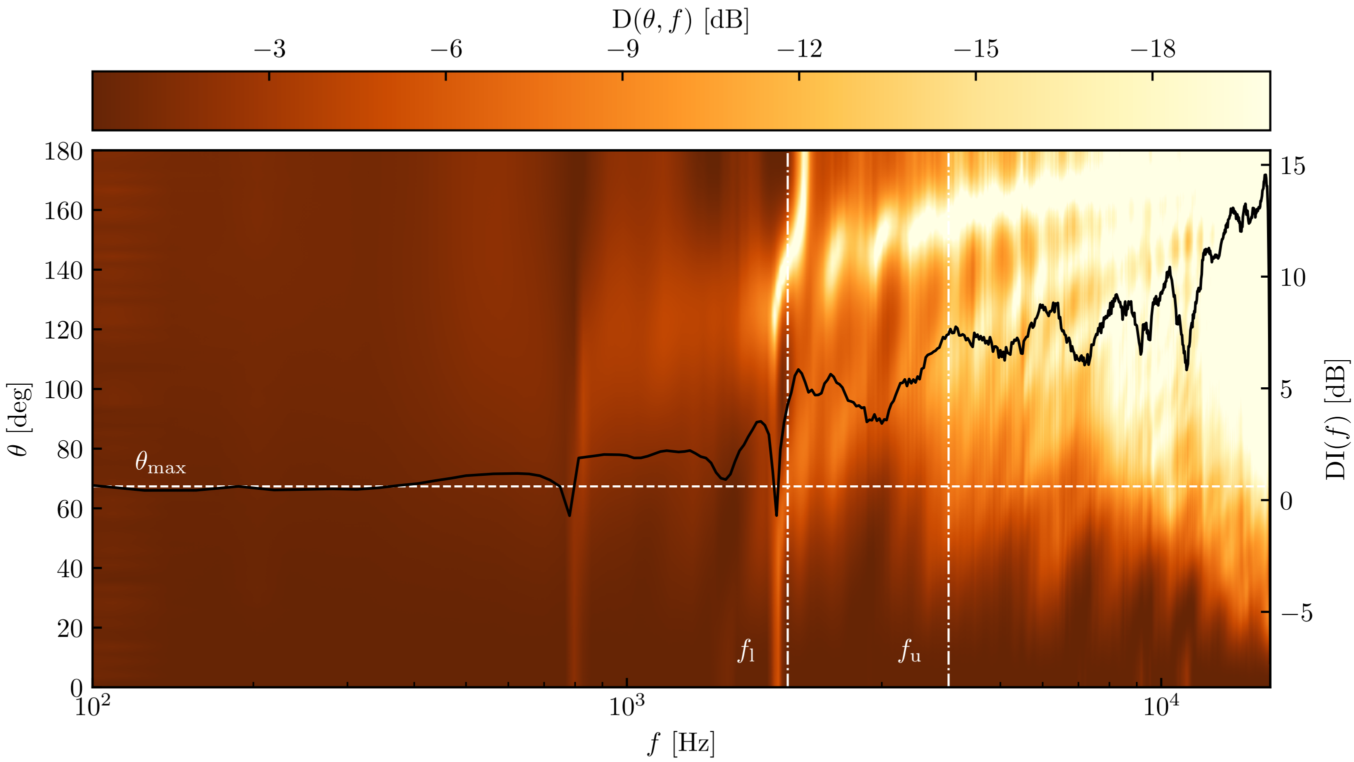

Figure 5: Directivity D and directivity index DI of the loudspeaker. The maximum opening angle across all experiments is denoted by θ_{max}.

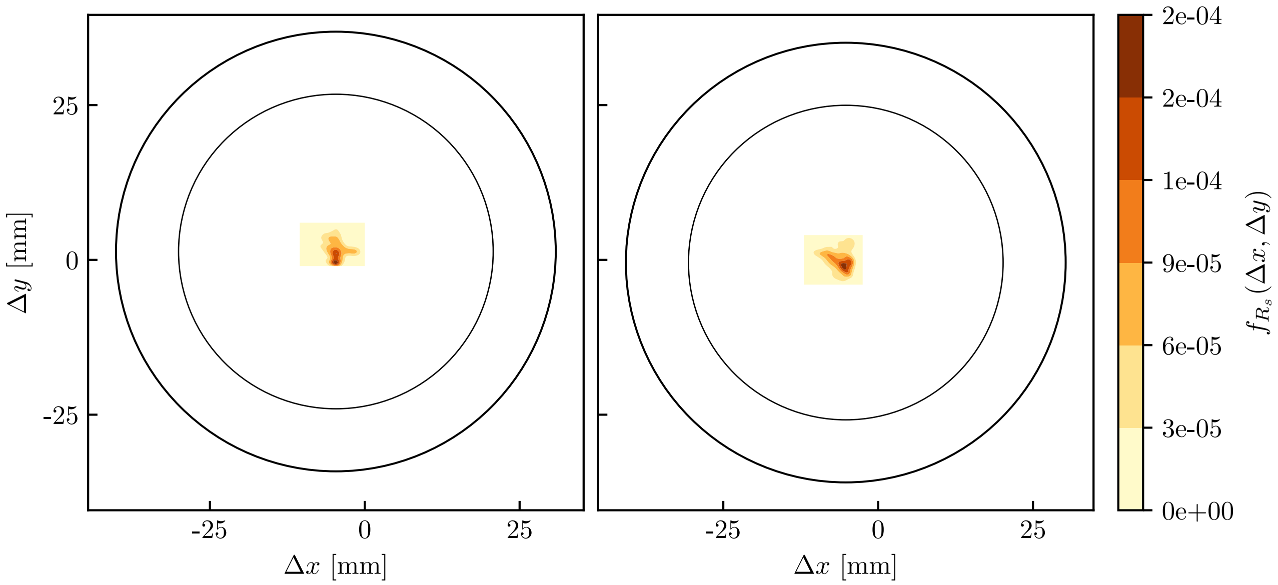

Figure 6: Estimated joint PDF of the positional deviations between the beamforming results and the assumed source positions. The inner black circle corresponds to the outer rim of the loudspeaker and the outer black circle indicates the outer rim of the enclosure box (left: Experiments {A1, D1}, right: Experiment A2)

was employed. The beamforming map was computed on the discretized region for every frequency in the range

$$\Omega = \left[2\pi \frac{f_\text{l}}{f_\text{d}}, 2\pi \frac{f_\text{u}}{f_\text{d}}\right]$$

which was chosen such that the lower frequency limit $f_\text{l} = 2\,\text{kHz}$ enabled a sufficiently large spatial resolution in the resulting beamforming map and the upper frequency limit $f_\text{u} = 4\,\text{kHz}$ ensures that the wavelength is larger than the loudspeaker diameter. The latter is important to ensure that the loudspeaker has a radiation pattern close to a monopole at relevant radiation angles in order to meet the monopole assumption needed for the steering vector formulation. As indicated by the dashed line in Figure 5, the radiation angle from the loudspeaker to any microphone in the array is bounded by $\theta_\text{max} = 67.3°$. The global spatial maximum is then determined by

$$\hat{x}_i = \underset{x_s}{\arg\max} \sum_{\omega \in \Omega} \frac{b(x_s, \omega)}{b(\hat{x}_s, \omega)},$$

whereby $b(\hat{x}_s, \omega)$ denotes the maximum value among all beamformed positions. To account for a potential mismatch of the source plane distance, the evaluation was conducted for different distances within a range of up to $\pm 120\,\text{mm}$ around the assumed source distance with a sampling interval of $\Delta z = 1\,\text{mm}$. Finally, the positional offset between the beamformer's prediction and the assumed source position is determined by $\Delta x_i = \hat{x}_i - x_i$.

Statistical Evaluation

The systematic positional offset between the center of the observation area and the microphone array in horizontal and vertical direction can be statistically determined by using the estimates $\Delta x_i \in \mathbb{R}^2$ for each individual measured source position. Thereby, each estimated positional deviation Δx_i can be seen as a realization of the jointly distributed random variables R_x, R_y with the joint probability density function (PDF) $f_{R_x,R_y}(\Delta x_i)$. It is assumed that the individual positional offset estimations Δx_i are symmetrically distributed around the true positional offset, due to the approximate symmetry of the microphone array and observation plane around the origin. Then, the true positional offset corresponds to the deviation associated with the greatest probability. A simple method to determine the joint PDF of jointly distributed random variables based on a finite set of samples is the kernel density estimation (KDE) [6], denoted by

$$\hat{f}_{R_x,R_y}(\Delta x_i) = \frac{1}{N} \sum_{n=1}^{N} K_h(\Delta x_i - \Delta x_i^{(n)}),$$

where N refers to the sample size and K_h is the so-called kernel. A bivariate Gaussian kernel with bandwidth h was used, whereby h was chosen according to the Silverman's rule of thumb [11].

Offset Correction

The first step of the correction procedure was to determine the true distance Δz between the loudspeaker and the microphone array plane for the experiments $\{A1, D1\}$ and $\{A2\}$. The estimation of the joint PDF was performed individually for each evaluated distance Δz. Note that source cases from experiment R2 were excluded from the statistical evaluation, since the ground plate reflections would introduce an additional disruptive factor in the positional estimation. It is assumed that the true distance minimizes the variance among any direction associated with $\hat{f}_{R_x,R_y}(\Delta x_i)$, i.e. the spectral norm of the covariance matrix $\Sigma_{\Delta x_i}(\Delta z)$ is minimized, such that

$$\underset{\Delta z}{\arg\min} ||\Sigma_{\Delta x_i}(\Delta z)||_2.$$

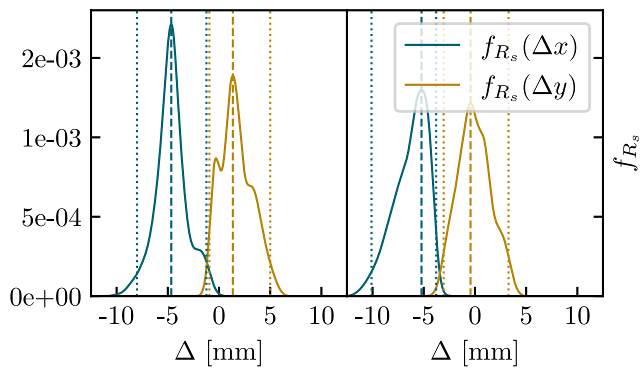

Figure 7: Marginal distribution functions characterizing the positional offset between the microphone array and the observation plane (left: Experiments {A1, D1}, right: Experiment A2). The dashed line indicates the positional offset corresponding to the maximum of the corresponding PDF. The dotted lines indicate the 2.5% and 97.5% percentiles.

Figure 6 shows the joint PDF with the smallest spectral norm for the experiments $\{A1, D1\}$ and $\{A2\}$. Based on the joint PDF corresponding to the optimal distance correction Δz, the true positional offset in vertical and horizontal direction is determined from the maximum of the corresponding marginal distributions depicted in Figure 7. Table 3 shows the positional offset correction values for each of the experiments.

Table 3: Positional correction values for each experiment.

Scenarios	Δx[mm]	Δy[mm]	Δz[mm]
A1, D1	$-4.7\,\text{mm}$	$1.4\,\text{mm}$	$5.0\,\text{mm}$
A2, R2	$-5.2\,\text{mm}$	$-0.4\,\text{mm}$	$7.0\,\text{mm}$

Acknowledgments

The authors also thank Arya Prasetya, Serdar Gareayaghi, Can Kurt Kayser and Roman Tschakert for their help with the experimental measurements and Fabian Brinkmann for valuable insights into sweep synthesis and experiment design.

The authors thankfully acknowledge the support of this research by Deutsche Forschungsgemeinschaft through projects 439144410 and 504367810.

References

[1] E. J. Arcondoulis et al. *Experimental validation and performance analysis of deep learning acoustic source imaging methods. 28th AIAA/CEAS Aeroacoustics Conference.* Southampton, UK, 2022, AIAA 2022-2852. DOI: 10.2514/6.2022-2852.

[2] A. Farina. *Simultaneous Measurement of Impulse Response and Distortion with Swept-sine Technique. 108th AES Convention, Preprint 5093.* Paris, 2000.

[3] A. Francl and J. McDermott. *Deep neural network models of sound localization reveal how perception is adapted to real-world environments.* Nature Human Behaviour 6 (2022), pp. 111–133. DOI: 10.1101/2020.07.21.214486.

[4] P. C. Hansen. *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion.* SIAM Monographs on Mathematical Modeling and Computation. Philadelphia: Siam, 1998, p. 247. ISBN: 978-0-89871-403-6.

[5] R. Merino-Martínez et al. *A review of acoustic imaging methods using phased microphone arrays.* CEAS Aeronautical Journal 10 (2019), pp. 197–230. ISSN: 18695590. DOI: 10.1007/s13272-019-00383-4.

[6] E. Parzen. *On Estimation of a Probability Density Function and Mode.* The Annals of Mathematical Statistics 33.3 (1962), pp. 1065–1076. ISSN: 00034851.

[7] K. Prawda, S. J. Schlecht, and V. Välimäki. *Robust selection of clean swept-sine measurements in non-stationary noise.* The Journal of the Acoustical Society of America 151.3 (2022), pp. 2117–2126. DOI: 10.1121/10.0009915.

[8] M. Rébillat et al. *Identification of Cascade of Hammerstein Models for the Description of Nonlinearities in Vibrating Devices.* Journal of Sound and Vibration 5 (2011), pp. 1018–1038. DOI: 10.1016/j.jsv.2010.09.012.

[9] E. Sarradj. *Three-Dimensional Acoustic Source Mapping with Different Beamforming Steering Vector Formulations.* Advances in Acoustics and Vibration, 292695 (2012). DOI: 10.1155/2012/292695.

[10] E. Sarradj. *A Generic Approach To Synthesize Optimal Array Microphone Arrangements. 6th Berlin Beamforming Conference.* Berlin, Germany: Gesellschaft zur Förderung angewandter Informatik (GFaI), 2016, pp. 1–12. ISBN: 978-3-942709-15-6.

[11] B. W. Silverman. *Density estimation for statistics and data analysis.* Chapman & Hall/CRC monographs on statistics and applied probability. London: Chapman and Hall, 1986. URL: https://cds.cern.ch/record/1070306.

A. Experiment Equipment

Table 4: Utilized hardware devices.

Device	Manufacturer	Type	Usage
Microphones	GRAS	40PL-1 Short CCP	Sound pressure acquisition
Temperature Sensor	OMNI SENSORS	OT60-B ($\pm 0.8\,°C$)	Temperature acquisition
Acquisition System	SINUS	Typhoon	Data acquisition
Stepper Motor	Steperonline	NEMA23	Axes positioning
Motor Control Unit	OpenBuilds	Blackbox X32	Control loudspeaker position
Amplifier	Klein & Hummel	Monoblock MB 80	Loudspeaker amplification
Turntable	Outline	ET2	Directivity measurement
Laser distance meter	PeakTech	2800A	Positional alignment
Cross line laser	Bosch	PCL20	Positional alignment

B. File Structure

The files `A1.h5`, `A2.h5` and `R2.h5` have a size of about 1.07 GB and `D1.h5` has a size of about 302.3 MB. Their contents are organized as follows:

<Dataset>

- `data`
  - `impulse_response` — `float32` array of shape $(n_\mathrm{i}, n_\mathrm{o}, n_t)$ - measured impulse responses
  - `location`
    - `receiver` — `float64` array of shape $(n_\mathrm{o}, 3)$ - microphone locations
    - `source` — `float64` array of shape $(n_\mathrm{o}, 3)$ - corrected source locations
    - `source_raw` — `float64` array of shape $(n_\mathrm{o}, 3)$ - uncorrected source locations
- `metadata`
  - `c0` — `float32` array of shape $(n_\mathrm{i},)$ - speed of sound
  - `temperature` — `float32` array of shape $(n_\mathrm{i},)$ - ambient temperature
  - `sampling_rate` — `int64` - sampling rate

We also supply the file `loudspeaker.h5` with a size of about 468 KB which contains the directivity measurements of the loudspeaker. Its contents are organzine as follows:

<Dataset>

- `data`
  - `angle` — `float32` array of shape $(73,)$ - measurement angles
  - `impulse_response` — `float32` array of shape $(73, n_t)$ - measured impulse responses
- `metadata`
  - `directivity` — `float32` array of shape $(73, 513)$ - directivity D
  - `directivity_index` — `float64` array of shape $(513,)$ - directivity index DI
  - `fftfreq` — `float64` array of shape $(513,)$ - corresponding frequencies
  - `sampling_rate` — `int64` - sampling rate