

EMPIRICAL RESEARCH

Open Access



MIRACLE—a microphone array impulse response dataset for acoustic learning

Adam Kujawski^{1*†} , Art J. R. Pelling^{1*†}  and Ennes Sarradj^{1*} 

Abstract

This work introduces a large dataset comprising impulse responses of spatially distributed sources within a plane parallel to a planar microphone array. The dataset, named MIRACLE, encompasses 856,128 single-channel impulse responses and includes four different measurement scenarios. Three measurement scenarios were conducted under anechoic conditions. The fourth scenario includes an additional specular reflection from a reflective panel. The source positions were obtained by uniformly discretizing a rectangular source plane parallel to the microphone for each scenario. The dataset contains three scenarios with a spatial resolution of 23 mm at two different source-plane-to-array distances, as well as a scenario with a resolution of 5 mm for the shorter distance. In contrast to existing room impulse response datasets, the accuracy of the provided source location labels is assessed and additional metadata, such as the directivity of the loudspeaker used for excitation, is provided. The MIRACLE dataset can be used as a benchmark for data-driven modelling and interpolation methods as well as for various acoustic machine learning tasks, such as source separation, localization, and characterization. Two timely applications of the dataset are presented in this work: the generation of microphone array data for data-driven source localization and characterization tasks and data-driven model order reduction.

Keywords Room impulse response, Dataset, Microphone array, Acoustics, Machine learning

1 Introduction

A room impulse response (RIR) characterizes the linear time-invariant acoustic propagation between a source and a receiver within a specific acoustic environment. RIRs are crucial for sound field auralization [1] as well as in the realm of room acoustics, where they are used for estimating acoustic properties of a room such as the reverberation time [2].

The emergence of data-driven methods in acoustics [3], particularly deep learning methods, has sparked increasing interest in the availability of rich, high-quality RIR datasets. These datasets play a pivotal role in the training of data-driven (interpolatory) sound field reconstruction methods [4–7], deep generative models [8, 9], and augmentation methods [10]. In addition, RIR datasets can be flexibly employed in order to synthesize acoustic training data for source localization and characterization [11], sound event detection, and speech separation tasks by convolving arbitrary source signals with RIRs [12–14]. The same synthesis procedure can be employed for data-driven acoustic parameter estimation problems, such as blind reverberation time estimation [15, 16] and others [17].

While data-driven methods often exhibit superior performance compared to conventional model-based methods, they require large amounts of realistic training data and are sensitive to variations of underlying probability

[†]Adam Kujawski and Art J. R. Pelling contributed equally to this work.

*Correspondence:

Adam Kujawski
adam.kujawski@tu-berlin.de

Art J. R. Pelling
a.pelling@tu-berlin.de

Ennes Sarradj
ennes.sarradj@tu-berlin.de

¹ Department of Engineering Acoustics, TU Berlin, Einsteinufer 25, Berlin 10587, Germany

distributions describing the data, also known as *dataset shift* [18]. Experimental data is oftentimes not available or too time-consuming to acquire. Many data-driven methods across various application areas, such as speech enhancement and recognition [19, 20], localization [11, 21], sound field reconstruction [22], room acoustic parameter estimation [23, 24], and acoustical engineering [25–27], are therefore trained with simulated data, whereby enhanced realism helps to improve generalization performance [28, 29]. However, without adaptation to or training with realistic data, the performance of data-driven methods can be significantly impaired [25, 30], which indicates the need for experimentally measured RIR datasets.

1.1 State of the art

In contrast to the diverse analytical and data-driven frameworks for the simulation of RIRs proposed in the past [8, 31–34], the availability of openly available experimental RIR datasets is limited. Moreover, the landscape of RIR datasets is rather heterogeneous because the experimental data is usually acquired with a certain application in mind. This manifests in application-specific environments, customized source-receiver arrangements, and non-standardized dataset annotations.

In some datasets, the focus lies on environmental variability through different room geometries and reverberation times [35–40]. The SoundCam dataset [35] poses an extreme example in which the environmental characteristics of four rooms are manipulated by placing humans at hundreds of locations in the rooms, leading to an enormous amount of environmental representations. These datasets are of particular interest in speech recognition, teleconferencing, and sound event detection applications. While in some of these datasets, the sensors are distributed over the room [35, 36], others rely on compact microphone arrangements with only a few channels and aperture sizes that are small compared to the source distance [39–41]. The latter is often the case in speech and sound event localization applications. Only very few datasets provide additional annotations of the environmental characteristics, such as echo annotations, which enable their use for acoustic parameter estimation problems [40].

Conversely, there are application fields where spatial diversity and accurate positional annotations are similarly or even more important than environmental variability. In acoustical engineering applications, microphone array measurements for determining source locations and individually induced sound pressure at a specific receiver position are widely used to implement noise reduction measures [42]. To accurately estimate these source characteristics, free-field conditions or laboratory

environments with a short reverberation time are preferred. Usually, the utilized microphone arrangements are planar and contain many channels spread over a large area to ensure accurate spatial filtering capabilities over a wide frequency range. Typical examples include the measurement of machinery noise and wind turbines. Oftentimes, multiple closely neighbouring sound sources can be expected. There exist datasets that employed mechatronic positioning devices, which offer a higher positional accuracy over manual positioning, to achieve a dense sampling of the acoustic space under reverberant [43–45] or anechoic conditions [46]. For example, the Pyramic dataset [46] was acquired by rotating a tetrahedral microphone array with a turntable device using fixed loudspeaker positions. However, the accuracy of the acquired positional information is rarely assessed in the literature [46]. A limitation of these datasets is that either the number of sources or receivers can be considered as small [39, 43, 44, 46], which prohibits their use in engineering acoustic applications where a dense spatial sampling of the source and receiver space is required.

Another discipline with similar data requirements is the spatial interpolation of room impulse responses [5–7, 47]. Due to the large spatial variability of RIRs [47], dense sampling grids have to be employed in order to adequately capture the entire range of dynamics, especially at high frequencies. The absence of large and spatially dense RIR datasets entails that thorough assessment of interpolation methods has only been performed with synthetic data so far [5, 7].

Regardless of the application, only a few large-scale datasets encompassing several thousand RIRs, the largest of which is presented in this work. These datasets are outlined in Table 1. In addition to these larger datasets, there are several popular datasets with a smaller number of RIRs, for example [38–40, 48–55]. Large and realistic RIR datasets are of great interest to many scientific communities, especially those concerned with the development of novel data-driven modelling algorithms. The mathematical field of model order reduction (MOR) [56], for example, is rapidly developing data-driven methods [57–59]. However, in the majority of cases, new methods are validated only with synthetic data of common benchmark problems [60–62] and the validation of these methods in realistic scenarios with high-dimensional data is still pending. Providing well-documented high-dimensional measurement data is an essential element that can enable these communities to increase the applicability of their methods to real problems.

In this work, we introduce a large measured room impulse response (RIR) dataset, which we call “Microphone Array Impulse Response Dataset for Acoustical Learning” (MIRACLE). The dataset is tailored to the

Table 1 Existing large-scale RIR datasets consisting of more than 2000 RIRs (A, anechoic; R, reflective; M, mechatronic positioning)

Dataset	No. RIRs	Sensors		Sources		Environment		M
		Geometry	No.	Area	No.	Type	No.	
MIRACLE	262,144	Vogel's Spiral	64	2D grid	4096	A	1	✓
	69,696				1089			
	262,144				4096			
	262,144				4096			
SoundCam [35]	18,000	Room enclosing	10	n/a	1	R	1800	✗
	10,000				1000			
	11,000				1100			
	11,000				1100			
ARNI [36]	132,037	n/a	5	n/a	1	R	5342	✗
MeshRIR [43]	3969	3D grid	3969	n/a	1	R	1	✓
	14,112	2D grid	441	2D grid	32			
Motus [37]	106,240	Eigenmike	32	n/a	4	R	830	✗
MIRaGe [44]	371,610	1D grid	6×5	3D grid	4104+25	R	3	✓
Pyramic [46]	25,920	Tetrahedron	48	Circular grid	3×180	A	1	✓
TUT Tietotalo Ambisonic IR [41]	10,368	Eigenmike	32	3D Grid	324	R	1	✗
	5760				180			

application field of sound source localization and characterization in acoustical engineering applications, and spatial interpolation of RIRs. The focus of the dataset lies on spatial rather than on environmental variability and accurate positional annotations.

Its key highlights are:

1. MIRACLE is the first RIR dataset with a dense sampling of the environment in both source and receiver space.
2. Particularly accurate spatial sampling is achieved by using a mechatronic positioning device to control the loudspeaker position. In contrast to most of the previously published datasets, the accuracy of the source positions (positional labels) is statistically validated. The assessment reveals outstandingly accurate and precise source locations with positional uncertainties in the order of a few millimetres.
3. With a total of 856,128 captured room impulse responses and dense spatial sampling of the observation area, the dataset can be well suited for machine learning and deep learning.
4. This paper presents two important applications of the dataset. As will be shown in Section 4.1, the dataset is well suited to construct acoustic source cases with possibly closely neighbouring sources by superimposing signals that have been convolved with the provided RIRs. In Section 4.2, reduced order state-space models are constructed from the MIRACLE dataset, demonstrating that it can serve as an excel-

lent real-world benchmark problem for data-driven reduced order modelling methods.

2 Materials and methods

2.1 Experimental setup

The experimental setup is illustrated in Fig. 1. Details on the utilized hardware are given in Table 2.

2.1.1 Microphone array

The phased microphone setup features a planar microphone array comprising $n_o = 64$ channels mounted in a $1.5 \text{ m} \times 1.5 \text{ m}$ aluminium plate. The microphone arrangement follows Vogel's spiral [63]. The maximum pairwise distance between the array microphones is referred to as the aperture size $d_a = 1.47 \text{ m}$. The microphone array data was acquired with a multichannel acquisition system (sampling rate = 51.2 kHz).

2.1.2 Sound source and excitation signal

A dynamic 2" cone loudspeaker in a cylindrical 3D-printed enclosure was employed as the sound source. An exponential sine sweep was used as the excitation signal because of its favourable properties with regard to crest factor and rejection of non-linearities [64]. It was designed according to [65–67] 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 low noise levels, it was possible to choose a relatively short sweep time of 3 s for the measurement. In order to ensure that the entire system response after excitation

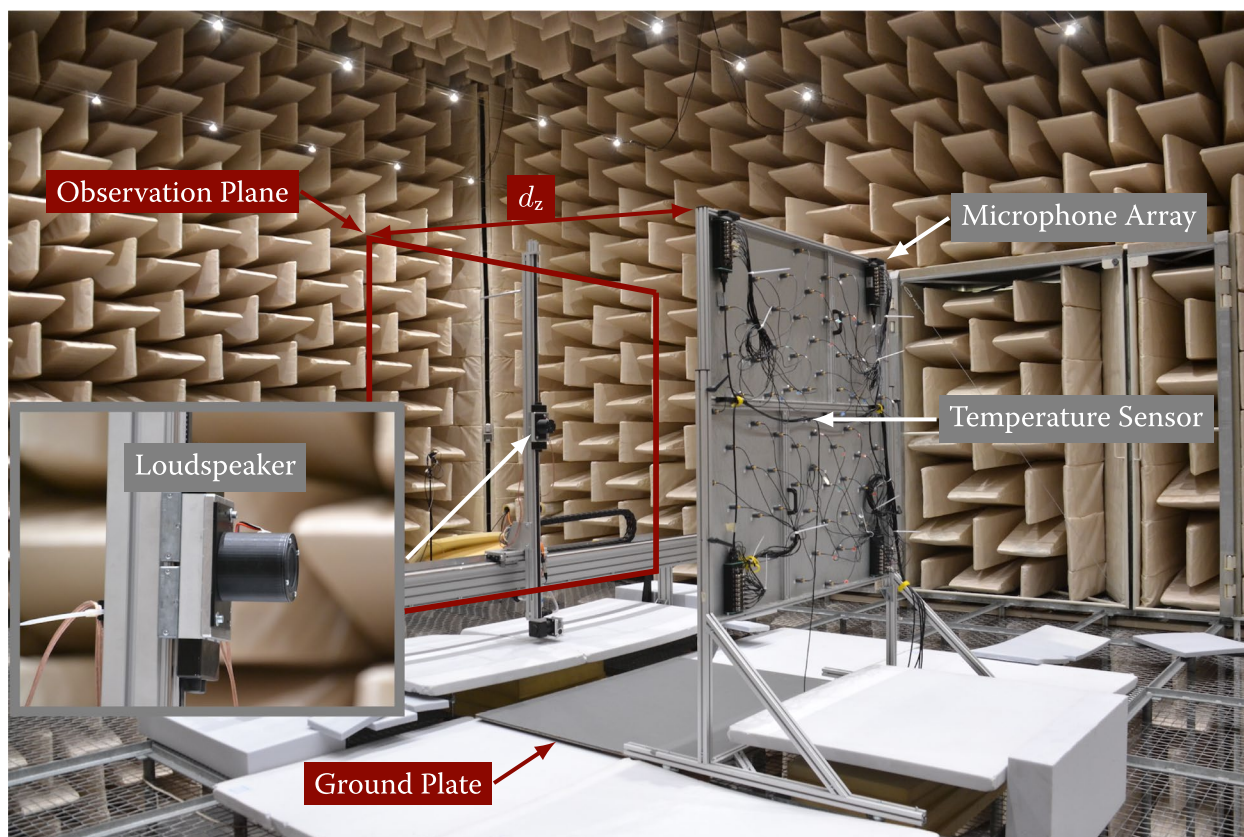


Fig. 1 Experimental setup for the main experiment (R2) with reflective ground plate

is captured, a safety window of 250 ms was added to the recording duration, resulting in $n_s = 166,400$ samples per measurement. The loudspeaker excitation signal was also fed back directly to the AD-converter and was synchronously recorded with the microphone signals as a reference signal for post-processing. It is referred to as the loopback excitation signal in the following.

2.1.3 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 using a laser distance meter and a cross-line laser, achieving only minor alignment errors of a few millimetres at worst. The loudspeaker dust cap at the membrane centre was used as reference in the manual alignment. During data post-processing, a spatial offset

Table 2 Utilized hardware devices

Device	Manufacturer	Type	Usage
Microphones	GRAS	40PL-1 Short CCP	Sound pressure acquisition
Temperature Sensor ^a	OMNI SENSORS	OT60-B (± 0.8 °C)	Temperature acquisition
Acquisition System	SINUS	Typhoon	Data acquisition
Stepper Motor	Stepperonline	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

^a Calibration was performed after the measurement campaign using a reference sensor with a temperature accuracy of ± 0.1 °C

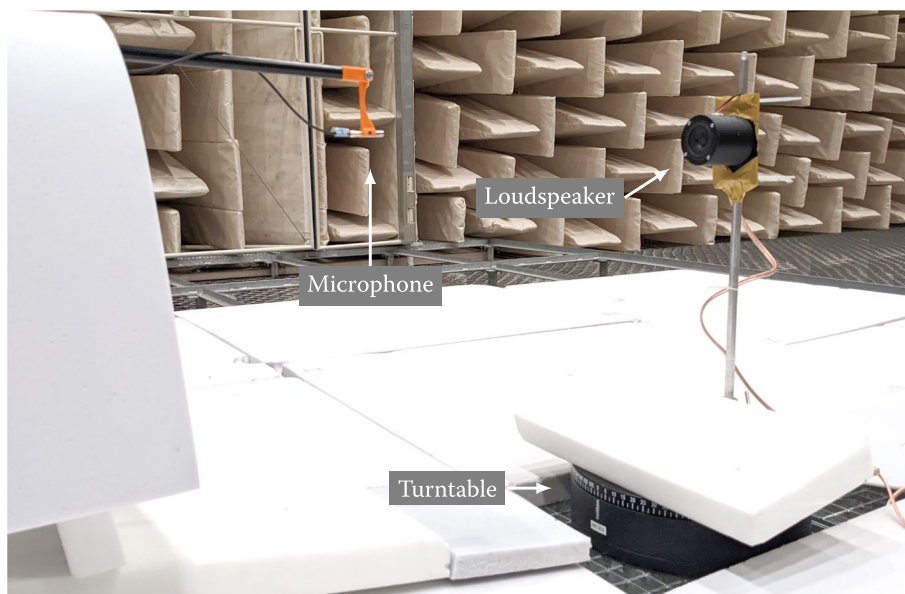


Fig. 2 Experimental setup for the directivity measurement

correction was applied based on a statistical evaluation given in Section 3.3. The corrected positions apply to the acoustical centre of the loudspeaker rather than the centre of the membrane.

2.1.4 Environment

All measurements were performed in the anechoic chamber of TU Berlin (room volume $V = 830 \text{ m}^3$, lower cut-off frequency $f_c = 63 \text{ Hz}$). Neither heating nor air conditioning was active, and the temperature was monitored at the microphone array centre throughout the experiment. A ground plate was placed between the loudspeaker and the microphone array in one of the experimental scenarios to enable a reflective environment. The supporting grid platform and the positioning system were clad with absorptive foam to minimize reflections.

2.2 Experimental procedure

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 min (the duration was determined in a dedicated experiment). This warm-up phase accounts for the weakly non-stationary dynamics of the loudspeaker's transfer function, e.g., changes of the properties of the loudspeaker magnet related to internal temperature fluctuations; see [68]. 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

(1 s each) and loudspeaker excitation measurements (3 s each) were performed using all n_o microphones at once. Subsequently, the cross-correlation between all n_i recorded channels was evaluated according to the *rule of two* [69]. 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.

Following the main measurement campaign, an additional measurement was conducted in the anechoic chamber to obtain the angle-dependent frequency response of the loudspeaker at discrete azimuth angles at a resolution of $\Delta\theta = 2.5^\circ$. A microphone was placed at a distance of 0.5 m from the loudspeaker centre. The latter was mounted on a motor-driven dispersion measurement turntable. A photograph of the measurement setup can be found in Fig. 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.

2.3 Post-processing

Several post-processing steps were performed to obtain a good estimate of the system impulse response from the measurements. Firstly, the loopback excitation and microphone signals were averaged across the two measurement repetitions to obtain a single averaged excitation signal $\tilde{u}_{i,j} \in \mathbb{R}^{n_s}$ and averaged 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 resampled to a sampling rate of $f_d = 32$ kHz since the loudspeaker transmission capability and excitation sweep have an upper frequency limit of 16 kHz. We applied the polyphase method for resampling (see [70] for details).

2.3.1 Deconvolution

In the following, let $n_d = 104,000$ denote the number of samples after resampling. An estimate of the frequency response was obtained by dividing the Discrete Fourier Transform (DFT) of the averaged and downsampled measurement signals $Y_{i,j} = \text{DFT}(\tilde{y}_{i,j}) \in \mathbb{C}^{n_d}$ by the corresponding DFT of the averaged and resampled loopback excitation signals $U_{i,j} = \text{DFT}(\tilde{u}_{i,j}) \in \mathbb{C}^{n_d}$, i.e.

$$H_{i,j}(e^{t\omega_k}) = Y_{i,j}(e^{t\omega_k})U_{i,j}^{-1}(e^{t\omega_k}) \in \mathbb{C}, \quad (1)$$

for the angular frequency $\omega_k = 2\pi k/n_d$ with $k \in [-n_d/2, n_d/2] \subset \mathbb{Z}$. The inverse spectra $U_{i,j}^{-1} \in \mathbb{C}^{n_d}$ were obtained by regularized inversion [67, 71–75]

$$U_{i,j}^{-1}(e^{t\omega_k}) = \frac{U_{i,j}^*(e^{t\omega_k})}{U_{i,j}^*(e^{t\omega_k})U_{i,j}(e^{t\omega_k}) + M\lambda(e^{t\omega_k})}, \quad (2)$$

where $M = \max_{k \in \{1, \dots, n_d\}} \{|U_{i,j}(e^{t\omega_k})|^2\} = 1$. Regularization is necessary to avoid instabilities in the deconvolved frequency response that arise from persistently exciting only over a limited frequency range. Practical considerations for choosing the regularization term in acoustic applications can be found in [67]. The regularization term $\lambda \in \mathbb{R}^{n_d}$ was chosen as

$$\lambda(e^{t\omega_k}) = \begin{cases} 1 & \text{for } |\omega_k| \in [0, \omega_{\text{fade}}] \\ \frac{1 + \cos\left(\frac{\omega_{\text{fade}} - |\omega_k|}{\omega_{\text{fade}} - \omega_{\text{cut}}}\right)}{2} & \text{for } |\omega_k| \in [\omega_{\text{fade}}, \omega_{\text{cut}}] \\ 0 & \text{for } |\omega_k| \in [\omega_{\text{cut}}, \pi] \end{cases} \quad (3)$$

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

$$\omega_{\text{cut}} = 2\pi \frac{100 \text{ Hz}}{f_d} \quad (4)$$

which is chosen according to the lower limit of the loudspeaker's frequency range of 100 Hz and equal to 1 below

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

$$h_{i,j} = \text{DFT}^{-1}(H_{i,j}). \quad (5)$$

2.3.2 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 $e \in \mathbb{R}^{n_d}$ given by

$$e(t) = \min_{i \in \{1, \dots, n_d\}} \sum_{\tau=1}^t |h_{i,j}(\tau)|^2, \quad t \in \{1, \dots, n_d\}, \quad (6)$$

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

$$n_t = 1,024 \geq \tilde{t} = \arg \max_{t \in \{1, \dots, n_d\}} \{e(t) \leq 0.999 \|e\|_{\infty}\}. \quad (7)$$

3 Results and discussion

3.1 Impulse responses

A total of four different experimental scenarios were realized, which are summarized in Table 3. The acquisition time for each of the large-scale scenarios A1, A2, and R2 was about 20 h. The total number of single-channel impulse responses across all scenarios is 856,128. The scenarios differ regarding the environment as well as the spatial dimension ($d_y = d_x$), sampling resolution ($\Delta d_y = \Delta d_x$), and distance d_z of the source plane. The two large anechoic scenarios A1 and A2 each include 4096 measured source positions on an equidistantly spaced 64×64 grid at different source-plane distances d_z . In addition, a densely sampled scenario D1 was acquired on a smaller 33×33 grid with a spacing of only 5 mm. Scenario R2 is based on the same geometric setup as scenario A2, but an aluminium plate on the floor introduces a specular reflection.

Table 3 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

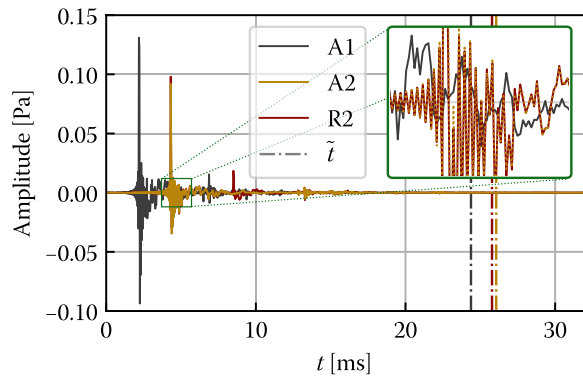


Fig. 3 Measured impulse responses for the scenarios A1, A2, and R2 and the centremost locations in the source and receiver plane. The dash-dotted vertical lines indicate the scenarios' index \tilde{t}

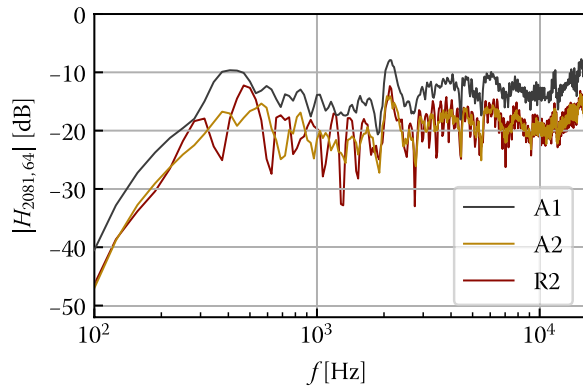


Fig. 4 Magnitude of the frequency response of the measured transfer functions for the scenarios A1, A2, and R2 at the centremost locations in the source and receiver plane

Figures 3 and 4 exemplarily show the measured impulse response and its magnitude spectrum for a single source-receiver combination for scenarios A1, A2, and R2, respectively. It can be readily verified that the doubling of the distance to the source is also reflected in a doubling of the delay shift and an attenuation of the magnitude spectrum by approximately -6 dB. Furthermore, the specular reflection for scenario R2 manifests in a prominent second peak in the impulse response and comb filtering in its magnitude spectrum. Additional reflections manifesting as spurious peaks in the impulse response are due to the structure of the positioning system and the supporting grid platform.

The mean and standard deviation of temperature and the speed of sound for each of the scenarios are given in Table 4. The speed of sound has been calculated

Table 4 Mean μ and standard deviation σ of the temperature and speed of sound for each experiment

Scenario	Temperature [$^{\circ}\text{C}$]		Speed of sound [ms^{-1}]	
A1	$\mu = 21.6$	$\sigma = 0.12$	$\mu = 344.8$	$\sigma = 0.07$
D1	$\mu = 21.8$	$\sigma = 0.01$	$\mu = 345.0$	$\sigma = 0.01$
A2	$\mu = 22.3$	$\sigma = 0.05$	$\mu = 345.3$	$\sigma = 0.03$
R2	$\mu = 22.5$	$\sigma = 0.02$	$\mu = 345.4$	$\sigma = 0.01$

according to [76, 77]¹. It reveals that the temperature and the speed of sound are almost identical across all scenarios with an absolute difference of $\Delta\mu < 1^{\circ}\text{C}$ and $\Delta\mu \leq 0.6 \frac{\text{m}}{\text{s}}$, respectively, which is expected due to the fairly constant environmental conditions inside the anechoic chamber.

3.2 Loudspeaker directivity

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 $p_{\text{RMS}}(\theta, f)$ at an angle θ and the maximum among all angles, i.e.

$$D(\theta, f) = 10 \log_{10} \left(\frac{p_{\text{RMS}}(\theta, f)}{\max_{\phi \in [0, 2\pi]} p_{\text{RMS}}(\phi, f)} \right), \quad (8)$$

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

$$DI(f) = 10 \log_{10} \left(\frac{4\pi p_{\text{RMS}}^2(0, f)}{2\pi \int_0^\pi p_{\text{RMS}}^2(\phi, f) \sin(\phi) d\phi} \right), \quad (9)$$

where $p_{\text{RMS}}^2(0, f)$ represents the squared sound pressure in front of the speaker.

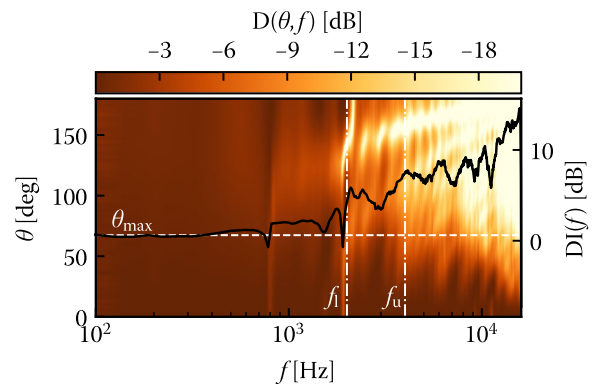


Fig. 5 Directivity D and directivity index DI of the loudspeaker. The maximum opening angle across all experiments is denoted by θ_{max}

¹ An atmospheric pressure of 101.325 kPa and a carbon dioxide mole fraction of 0.0004 was used. A generic value of 38% was used for the relative humidity approximating the humidity conditions throughout the experiments

It is seen that the loudspeaker exhibits a radiation pattern similar to a monopole until an upper frequency of 2 kHz. Above this frequency, the directivity index increases. Still, the directivity observed by the microphone array is close to a monopole at relevant radiation angles, i.e. $\theta \leq \theta_{\max} = 67.3^\circ$, as indicated by the dashed line in Fig. 5.

3.3 Positional validation

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

Due to the anechoic environment and the use of a large-scale microphone array enabling an excellent spatial resolution, Conventional Frequency Domain Beamforming [42] serves as an appropriate method to obtain an estimate of the actual source location. The large number of acoustic cases also permits a statistical approach to determine the spatial offset for a measurement scenario and to quantify the uncertainty regarding the source position information.

3.3.1 Beamforming

Let $\omega_k = 2\pi k/n_d$ with $k \in [-n_d/2, n_d/2] \subset \mathbb{Z}$ and let

$$H(e^{i\omega_k}) = [H_{i,1}(e^{i\omega_k}) \quad \dots \quad H_{i,n_o}(e^{i\omega_k})] \in \mathbb{C}^{n_o} \quad (10)$$

denote the transfer function measurements from the i -th source at location x_s for $i \in \{1, \dots, n_i\}$ to each of the n_o microphones. The cross-spectral matrix induced by a sound source with unit strength is then given by

$$C(\omega_k) = H(e^{i\omega_k})H(e^{i\omega_k})^* \in \mathbb{C}^{n_o \times n_o}. \quad (11)$$

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

$$b(x_s, \omega_k) = \|a(x_s, \omega_k)\|_{C(\omega_k)}^2 = a(x_s, \omega_k)^* C(\omega_k) a(x_s, \omega_k). \quad (12)$$

Many formulations of the steering vector can be found in the literature. The formulations I and IV in [78] 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

$$\{a(x_s, \omega)\}_j = \frac{e^{i\omega(r_j - r_0)/c}}{r_j \sqrt{n_o \sum_{k=1}^{n_o} r_k^{-2}}}, \quad (13)$$

where $r_j = \|x_s - x_j\|_2$ is the distance between the assumed source location x_s and the j -th microphone location x_j , and $r_0 = \|x_s - x_0\|_2$ is the distance between x_s and the reference position, in this 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 was employed. The beamforming map was computed on the discretized region for every frequency in the range

$$\Omega = \left[2\pi \frac{f_l}{f_d}, 2\pi \frac{f_u}{f_d} \right] \quad (14)$$

which was chosen such that the lower frequency limit $f_l = 2$ kHz enabled a sufficiently large spatial resolution in the resulting beamforming map, and the upper frequency limit $f_u = 4$ 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 Fig. 5, the radiation angle from the loudspeaker to any microphone in the array is bounded by $\theta_{\max} = 67.3^\circ$. The global spatial maximum is then determined by

$$\hat{x}_i = \arg \max_{x_s} \sum_{\omega \in \Omega} \hat{b}(x_s, \omega), \quad (15)$$

where $\hat{b}(x_s, \omega)$ denotes the amplitude normalized beamforming result

$$\hat{b}(x_s, \omega) = \frac{b(x_s, \omega)}{b(\hat{x}_s, \omega)}, \quad (16)$$

with $b(\hat{x}_s, \omega)$ being the beamformer's maximum output among all source locations x_s at a given frequency ω . The evaluation was conducted for different distances within a range of up to ± 12 mm around the assumed source distance with a sampling interval of $\Delta z = 1$ mm to account for a potential mismatch of the source plane distance. Finally, the positional offset between the beamformer's prediction and the assumed source position is determined by $\Delta x_i = \hat{x}_i - x_i$.

3.3.2 Statistical evaluation

The systematic positional offset between the centre of the observation area and the microphone array in the 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 [79], denoted by

$$\hat{f}_{R_x, R_y}(\Delta x_i) = \frac{1}{N} \sum_{n=1}^N K_h(\Delta x_i - \Delta x_i^{(n)}), \quad (17)$$

where N refers to the sample size and K_h is the so-called kernel. A bivariate Gaussian kernel with bandwidth h was used, where h was chosen according to the *Silverman's rule of thumb* [80].

3.3.3 Offset correction

The correction procedure's first step was determining the distance Δz between the loudspeaker and the microphone array plane for the experiments {A1, D1} and {A2}. The joint PDF was estimated 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

$$\arg \min_{\Delta z} \|\Sigma_{\Delta x_i}(\Delta z)\|_2. \quad (18)$$

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

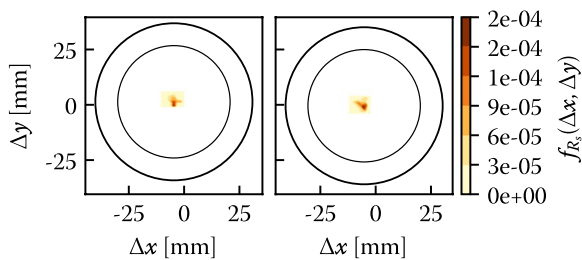


Fig. 6 Estimated joint PDF of the positional deviations between the beamforming results and the assumed source positions. The inner black circle corresponds to the outer rim of the loudspeaker and the outer black circle indicates the outer rim of the enclosure box (left: Experiments {A1, D1}, right: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

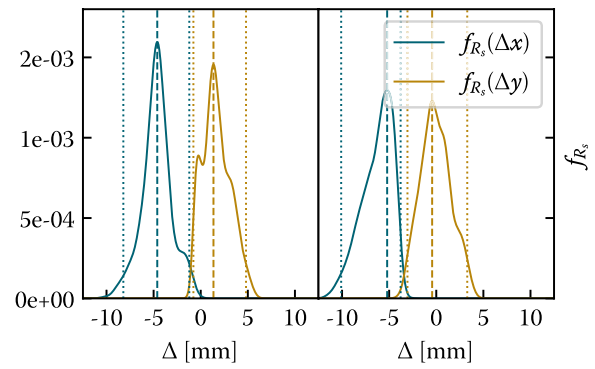


Fig. 7 Marginal distribution functions characterizing the positional offset between the microphone array and the observation plane (left: Experiments {A1, D1}, right: Experiment A2). The dashed line indicates the positional offset corresponding to the maximum of the corresponding PDF. The dotted lines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

Table 5 Positional correction values for each experiment

Scenarios	Δx [mm]	Δy [mm]	Δz [mm]
A1, D1	−4.6 mm	1.4 mm	4.0 mm
A2, R2	−5.2 mm	−0.4 mm	6.0 mm

correction Δz , the true positional offset in vertical and horizontal direction is determined from the maximum of the corresponding marginal distributions depicted in Fig. 7. Table 5 shows the positional offset correction values for each of the experiments.

With the correction offset applied, one can conclude that the positional uncertainties regarding the true source positions are in the order of a few millimetres. Given the 2.5 and 97.5 percentiles of the marginal distributions, the positional uncertainty is in the range of $[-3.6 \text{ mm}, 3.4 \text{ mm}]$ in x-direction and $[-2.1 \text{ mm}, 3.5 \text{ mm}]$ in y-direction for the experiments {A1, D1}. Regarding the experiments {A2, R2}, the positional uncertainty is in the range of $[-4.9 \text{ mm}, 1.4 \text{ mm}]$ in x-direction and $[-2.6 \text{ mm}, 3.7 \text{ mm}]$ in y-direction.

4 Applications

Experimental measurement of acoustical systems is rarely an end in itself. Usually, one utilizes the measurements to infer system properties or data-driven models of the system at hand. Given its substantial size and precision, the MIRACLE dataset can be applied to a variety of research applications. In the following, two timely and crucial acoustic research areas are addressed: firstly, source localization and characterization, and secondly, data-driven reduced order modelling.

4.1 Acoustic source localization and characterization

In recent years, there has been extensive investigation into data-driven methods, particularly deep learning models, to solve source localization and characterization problems [11]. Localization focuses on determining the positions of one or multiple sound sources, whereas source characterization entails quantifying additional characteristics, such as their individual strength at a specific spatial reference position. Due to the scarcity of suitable experimental data, it is common to leverage synthetically generated microphone array data, with only few studies considering real-world data already during training [25]. It remains an open research question how strongly the performance of data-driven methods is affected by this oversimplification and how this performance degradation could be mitigated.

4.1.1 Microphone array data generation

One way to overcome the lack of experimental data is to generate microphone array data via a convolution of experimental RIRs with arbitrary source signals in a Monte Carlo simulation. In order to facilitate reproducibility in machine learning, the training data has to be published alongside the source code. Providing the immutable raw microphone array data of the Monte Carlo simulation would lead to infeasibly large storage requirements. As a lightweight alternative, a code environment of the data simulation process can be published. This enables the generation of microphone array data on demand, which not only alleviates the problem of storage but can also greatly accelerate the training process. This approach is taken by the *AcouPipe framework* [81], a previous work of the authors.

AcouPipe is an open-source project written in Python programming language that incorporates a default dataset generation method that constructs source cases with an arbitrary number of sound sources. As of AcouPipe v24.04 [82], the AcouPipe framework includes an additional experimental dataset generation method that utilizes the MIRACLE dataset per default. Both the synthetic and the experimental data generation method rely on similar measurement setups. A Monte Carlo simulation is used to randomly generate multi-source scenarios by superposition. By default, the number of sources and their respective positions are drawn from a Poisson distribution $I \sim \mathcal{P}(\lambda = 3)$ and a bivariate normal distribution $(x_i, y_i) \sim \mathcal{N}(\mu = 0, \sigma = 0.1688 d_a)$ for each individual source case. Artificial white noise is used as the source signal with a signal length of five s. The signal amplitudes are chosen according to the desired squared RMS sound pressure at the reference position. The latter is drawn from a Rayleigh distribution, such that $p_{\text{RMS},i}^2 \sim \mathcal{R}(\sigma_R = 5)$. Finally, the sound pressure signals at the microphones are

obtained by superposition and convolution of the respective RIRs with the source signals

$$p_j(t) = \sum_{i=1}^I h_{i,j}(t) * q_i(t) + n_j(t). \quad (19)$$

Uncorrelated white noise $n_j(t)$ is added to each microphone channel. A more detailed description of the underlying dataset generation procedure can be found in [81].

```

from acoupipe.datasets import experimental

dataset = experimental.DatasetMIRACLE(
    scenario='A1')

# setup data generator
data_generator = dataset.generate(
    features=['sourcemap', 'loc'],
    f=2500,
    size=1000,
    split='validation')

# Iterate over / generate the data
source_case_data = next(data_generator)

```

Listing 1 Python source code snippet demonstrating multi-source scenario generation based on the MIRACLE dataset. To demonstrate the ease of use, Listing 1 exhibits a Python source code example that generates multi-source scenarios based on the MIRACLE dataset. Here, a `Dataset` instance is constructed, utilizing the RIRs from scenario A1 for the generation of 1000 validation samples. With AcouPipe, it is not only possible to obtain the raw time data but also to extract several features. The features include a source mapping of the sources and their locations. By default, conventional beamforming is used to calculate the source mapping. Figure 8 shows the beamforming map related to the seventh sample of the validation dataset.

Theoretically, very large numbers of experimental scenarios can be created by superimposing the individual measurements of the MIRACLE dataset. For example, when considering unordered sampling without replacement

$$\binom{n_i}{I} = \frac{n_i!}{I!(n_i - I)!}, \quad (20)$$

more than eight million different source cases can be constructed when considering scenario A1 and two sources. The number of possible combinations increases significantly when sources of different strengths or even more than two sources are combined.

We believe that both the synthetic and the experimental data provide an excellent basis to address and answer

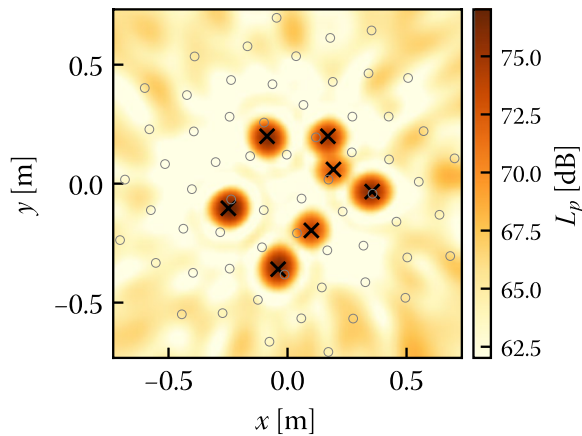


Fig. 8 Source mapping at $f = 2.5$ kHz for the seventh sample of the experimental validation dataset. Grey circles indicate the microphone position and black crosses indicate the source positions

up-to-date research questions related to data-driven source localization and characterization tasks, such as: How strong are synthetically trained source characterization models influenced by the domain shift when evaluated with data acquired in a real environment? Which domain generalization and adaptation methods are suitable to overcome the performance degradation due to the domain shift? How do existing source localization and characterization methods generalize to real-world scenarios?

4.2 Reduced order modelling

System measurements are usually performed with a certain goal in mind, e.g. inferring system properties from the data, such as the reverberation time. In many cases, one would like to obtain a model for the dynamical system transmission, i.e. the mapping from an input signal to an output signal. A common and apparent approach to obtaining a model for the system dynamics from RIR measurements is to employ them as a convolution kernel $h \in \ell_2^{n_o \times n_i}$ via the convolution sum

$$y(t) = \sum_{k \in \mathbb{N}} h(k-t)u(t), \quad (21)$$

where $u \in \ell_2^{n_i}$ and $y \in \ell_2^{n_o}$ are the input and output signal, respectively, and ℓ_2 denotes the Hilbert space of square-summable sequences. This approach was also pursued in the previous section. However, it can be unfavourable with regard to computational effort and storage requirements, especially for larger systems with many inputs and/or outputs, such as densely sampled RIRs, because the computational effort scales with the product of the

number of inputs and outputs [83]. Furthermore, systems whose dynamics are governed by a low number of weakly damped modes, e.g. small room acoustics, possess long impulse responses but are otherwise described by a simple model. In a convolution-based approach, there are no effective means to eliminate redundancies in the data, making the dynamical transmission of aforementioned systems unnecessarily expensive to compute.

As argued in [83, 84], state-space models can be computationally advantageous, especially in real-time scenarios, and provide access to powerful system-theoretic MOR methods [56]. MOR is a very active research field that aims to find small models that approximate the dynamics of a large and potentially infeasible so-called *full order model* as closely as possible. Instead of firstly identifying a large full order model from data and then applying MOR methods, a reduced order model can also be inferred directly from the data. This is oftentimes referred to as *data-driven model order reduction* or *reduced order modelling*.

In the following, discrete-time state-space models are briefly introduced, and the applicability of the Eigensystem Realization Algorithm (ERA) [85–87] to the MIRACLE dataset is demonstrated.

4.2.1 State-space models

A discrete-time linear time-invariant dynamical system with $m \in \mathbb{N}$ inputs, $p \in \mathbb{N}$ outputs can be described by the matrix quadruple (A, B, C, D) with $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{p \times n}$, and $D \in \mathbb{R}^{p \times m}$, where $n \in \mathbb{N}$ is the order of the system. Its dynamics are governed by the discrete-time state equations

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) + Du(t), \end{aligned} \quad (22)$$

for input $u \in \ell_2^m$, output $y \in \ell_2^p$, and state $x \in \ell_2^n$. In our concrete case, u denotes the n_i -dimensional input signal at the source locations, and y denotes the n_o -dimensional output signal at the array microphones. Being an LTI system, the input-output behaviour of the state-space system is fully described by the convolution sum (21), where the impulse response $h \in \ell_2^{p \times m}$ is given by $h(t) = CA^{t-1}B$ in discrete-time. This can be easily verified from (22) by applying an impulse input $u(t) = \delta_{t0}$ with zero initial condition $x(0) = 0$. A frequency domain description can be obtained by applying the z -transform to (22). The z -transform of output $Y(z) \in \mathbb{C}^p$ is given by a multiplication of the z -transform of the input $U(z) \in \mathbb{C}^m$ with the transfer function

$$H(z) = C(zI - A)^{-1}B \in \mathbb{C}^{p \times m}. \quad (23)$$

4.2.2 Eigensystem Realization Algorithm

The transfer function in (23) is a matrix-valued rational function that can be approximated by a Padé approximation [88]. The moments of the transfer function for expansion at infinity, i.e. $h_i = \frac{d^i}{dz^i} H(z)|_{z=\infty}$, are referred to as the Markov parameters of the system. Intriguingly, the impulse response is equivalent to the sequence of Markov parameters in discrete time, i.e. $h(t) = h_t = CA^{t-1}B$. This connection is leveraged by ERA in the following way: Given a finite sequence $h \in \ell_2^{p \times m}$ of $2s - 1$, $s \in \mathbb{N}$ Markov parameters, a matching state-space model can be identified via the Hankel matrix of Markov parameters

$$\mathcal{H} = \begin{bmatrix} CB & CAB & \dots & CA^{s-1}B \\ CAB & CA^2B & \dots & CA^sB \\ \vdots & \vdots & \ddots & \vdots \\ CA^{s-1}B & CA^sB & \dots & CA^{2s-2}B \end{bmatrix} \in \mathbb{R}^{ps \times ms}. \quad (24)$$

For discrete-time systems, the Markov parameters can be conveniently obtained by measurements of the system impulse response. The Hankel matrix can be factored into the observability and controllability matrix $\mathcal{O} \in \mathbb{R}^{ps \times n}$ and $\mathcal{C} \in \mathbb{R}^{n \times ms}$:

$$\mathcal{H} = \underbrace{\begin{bmatrix} C \\ \vdots \\ CA^{s-1} \end{bmatrix}}_{=: \mathcal{O}} \underbrace{\begin{bmatrix} B & \dots & A^{s-1}B \end{bmatrix}}_{=: \mathcal{C}}. \quad (25)$$

From this factorization, a realization can be constructed via

$$A = \mathcal{O}_f^* \mathcal{O}_l, \quad B = \mathcal{C} \begin{bmatrix} I_m \\ 0 \end{bmatrix}, \quad C = [I_p \quad 0] \mathcal{O}, \quad (26)$$

where \mathcal{O}_f and \mathcal{O}_l denote the first and last $p(s-1)$ rows of the observability matrix $\mathcal{O} \in \mathbb{R}^{ps \times n}$, respectively [86, 87]. Classically, the factorization is computed from a (truncated) singular value decomposition (SVD), i.e. for a truncation order $r \leq \min\{ms, ps\}$

$$\mathcal{H} = [U_r \quad *] \begin{bmatrix} \Sigma_r & 0 \\ 0 & * \end{bmatrix} \begin{bmatrix} V_r^T \\ * \end{bmatrix}, \quad (27)$$

where $U_r \in \mathbb{R}^{ps \times r}$ comprises the first r left singular vectors as columns, $\Sigma_r = \text{diag}(\sigma_1, \dots, \sigma_r)$ the first r singular values with $\sigma_1 \leq \dots \leq \sigma_r$ and $V_r \in \mathbb{R}^{ms \times r}$ the first r right singular vectors as columns. By choosing $\mathcal{O} = U_r \Sigma_r^{1/2}$ and $\mathcal{C} = \Sigma_r^{1/2} V_r^T$, a reduced order realization (A_r, B_r, C_r) can be constructed via (26). For a basic stability and error analysis, the reader is referred to the classical sources [85–87]. A recent and thorough error analysis can be found in [89].

Using above theory, ERA is applied to scenario D1 of the MIRACLE dataset with $m = 1089$ inputs, $p = 64$ outputs. In order to improve the model quality, as argued in [83], the common dead time in the measurements is removed conservatively by truncating the first 68 samples. Additionally, to avoid unnecessary computations, only the first 20 ms of the impulse response are considered. Thus, after truncation, a total of $2s - 2 = 572$ samples are taken as input data for ERA. It is immediately obvious from (24) that the resulting Hankel matrix of dimension $18,240 \times 310,365$ can lead to computational problems with the outlined algorithm. One would already need 45.3GB to construct this matrix explicitly, and computing the SVD of such a large matrix is infeasible. There are several strategies [90–92] that can alleviate this computational burden in classical ERA by exploiting low-rank structures in the data and/or the Hankel matrix. A randomized SVD algorithm [93] was employed here as an approximate orthogonal decomposition instead of (27) by sampling the range of \mathcal{H} with 5000 normally distributed random vectors and a single subspace iteration. For details of this randomized ERA variant, see [83, 92].

It is important to emphasize that *multiple-input-multiple-output* reduced order models are identified for all input-output transmissions simultaneously. In order to facilitate a qualitative analysis of the model quality, Fig. 9 depicts the magnitude responses of the measurement data alongside the identified reduced order models for the single-channel transmission from the centremost source to the centremost microphone. It can be seen that the accuracy increases with increasing model order and decreases with frequency. Furthermore, we have noticed that the single-channel approximation accuracy tends to deteriorate for source or receiver locations with greater distances to the source and receiver planes, respectively.

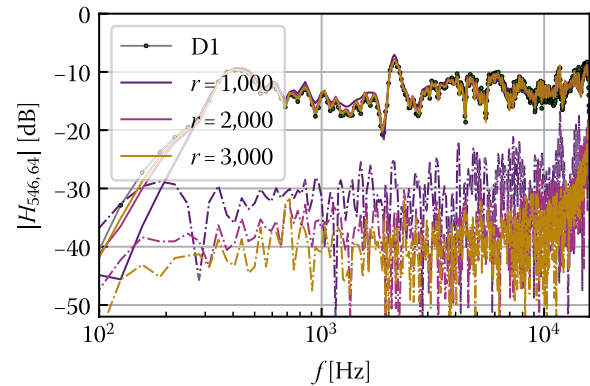


Fig. 9 Frequency response of the measured and modelled transfer functions (solid lines) and the error ε_r (dash-dotted lines) of different model orders for scenario D1 at the centremost location in the source and receiver plane

Therefore, an appropriate quantitative error criterion must encompass all transmission channels simultaneously. To this end, denote

$$h_{\text{avg}}(t) = \left(\frac{1}{n_i n_o} \sum_{i=1}^{n_i} \sum_{j=1}^{n_o} |h_{ij}(t)|^2 \right)^{1/2} = \frac{\|h(t)\|_F}{\sqrt{n_i n_o}} \quad (28)$$

and

$$\varepsilon_r(t) = \frac{\|h(t) - C_r A_r^{t-1} B_r\|_F}{\sqrt{n_i n_o}}, \quad (29)$$

as the RMS-averaged system and error impulse response, respectively, where $h(t) \in \ell_2^{64 \times 1,089}$ denotes the measured multichannel impulse response of the D1 scenario and A_r, B_r, C_r are the reduced order system matrices obtained by ERA for model orders $r = 1000, 2000, 3000$. The averaged responses are shown in Fig. 10 and it can be observed that the averaged error ε_r is still substantial around 3 ms for $r = 1000$, which is due to the aforementioned deterioration of approximation quality towards the edges of the source and receiver planes. For the higher orders $r = 2000, 3000$, this deviation is much less pronounced.

The potential benefit for model compression and real-time application is showcased in Table 6 in which the required memory and computational cost in Mega Floating Point Operations (MFLOPs) is listed alongside the relative approximation error

$$\hat{\varepsilon}_r = 10 \log_{10} \left(\frac{\sum_{t=1}^{n_d} \|h(t) - C_r A_r^{t-1} B_r\|_F^2}{\sum_{t=1}^{n_d} \|h(t)\|_F^2} \right). \quad (30)$$

In order to compare the cost of solving (22) with the state-space approach over the convolution sum in (21), the memory demand and computational cost of the Uniformly Partitioned Overlap-save (UPOLS) method [94] are also listed². The UPOLS method and the variants described in [94] are well-established and solve (21) via spectral multiplication of overlapping blocks of signals. It can be seen that the reduced order models are more memory efficient than the UPOLS method by factors between 5 – 30. A computational speedup for real-time simulation over the UPOLS method by factors between roughly 1.3 – 3.9 can be achieved for all models if the models are transformed into so-called *modal* or (*quasi-*) *diagonal* form which can be achieved via an eigenvalue decomposition of A_r . Details on the transformation can be found in [83].

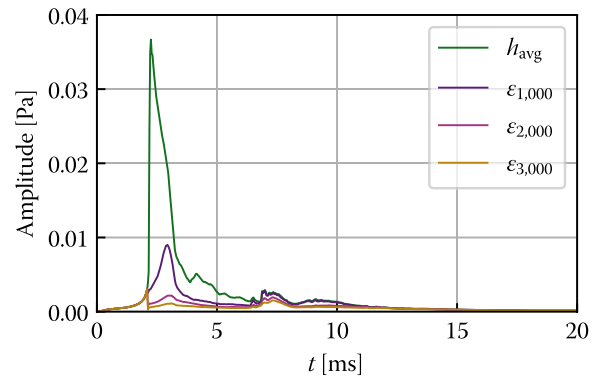


Fig. 10 RMS-averaged impulse response of the data and error systems for different model orders

Table 6 Memory demand in MB, computational cost in MFLOPs, and approximation error $\hat{\varepsilon}_r$ according to (30) for the UPOLS method with block-size $B = 64$ and $P = 16$ filter partitions and the reduced order models in different state-space representations. □: standard form, \: (quasi-) diagonal form

Model	Memory	Compr.	$\hat{\varepsilon}_r$	MFLOPs	Speedup
UPOLS	285.47MB	n/a	n/a	8.95	n/a
SS-□-1000	17.22MB	16.57	-10.45dB	4.30	2.08
SS-\-1000	9.23MB	30.92	-10.45dB	2.31	3.87
SS-□-2000	50.45MB	5.66	-17.28dB	12.61	0.71
SS-\-2000	18.46MB	15.46	-17.28dB	4.62	1.94
SS-□-3000	99.67MB	2.86	-19.78dB	24.91	0.36
SS-\-3000	27.07MB	10.31	-19.78dB	6.93	1.29

5 Conclusion

A publicly available large-scale RIR dataset of 856,128 single-channel impulse responses has been presented, which is particularly suited for applications in the field of microphone array signal processing and sound field reconstruction because of the dense spatial sampling of the receiver and source spaces. To the authors' knowledge, the MIRACLE dataset is the largest openly available RIR dataset yet.

The measured data have been shown to be spatially accurate and therefore provide an excellent basis for the development and statistical evaluation of data-driven modelling methods in acoustical engineering. The relevance, versatility, and applicability of the MIRACLE dataset have been demonstrated by means of two timely and diverging application examples.

In contrast to existing datasets, the MIRACLE dataset exhibits a notable limitation in environmental diversity, primarily constrained to free-field sound propagation scenarios with and without specular reflection. While the characteristics of the source and receiver predominantly shape the impulse response representation, enhancing environmental variability may be feasible through a

² For details with regard to the computational cost, we refer the reader to ([94], Sec. 5.2.1) for the UPOLS method and [83] for the state-space models.

semi-synthetic approach. This involves synthesizing corresponding RIRs and integrating them with the measured RIRs. To enrich the dataset, it is envisaged that the dataset will be extended in the near future by similar measurements of reverberant rooms. A persistent challenge lies in calibrating the source positions with comparable precision, mainly due to the diminished reliability of beamforming in reverberant environments. Addressing this challenge and broadening the dataset will be the focus of a forthcoming publication.

Abbreviations

DFT	Discrete Fourier Transform
ERA	Eigensystem Realization Algorithm
MFLOPs	Mega Floating Point Operations
MOR	Model order reduction
PDF	Probability density function
RIR	Room impulse response
SVD	Singular value decomposition
UPOLS	Uniformly Partitioned Overlap-save

Acknowledgements

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.

Authors' contributions

Experimental design, measurements, and post-processing were performed jointly by AK and AP. The manuscript was written by AK and AP. ES supervised the project. All authors read and approved the final manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. This research work was funded through the German Research Foundation (DFG) through projects [439144410](https://doi.org/10.1007/978-3-642-00000-0) and [504367810](https://doi.org/10.1007/978-3-642-00000-0).

Availability of data and materials

The dataset presented in this paper can be obtained from [doi:10.14279/depositonce-20106](https://doi.org/10.14279/depositonce-20106) under the CC BY-NC-SA 4.0 license authored by Adam Kujawski, Art J. R. Pelling, and Ennes Sarradj. The AcouPipe framework utilizing the MIRACLE dataset can be obtained from [10.5281/zenodo.10908288](https://doi.org/10.5281/zenodo.10908288).

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 22 December 2023 Accepted: 8 May 2024

Published online: 18 June 2024

References

1. K. Müller, F. Zotter, Auralization based on multi-perspective ambisonic room impulse responses. *Acta Acustica* **4**, (2020). <https://doi.org/10.1051/aacus/2020024>
2. M.R. Schroeder, New Method of Measuring Reverberation Time. *J. Acoust. Soc. Am.* **37**(3), 409–412 (2005). <https://doi.org/10.1121/1.1909343>
3. M.J. Bianco, P. Gerstoft, J. Traer, E. Ozanich, M.A. Roch, S. Gannot, C.-A. Deledalle, Machine learning in acoustics: theory and applications. *J. Acoust. Soc. Am.* **146**(5), 3590–3628 (2019). <https://doi.org/10.1121/1.5133944>
4. E. Fernandez-Grande, X. Karakonstantis, D. Caviedes-Nozal, P. Gerstoft, Generative models for sound field reconstruction. *J. Acoust. Soc. Am.* **153**(2), 1179–1190 (2023). <https://doi.org/10.1121/10.0016896>
5. A. Geldert, N. Meyer-Kahlen, S.J. Schlecht, in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Interpolation of Spatial Room Impulse Responses Using Partial Optimal Transport (IEEE, Rhodes Island, Greece, 2023), pp. 1–5. <https://doi.org/10.1109/ICASSP49357.2023.10095452>
6. Y. Haneda, Y. Kaneda, N. Kitawaki, Common-acoustical-pole and residue model and its application to spatial interpolation and extrapolation of a room transfer function. *IEEE Trans. Speech Audio Process.* **7**(6), 709–717 (1999). <https://doi.org/10.1109/89.799696>
7. F. Katzberg, R. Mazur, M. Maass, M. Böhme, A. Mertins, in *2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC)*. Spatial interpolation of room impulse responses using compressed sensing (Tokyo, 2018), pp. 426–430. <https://doi.org/10.1109/IWAENC.2018.8521390>
8. A. Ratnarajah, Z. Tang, R. Aralikatti, D. Manocha, in *Proceedings of the 30th ACM International Conference on Multimedia*. Mesh2ir: Neural acoustic impulse response generator for complex 3d scenes (Association for Computing Machinery, New York, Lisboa Portugal, 2022), pp. 924–933. <https://doi.org/10.1145/3503161.3548253>
9. S. Lee, H.-S. Choi, K. Lee, Yet another generative model for room impulse response estimation. in *2023 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)* (New Paltz, 2023), pp. 1–5. <https://doi.org/10.1109/WASPAA58266.2023.10248189>
10. N.J. Bryan, in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Impulse Response Data Augmentation and Deep Neural Networks for Blind Room Acoustic Parameter Estimation (IEEE, Barcelona, 2020), pp. 1–5. <https://doi.org/10.1109/ICASSP40776.2020.9052970>
11. P.-A. Grumiaux, S. Kitić, L. Girin, A. Guérin, A Survey of Sound Source Localization with Deep Learning Methods. *J. Acoust. Soc. Am.* **152**(1), 107–151 (2022). <https://doi.org/10.1121/10.0011809>
12. E. Guizzo, R.F. Gramaccioni, S. Jamili, C. Marinoni, E. Massaro, C. Medaglia, G. Nachira, L. Nucciarelli, L. Paglialonga, M. Pennese, S. Pepe, E. Rocchi, A. Uncini, D. Comminiello, in *Proceedings of the International Workshop on Machine Learning for Signal Processing (MLSP)*. L3DAS21 Challenge: Machine Learning for 3D Audio Signal Processing (IEEE, Gold Coast, 2021). <https://doi.org/10.1109/MLSP52302.2021.9596248>
13. E. Guizzo, C. Marinoni, M. Pennese, X. Ren, X. Zheng, C. Zhang, B. Masiero, A. Uncini, D. Comminiello, in *Proceedings of the ICASSP*. L3DAS22 Challenge: Learning 3D Audio Sources in a Real Office Environment (IEEE, Singapore, 2022), pp. 9186–9190. <https://doi.org/10.1109/ICASSP43922.2022.9746872>
14. K. Nagatomo, M. Yasuda, K. Yatabe, S. Saito, Y. Oikawa, in *Proceedings of the ICASSP*. Wearable Seld Dataset: Dataset For Sound Event Localization And Detection Using Wearable Devices Around Head (IEEE, Singapore, 2022), pp. 156–160. <https://doi.org/10.1109/ICASSP43922.2022.9746544>
15. M. Lee, J.-H. Chang, Deep neural network based blind estimation of reverberation time based on multi-channel microphones. *Acta Acustica U. Acustica* **104**, 486–495 (2018). <https://doi.org/10.3813/AAA.919191>
16. H. Gamper, I.J. Tashev, in *2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC)*. Blind reverberation time estimation using a convolutional neural network. (Tokyo, 2018), pp. 136–140. <https://doi.org/10.1109/IWAENC.2018.8521241>
17. M. Cobos, J. Ahrens, K. Kowalczyk, A. Politis, An overview of machine learning and other data-based methods for spatial audio capture, processing, and reproduction. *EURASIP J. Audio Speech Music. Process.* **2022**(1), 10 (2022). <https://doi.org/10.1186/s13636-022-00242-x>
18. J.G. Moreno-Torres, T. Raeder, R. Alaiz-Rodríguez, N.V. Chawla, F. Herrera, A unifying view on dataset shift in classification. *Pattern Recog.* **45**(1), 521–530 (2012). <https://doi.org/10.1016/j.patcog.2011.06.019>
19. S. Gannot, E. Vincent, S. Markovich-Golan, A. Ozerov, A consolidated perspective on multimicrophone speech enhancement and source separation. *IEEE/ACM Trans. Audio Speech Lang. Process.* **25**(4), 692–730 (2017). <https://doi.org/10.1109/TASLP.2016.2647702>
20. J. Huang, T. Bocklet, in *Proc. Interspeech 2019*. Intel Far-Field Speaker Recognition System for VOICES Challenge 2019 (2019), pp. 2473–2477. <https://doi.org/10.21437/Interspeech.2019-2894>

21. G. Bologni, R. Heusdens, J. Martinez, in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Acoustic reflectors localization from stereo recordings using neural networks (2021), pp. 1–5. <https://doi.org/10.1109/ICASSP39728.2021.9414473>
22. F. Lluis, P. Martinez-Nuevo, M. Bo Møller, S. Ewan Shepstone, Sound field reconstruction in rooms: Inpainting meets super-resolution. *J. Acoust. Soc. Am.* **148**(2), 649–659 (2020). <https://doi.org/10.1121/10.0001687>
23. S. Dilungana, A. Deleforge, C. Foy, S. Faisan, in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*. Learning-based estimation of individual absorption profiles from a single room impulse response with known positions of source, sensor and surfaces, vol. 263 (2021), pp. 5623–5630. <https://doi.org/10.3397/IN-2021-3186>
24. W. Yu, W.B. Kleijn, Room acoustical parameter estimation from room impulse responses using deep neural networks. *IEEE/ACM Trans. Audio Speech Lang. Process.* **29**, 436–447 (2021). <https://doi.org/10.1109/TASLP.2020.3043115>
25. E.J.G. Arcondoulis, Q. Li, S. Wei, Y. Liu, P. Xu, in *28th AIAA/CEAS Aeroacoustics Conference*. Experimental validation and performance analysis of deep learning acoustic source imaging methods (Southampton, 2022), <https://doi.org/10.2514/6.2022-2852>
26. A. Kujawski, E. Sarradj, Fast grid-free strength mapping of multiple sound sources from microphone array data using a Transformer architecture. *J. Acoust. Soc. Am.* **152**(5), 2543–2556 (2022). <https://doi.org/10.1121/10.0015005>
27. T. Lobato, R. Sottek, M. Vorländer, Deconvolution with neural grid compression: A method to accurately and quickly process beamforming results. *J. Acoust. Soc. Am.* **153**(4), 2073–2089 (2023). <https://doi.org/10.1121/10.0017792>
28. P. Srivastava, A. Deleforge, A. Politis, E. Vincent, in *Proc. INTERSPEECH 2023*. How to (Virtually) Train Your Speaker Localizer (ISCA, Dublin, 2023), pp. 1204–1208. <https://doi.org/10.21437/Interspeech.2023-1065>
29. P. Srivastava, Realism in virtually supervised learning for acoustic room characterization and sound source localization. Theses, Université de Lorraine (2023). <https://theses.hal.science/tel-04313405>. Accessed 22.05.24
30. A. Francl, J. McDermott, Deep neural network models of sound localization reveal how perception is adapted to real-world environments. *Nat. Hum. Behav.* **6**, 111–133 (2022). <https://doi.org/10.1101/2020.07.21.214486>
31. R. Scheibler, E. Bezzam, I. Dokmanic, in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Pyroomacoustics: A Python Package for Audio Room Simulation and Array Processing Algorithms (IEEE, Calgary, 2018), pp. 351–355. <https://doi.org/10.1109/ICASSP.2018.8461310>
32. Z. Tang, R. Aralikatti, A.J. Ratnarajah, D. Manocha, in *ACM SIGGRAPH 2022 Conference Proceedings*. Gwa: A large high-quality acoustic dataset for audio processing (Association for Computing Machinery, New York, 2022). <https://doi.org/10.1145/3528233.3530731>
33. C. Chen, C. Schissler, S. Garg, P. Kobernik, A. Clegg, P. Calamia, D. Batra, P.W. Robinson, K. Grauman, in *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks, (Online Conference)*. Sound-spaces 2.0: A simulation platform for visual-acoustic learning (2022). <https://doi.org/10.48550/arXiv.2206.08312>
34. D. Diaz-Guerra, A. Miguel, J.R. Beltran, gpuRIR: A python library for room impulse response simulation with GPU acceleration. *Multimedia Tools Appl.* **80**, 5653–5671 (2021). <https://doi.org/10.1007/s11042-020-09905-3>
35. M. Wang, S. Clarke, J.-H. Wang, R. Gao, J. Wu, in *Advances in Neural Information Processing Systems*, ed by A. Oh, T. Neumann, A. Globerson, K. Saenko, M. Hardt, S. Levine. SoundCam: A dataset for finding humans using room acoustics, vol. 36 (Curran Associates, Inc., New Orleans, 2023), pp. 52238–52264. <https://doi.org/10.25740/xq364hd5023>
36. P. Karolina, S.J. Schlecht, V. Välimäki, Dataset of impulse responses from variable acoustics room Arni at Aalto Acoustic Labs. Zenodo. (2022). <https://doi.org/10.5281/zenodo.6985104>
37. G. Götz, S.J. Schlecht, V. Pulkki, in *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*. A dataset of higher-order Ambisonic room impulse responses and 3D models measured in a room with varying furniture (IEEE, Bologna, 2021), pp. 1–8. <https://doi.org/10.1109/I3DA48870.2021.9610933>
38. E. Hadad, F. Heese, P. Vary, S. Gannot, in *Proceedings of the IWAENC*. Multichannel audio database in various acoustic environments (IEEE, Juan-les-Pins, 2014), pp. 313–317. <https://doi.org/10.1109/IWAENC.2014.6954309>
39. T. Dietzen, R. Ali, M. Taseska, T. Waterschoot, MYRIAD: a multi-array room acoustic database. *EURASIP J. Audio Speech Music Process.* (2023) <https://doi.org/10.1186/s13636-023-00284-9>
40. D. Di Carlo, P. Tandetnik, C. Foy, N. Bertin, A. Deleforge, S. Gannot, dEchorate: a calibrated room impulse response dataset for echo-aware signal processing. *EURASIP J. Audio Speech Music Process.* (2021). <https://doi.org/10.1186/s13636-021-00229-0>
41. S. Adavanne, J. Nikunen, A. Politis, T. Virtanen, TUT Tietotalo Ambisonic Impulse Response. Zenodo. 2018. <https://doi.org/10.5281/zenodo.1443539>
42. R. Merino-Martínez, P. Sijtsma, M. Snellen, T. Ahlefeldt, J. Antoni, C.J. Bahr et al., A review of acoustic imaging methods using phased microphone arrays. *CEAS Aeronaut. J.* **10**, 197–230 (2019). <https://doi.org/10.1007/s13272-019-00383-4>
43. S. Koyama, T. Nishida, K. Kimura, T. Abe, N. Ueno, J. Brunström, in *2021 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. MESHRRIR: A dataset of room impulse responses on meshed grid points for evaluating sound field analysis and synthesis methods (New Paltz, 2021), pp. 1–5. <https://doi.org/10.1109/WASPAA52581.2021.9632672>
44. J. Čmejla, T. Kounovský, S. Gannot, Z. Koldovský, P. Tandetnik, in *2020 28th European Signal Processing Conference (EUSIPCO)*. Mirage: Multichannel database of room impulse responses measured on high-resolution cubeshaped grid (2021), pp. 56–60. <https://doi.org/10.23919/Eusipco47968.2020.9287646>
45. A. Lu, E. Moore, A. Nallanthighal, K. Sarkar, M. Mittal, R.M. Corey, P. Smaragdis, A. Singer, in *2022 International Workshop on Acoustic Signal Enhancement (IWAENC)*. Mechatronic generation of datasets for acoustics research (IEEE, Bamberg, 2022), pp. 1–5. <https://doi.org/10.1109/IWAENC53105.2022.9914771>
46. R. Scheibler, J. Azcarreta, R. Beuchat, C. Ferry, in *2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC)*. Pyramic: Full Stack Open Microphone Array Architecture and Dataset (IEEE, Tokyo, 2018), pp. 226–230. <https://doi.org/10.1109/IWAENC.2018.8521337>
47. T. Ajdler, L. Sbaiz, M. Vetterli, The plenacoustic function and its sampling. *IEEE Trans. Signal Process.* **54**(10), 3790–3804 (2006). <https://doi.org/10.1109/TSP.2006.879280>
48. I. Szöke, M. Skácel, L. Mošner, J. Paliesek, J. Černocký, Building and evaluation of a real room impulse response dataset. *IEEE J. Sel. Top. Signal Process.* **13**(4), 863–876 (2019). <https://doi.org/10.1109/JSTSP.2019.2917582>
49. R. Stewart, M. Sandler, in *Proceedings of the ICASSP*. Database of omnidirectional and B-format room impulse responses. (IEEE, Dallas, 2010), pp. 165–168. <https://doi.org/10.1109/ICASSP2010.5496083>
50. A.V. Venkatakrishnan, P. Pertilä, M. Parviainen, in *Proceedings of the EUSIPCO*. Tampere university rotated circular array dataset (IEEE, Dublin, 2021), pp. 201–205. <https://doi.org/10.23919/EUSIPCO54536.2021.9616072>
51. J. Eaton, N.D. Gaubitch, A.H. Moore, P.A. Naylor, Estimation of room acoustic parameters: The ace challenge. *IEEE/ACM Trans. Audio Speech Lang. Process.* **24**(10), 1681–1693 (2016). <https://doi.org/10.1109/TASLP.2016.2577502>
52. K. Kinoshita, M. Delcroix, T. Yoshioka, T. Nakatani, E. Habets, R. Haeb-Umbach, V. Leutnant, A. Sehr, W. Kellermann, R. Maas, S. Gannot, B. Raj, in *2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. The verb challenge: A common evaluation framework for dereverberation and recognition of reverberant speech (2013), pp. 1–4. <https://doi.org/10.1109/WASPAA.2013.6701894>
53. D.T. Murphy, S. Shelley, in *Audio Engineering Society Convention OpenAIR: an interactive auralization web resource and database*. 129 (2010). <https://secure.aes.org/forum/pubs/conventions/?elib=15648>
54. M. Jeub, M. Schafer, P. Vary, in *2009 16th International Conference on Digital Signal Processing*. A binaural room impulse response database for the evaluation of dereverberation algorithms (2009). pp. 1–5. <https://doi.org/10.1109/ICDSP.2009.5201259>
55. S. Nakamura, K. Hiyane, F. Asano, T. Endo, Sound scene data collection in real acoustical environments. *J. Acoust. Soc. Jpn. (E)* **20**(3), 225–231 (1999). <https://doi.org/10.1250/ast.20.225>
56. P. Benner, A. Cohen, M. Ohlberger, K. Willcox (eds.) *Model Reduction and Approximation: Theory And Algorithms*. Computational Science and

- Engineering*, vol. 15 (Society for Industrial and Applied Mathematics, Philadelphia, 2017). <https://doi.org/10.1137/1.9781611974829>
57. S.L. Brunton, J.N. Kutz, in ed. by P. Benner, S. Griev-Talocia, A. Quarteroni, G. Rozza, W. Schilders, L.M. Silveira (eds.) *7 Data-driven methods for reduced-order modeling*, pp. 307–344. (De Gruyter, Berlin, Boston, 2021). <https://doi.org/10.1515/9783110671490-007>
 58. I.V. Gosea, S. Gugercin, C. Beattie, Data-Driven Balancing of Linear Dynamical Systems. *SIAM J. Sci. Comput.* **44**(1), 554–582 (2022). <https://doi.org/10.1137/21M1411081>
 59. B. Peherstorfer, S. Gugercin, K. Willcox, Data-driven reduced model construction with time-domain Loewner models. *SIAM J. Sci. Comput.* **39**(5), 2152–2178 (2017). <https://doi.org/10.1137/16M1094750>
 60. The MORwiki Community: MORwiki - Model Order Reduction Wiki. <http://modelreduction.org>. Accessed 05 Oct 2021
 61. J.G. Korvink, E.B. Rudnyi, in *Dimension Reduction of Large-Scale Systems. Lecture Notes in Computational Science and Engineering*, ed. by P. Benner, D.C. Sorensen, V. Mehrmann, Oberwolfach Benchmark Collection (Springer, Berlin, 2005), pp. 311–315. https://doi.org/10.1007/3-540-27909-1_11
 62. Benchmark Examples for Model Reduction - SLICOT. <http://slicot.org/20-site/126-benchmark-examples-for-model-reduction>. Accessed 29 Sept 2021
 63. E. Sarradj, in *6th Berlin Beamforming Conference. A Generic Approach To Synthesize Optimal Array Microphone Arrangements*. (Gesellschaft zur Förderung angewandter Informatik (GFal), Berlin, 2016), pp. 1–12
 64. M. Rébillat, R. Hennequin, É. Corteel, B.F.G. Katz, Identification of cascade of hammerstein models for the description of nonlinearities in vibrating devices. *J. Sound Vib.* **330**(5), 1018–1038 (2011). <https://doi.org/10.1016/j.jsv.2010.09.012>
 65. A. Farina, in *108th AES Convention. Simultaneous Measurement of Impulse Response and Distortion with Swept-sine technique*. (Paris, 2000). <https://secure.aes.org/forum/pubs/conventions/?elib=10211>
 66. S. Müller, P. Massarani, Transfer-function measurement with sweeps. *J. Audio Eng. Soc.* **49**(6), 443–471 (2001)
 67. M. Müller-Trapel, On the practical application of the impulse response measurement method with swept-sine signals in building acoustics. *J. Acoust. Soc. Am.* **148**(4), 1864–1878 (2020). <https://doi.org/10.1121/10.0001916>
 68. L.L. Beranek, T.J. Mellow, *Acoustics: Sound Fields and Transducers*, 1st edn (Academic Press, and imprint of Elsevier Amsterdam, 2012)
 69. K. Prawda, S.J. Schlecht, V. Välimäki, Robust selection of clean swept-sine measurements in non-stationary noise. *J. Acoust. Soc. Am.* **151**(3), 2117–2126 (2022). <https://doi.org/10.1121/10.0009915>
 70. SciPy v1.11.4 Manual. https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.resample_poly.html. Accessed 18 Dec 2023
 71. P.C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion. SIAM Monographs on Mathematical Modeling and Computation*, p. 247. (SIAM, Philadelphia, 1998). <https://doi.org/10.1137/1.9780898719697>
 72. A. Farina, in *122nd AES Convention. Advancements in impulse response measurements by sine sweeps* (Vienna, 2007), p. 21. <https://www.aes.org/e-lib/browse.cfm?elib=14106>
 73. M. Holters, T. Corbach, U. Zölzer, in *12th Int. Conference on Digital Audio Effects (DAFx-09). Impulse response measurement techniques and their applicability in the real world*. (2009). https://www.dafx.de/paper-archive/details.php?id=1u-OdqvbtbweDYmNY2_kuA. Accessed 13 Mar 2024
 74. H. Tokuno, O. Kirkeby, P.A. Nelson, H. Hamada, Inverse Filter of Sound Reproduction Systems Using Regularization. *IEICE Trans. Fundam. A* **80**(5), 809–820 (1997)
 75. S.G. Norcross, M. Bouchard, G.A. Soulodre, in *Audio Engineering Society Convention 121. Inverse filtering design using a minimal-phase target function from regularization* (Audio Engineering Society, San Francisco, 2006). <https://www.aes.org/e-lib/browse.cfm?elib=13763>. Accessed 13 Mar 2024
 76. O. Cramer, The variation of the specific heat ratio and the speed of sound in air with temperature, pressure, humidity, and CO₂ concentration. *J. Acoust. Soc. Am.* **93**(5), 2510–2516 (1993). <https://doi.org/10.1121/1.405827>
 77. R.S. Davis, Equation for the determination of the density of moist air (1981/91). *Metrologia* **29**(1), 67 (1992). <https://doi.org/10.1088/0026-1394/29/1/008>
 78. E. Sarradj, Three-dimensional acoustic source mapping with different beamforming steering vector formulations. *Adv. Acoust. Vib.* (2012). <https://doi.org/10.1155/2012/292695>
 79. E. Parzen, On estimation of a probability density function and mode. *Ann. Math. Stat.* **33**(3), 1065–1076 (1962)
 80. B.W. Silverman, *Density Estimation for Statistics and Data Analysis. Chapman & Hall/CRC monographs on statistics and applied probability* (Chapman and Hall, London, 1986). <https://cds.cern.ch/record/1070306>
 81. A. Kujawski, A.J.R. Pelling, S. Jekosch, E. Sarradj, A framework for generating large-scale microphone array data for machine learning. *Multimed. Tools Appl.* (2023). <https://doi.org/10.1007/s11042-023-16947-w>
 82. A. Kujawski, A.J.R. Pelling, S. Jekosch, C. Kayser, E. Sarradj, adku1173/acoupipe: v23.11. Zenodo. (2023). <https://doi.org/10.5281/zenodo.10405973>
 83. A.J.R. Pelling, E. Sarradj, Efficient forced response computations of acoustical systems with a state-space approach. *Acoustics* **3**(3), 581–593 (2021). <https://doi.org/10.3390/acoustics3030037>
 84. G. Nijse, M. Verhaegen, B. Schutter, D. Westwick, N. Doelman, in *Proceedings of the 1999 International Symposium on Active Control of Sound and Vibration (ACTIVE 99)*, ed. by S. Douglas. State space modeling in multichannel active control systems (Fort Lauderdale, 1999), pp. 909–920. https://www.dsc.tudelft.nl/~bdeschutter/pub/rep/99_11.pdf
 85. L. Silverman, Realization of linear dynamical systems. *IEEE Trans. Autom. Control.* **16**(6), 554–567 (1971). <https://doi.org/10.1109/TAC.1971.1099821>
 86. S. Kung, in *Proceedings of the 12th Asilomar Conference on Circuits, Systems and Computers. A new identification and model reduction algorithm via singular value decomposition* (1978). pp. 705–714
 87. J.-N. Juang, R.S. Pappa, An eigensystem realization algorithm for modal parameter identification and model reduction. *J. Guid. Control. Dyn.* **8**(5), 620–627 (1985). <https://doi.org/10.2514/3.20031>
 88. A.C. Antoulas, *Approximation of Large-Scale Dynamical Systems. Advances in Design and Control* (Society for Industrial and Applied Mathematics, Philadelphia, 2005). <https://doi.org/10.1137/1.9780898718713>
 89. S. Oymak, N. Ozay, Revisiting ho-kalman-based system identification: robustness and finite-sample analysis. *IEEE Trans. Autom. Control.* **67**(4), 1914–1928 (2022). <https://doi.org/10.1109/TAC.2021.3083651>
 90. B. Kramer, S. Gugercin, Tangential interpolation-based eigensystem realization algorithm for mimo systems. *Math. Comput. Model. Dyn. Syst.* **22**(4), 282–306 (2016). <https://doi.org/10.1080/13873954.2016.1198389>
 91. B. Kramer, A.A. Gorodetsky, System identification via cur-factored hankel approximation. *SIAM J. Sci. Comput.* **40**(2), 848–866 (2018). <https://doi.org/10.1137/17M1137632>
 92. R. Minster, A.K. Saibaba, J. Kar, A. Chakraborty, Efficient algorithms for eigensystem realization using randomized svd. *SIAM J. Matrix Anal. Appl.* **42**(2), 1045–1072 (2021). <https://doi.org/10.1137/20M1327616>
 93. N. Halko, P.G. Martinsson, J.A. Tropp, Finding Structure with Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions. *SIAM Rev.* **53**(2), 217–288 (2011). <https://doi.org/10.1137/090771806>
 94. F. Wefers, *Partitioned Convolution Algorithms for Real-Time Auralization. Aachener Beiträge Zur Technischen Akustik*, vol. Band 20. (Logos Verlag Berlin GmbH, Berlin, 2015)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.