Database of Simulated Room Impulse Responses for Acoustic Sensor Networks Deployed in Complex Multi-Source Acoustic Environments

Rene Glitza, Luca Becker, Alexandru Nelus, and Rainer Martin
Institute of Communication Acoustics
Ruhr-Universität Bochum
44780 Bochum, Germany
{firstname.lastname}@ruhr-uni-bochum.de

***Abstract*—In this work we present a large set of simulated room impulse responses for a multi-room apartment. The simulated apartment models a real vacation apartment for which a recorded set of audio data has already been made available in the context of the DCASE challenges. The impulse responses were rendered using a dense grid of sources and receivers by means of a hybrid auralization algorithm based on a low-order image-source method and deterministic cone tracing. The proposed data set can be used to generate a wide variety of acoustic scenes which, in turn, can benefit numerous data-demanding machine-learning algorithms. An example application is provided in the form of an unsupervised clustering algorithm that groups microphone nodes around dominant sound sources across the entire apartment.**

***Index Terms*—data set, virtual acoustics, machine learning, clustering**

I. Introduction

With the wide dissemination of wireless communications and affordable, embedded acoustic sensors, (wireless) acoustic sensor networks (ASNs) have gained increased scientific interest in recent years. Challenging applications of ASNs [3], [22] are, among others, acoustic source localization [18], acoustic event localization, detection, and classification [1], [10], [12], and speech enhancement [16]. These can be deployed in a variety of environments, such as smart-cities [6] and smart-homes [9], where additional information regarding the spatial relation between microphones and active sound sources by means of clustering algorithms can further improve their utility [2], [13], [19], [20].

The development of these ASN-based applications is critically dependent on the availability of recorded or simulated data. For comprehensive and realistic acoustic scenarios, multiple sources and microphones are required along with a complex environmental layout, thus making the recording or simulation task a non-trivial endeavor.

Open source projects like the RIR Generator [11] or Pyroomacoustics [24] aim to provide simulated room impulse responses by implementing the image source method. While the RIR Generator only allows shoebox-shaped rooms, Pyroomacoustics features an extension of the image model to arbitrary polyhedra [5]. With increasing the number of bounding planes in the geometric model and the order of reflections, the computational time also increases exponentially. Therefore, ray tracing [28] has been added recently for a hybrid impulse response generation. The implementation, however, is not thoroughly tested and does not feature all physical effects which are, among others, diffraction, transmission, and source directivity.

Therefore, many machine learning algorithms, such as [19], are developed and tested using the room impulse responses (RIRs) of rather simple shoebox-shaped rooms. These simulations are often sufficient for validating an algorithm's core function, but are not suitable for assessing its real-world capabilities. Recording RIRs or sounds and voices in real environments [14], [17], [26] is a frequent solution which can be a very resource-consuming. One database created for this purpose is the SINS database [8]. It consists of continuous audio recordings of one person living in a vacation home over a period of one week. Parts of this data set have been successfully used in the 2018 DCASE challenge [9] and in the development of algorithms for detecting and classifying daily activities [10].

However, when it comes to the development of state-of-the-art machine learning algorithms, especially deep neural networks, a large amount of diverse data is usually required. This cannot be easily provided by audio recordings of individual people or groups. Therefore, the data needs to be simulated using realistic conditions that go beyond simple shoebox rooms. Then, data sets containing clean speech, such as LibriSpeech [21], or other audio sources can be employed in conjunction with simulated RIRs to generate extensive and more realistic acoustic environments.

The main contribution of this paper is the introduction of a free and open source data set containing room impulse responses from a complex acoustic environment, based on the apartment layout described in [8], which can be used to simulate an acoustic sensor network with a dense grid of sources and microphones. It features multiple rooms that have different reverberation conditions, open doors, furniture, and decorations along with a grid of sources ($\geq$ 200) and

This work has been supported by the German Research Foundation (DFG) - Project Number 282835863.

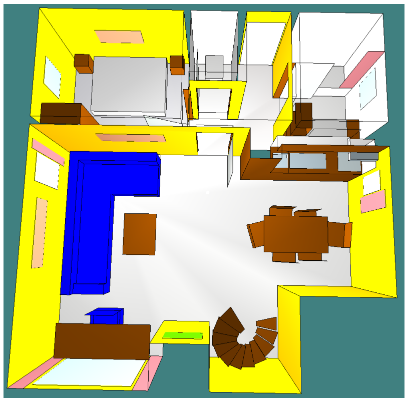


Fig. 1: 3-D apartment model rendered by CATT-Acoustic. Different materials are highlighted using different colors.

sensors (≥ 300), with the latter grouped in small quadratic arrays of four sensors each. The impulse responses have been simulated using a geometric cone tracing algorithm integrated in the CATT Acoustics software package [7], a thoroughly validated industry-grade acoustic simulation package, requiring extensive computational resources over a period of several months. To the best knowledge of the authors, this is the first such data set that provides simulated room impulse responses and handles a complex apartment layout with multiple rooms, furniture, and a variety of absorbing surfaces. Adaptations of the data set have already been incorporated in several research works, featuring ASN sensor node clustering [2], [19], [20], or wake-word detection (WWD) systems [13].

II. Acoustic Simulation and Settings

Prior to the creation of this database, the strengths and weaknesses of several geometrical and wave-based acoustic modeling approaches were evaluated. While wave-based methods [28] like finite element method (FEM), boundary element method (BEM), and finite-difference time-domain (FDTD) are based on solving the wave equations, they are computationally very expensive and less suitable for large audio bandwidths and highly complex environments [25]. Geometrical acoustics (GA) algorithms, such as image-source method (ISM), ray tracing (RT), and cone tracing (CT), rely on the assumption that sound propagates as rays to reduce computation time. This simplification is valid at higher frequencies, where the wavelength of sound is short compared to surface dimensions, but introduces potential approximation errors at lower frequencies [23].

In order to simulate thousands of impulse responses up to a frequency of 16 kHz, wave-based methods are, due to their calculation effort, currently not a suitable option. Instead, the Universal Cone Tracer (TUCT2) algorithm provided by CATT-Acoustic [7] appears to offer a suitable framework for this work's purposes, because it takes care of many GA limitations such as diffraction and interference. It features a combination of the deterministic image source method for complex acoustic scenarios, along with specular and diffuse cone tracing for higher orders of reflections. The diffuse reflections are computed for each 1/1 octave band from 125 Hz up to 16 kHz, along with auto-edge scattering for edge diffraction effects. Direct sound and first-order specular reflections are deterministic, and the first-order reflection will interfere with the direct sound for pressure impulse response. The first order diffuse reflections are deterministic but are incoherent, so they will not directly interfere with the direct sound and first-order specular reflections. Transmission through walls is frequency-dependent as well, and handled the same way as the diffuse reflection bouncing back on the other side of a room boundary.

The following simulation settings have been used to create the database: Scattering for both surface and edges, detailed auralization (alg. 2) with cone split-up [7], $15\,\mathrm{k}$ initial cones, length of impulse responses as suggested by an automatic method, air absorption turned on, 1st order diffraction [27], and a sampling rate of $f_S = 44.1\,\mathrm{kHz}$.

III. Simulation Environment and Setup

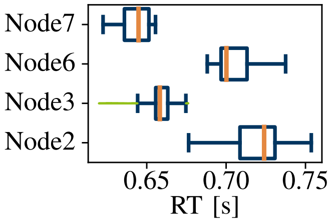
A. 3D Apartment Model

The database was simulated using a 3D model of a vacation home with five different rooms: a combined living room and kitchen, bathroom, toilet, bedroom, and hall. This model is based on the apartment used to record the SINS database [8], but is not an exact replica. Information made available by the authors of the SINS database include the floor plan, total floor area, a video of the apartment[1], and the recordings from the SINS database.

For achieving most realistic outcomes with geometrical acoustics, among other considerations, objects and wall boundaries should not be broken up into many small segments but rather kept in one large surface. Then, scattering and absorption coefficients need to be adjusted to obtain the same acoustical properties.

In order to implement the 3D model as closely as possible w.r.t. the real apartment, the following steps have been carried out: 1) The floor plan from [8] has been loaded into SketchUp and resized so the floor area of the entire apartment equals $S = 50.1\,\mathrm{m}^2$. 2) To this first model, walls, doors, stairs, and big furniture, which are shown on the floor plan, have been added. The stairs, chairs, tables, and doors are implemented as double-sided planes. 3) Supplementary objects and furniture are added as visible in a video[1]. 4) Standard absorption coefficients for assumed materials have been added to all surfaces. 5) The reverberation time (RT) of the rooms based on the recordings has been estimated. For the speech samples in the SINS Database recorded in the living room, a blind reverberation time estimator based on a simple statistical model of sound decay [15] has been used. The resulting blind estimations are shown in Fig. 2. Note that in the presence of ambient noise (as in these recordings), the algorithm in [15] tends to overestimate RTs. 6) The model has then been fine-tuned by informal listening experiments and reverberation time comparisons in multiple iterations by expert listeners.

[1] G. Dekkers, https://youtu.be/ylXM4KmoIIc



Node	Position
Node7	Left of living room door
Node6	Right of living room door
Node3	Left of living room stairs
Node2	Right of living room desk

Fig. 2: RT estimations using [15] for signals obtained from data set [8]. For each node, 22 speech signals are considered.

TABLE I: Room areas, broadband T_{30} in seconds, and T_{30} for selected octave bands. RTs are calculated from nodes close to the microphones positions of the SINS Database. Reverberation times are averaged over multiple sources.

	Area	Mean	64 Hz	250 Hz	1 kHz	4 kHz
Living room	$33.09\,\mathrm{m}^2$	0.46	0.44	0.44	0.50	0.50
Bedroom	$7.47\,\mathrm{m}^2$	0.31	0.33	0.36	0.36	0.35
Toilet	$1.40\,\mathrm{m}^2$	0.35	0.35	0.38	0.40	0.41
Hall	$3.69\,\mathrm{m}^2$	0.41	0.54	0.47	0.47	0.46
Bathroom	$4.32\,\mathrm{m}^2$	0.44	0.60	0.45	0.49	0.49

In these listening experiments, we selected speakers from the LibriSpeech data set [21] with similar long term spectra as the speakers available in the SINS data set.

The final result of the apartment is displayed in Fig. 1 with the floor area and reverberation times of several octave bands stated in Table I.

B. Microphone Grid

In the apartment as shown in Fig. 1, a grid of microphone array nodes and sources have been added. Each node consists of four omnidirectional microphones with a flat frequency response, which are positioned at the corners of a $5\,\mathrm{cm}^2$ square (radius $r \approx 3.536\,\mathrm{cm}$). They are rotated randomly for every node with an orientation angle defined relative to the receiver with ID 0. This also allows for applications using uniform circular arrays, as described in [29]. Sources are single omnidirectional speakers with a flat frequency response.

The grid of sensor nodes has a higher resolution in the living room than in the other rooms. In the living room microphone nodes are spaced $60\,\mathrm{cm}$ apart and they are spaced $75\,\mathrm{cm}$ in the other rooms. All nodes which are closer than $20\,\mathrm{cm}$ to surfaces have been deleted. Sources are spaced $40\,\mathrm{cm}$ apart in the living room and $1\,\mathrm{m}$ in the other rooms. All sources which are closer than $20\,\mathrm{cm}$ to walls are not considered. All grids are centered within the x and y boundaries of each room. Additional sources have been added in the middle of each door frame. The source-node grids are illustrated in Fig. 3 with the respective random orientations of nodes. The height (z-coordinate) of any node and source is $1.35\,\mathrm{m}$.

IV. Database Content

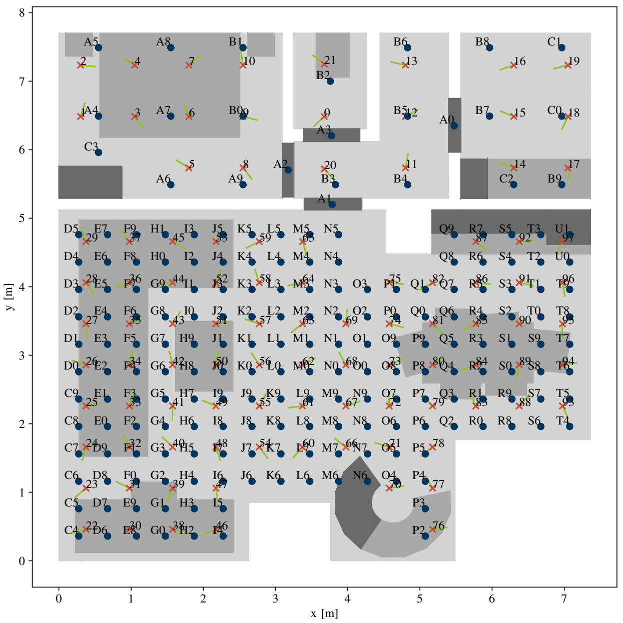


Fig. 3: 2D top view of sources (blue dot) and receiver (red cross) grid, including node array orientation (green line). Shades indicate geometry with light gray being the floor, darker gray indicating objects below $1.35\,\mathrm{m}$, and darkest gray for objects above $1.35\,\mathrm{m}$. IDs for sources start with a letter and for nodes with a number.

The RIRs are exported as .MAT, .WAV, and .BIN files, whereby the length of each RIR is different and based on reverberation time estimations. All audio files are single channel and have a sample rate of $f_S = 44.1\,\mathrm{kHz}$ and an amplitude resolution of $32\,\mathrm{bit}$.

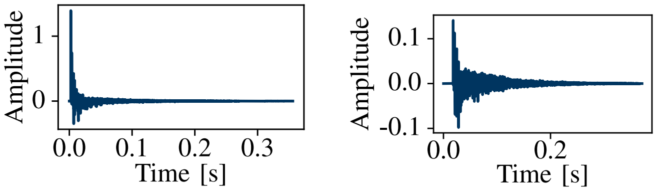
All three versions of the data set (WAV, BIN, MAT) are available separately. They each contain room impulse responses from 202 sources and 98 nodes, resulting in 19796 source-receiver pairs and 79184 RIRs in their respective file format. The file size for the WAV and BIN version is $4.9\,\mathrm{GB}$ each, and $9.7\,\mathrm{GB}$ for the MAT version. In addition to the impulse responses, MAT files are available which contain meta data on the simulation runs. A JSON file contains information about every source receiver pair. This includes the room of the receiver node, a unique microphone-source pair ID, node ID, source ID, node position, node orientation in rad, source position, position of all four microphones belonging to the node, and file names of all four RIRs belonging to the source-receiver pair.

V. Evaluation

A. Analysis of Room Impulse Responses

Time-domain plots of two exemplary RIRs are shown in Fig. 4, where a significant decrease in direct sound amplitude can be noticed from node 60 to node 21 (see Fig. 3 for node numbers). For an analysis of the RIRs, we compute the direct-to-reverberant ratio (DRR). All DRR values are estimated according to

$$\mathrm{DRR} = 10\log_{10}\left(\textstyle\sum_{k=k_0-\sigma}^{k_0+\sigma} h^2(k) / \sum_{k=k_0+\sigma}^{\infty} h^2(k)\right)\,, \qquad (1)$$



(a) RIR from M6 to 60.

(b) RIR from M6 to 21.

Fig. 4: RIRs from source M6 to receivers in various distances.

where $h(k)$ denotes the RIR, k_0 is the discrete time index of the direct sound component, and $\sigma = 2$ ms is half the width of the direct sound summation window. Along with DRR values, we also consider the critical distances relative to sources in the living room. At the critical distance, the sound pressure level of the direct sound is equal to the sound pressure level of the reverberated sound field. It can be approximated using Sabine's equation [4] as

$$r_H = 0.1\,\mathrm{m} \cdot \sqrt{V\,\mathrm{m}^{-3}/\pi \cdot T_{30}\,\mathrm{s}^{-1}}, \qquad (2)$$

where $V = 79.27\,\mathrm{m}^3$ is the volume of the living room and $T_{30} = 0.46\,\mathrm{s}$ is its reverberation time.

As can be seen in Fig. 5, distant receivers experience smaller DRR values, while receivers in close proximity to the critical distance exhibit higher DRR values. Nodes inside the critical distance of a source have DRRs greater than $0\,\mathrm{dB}$. It can be observed that receivers with a direct line of sight to, e.g., source N3 have significantly higher DRR than the ones obscured by walls. Additionally, Fig. 6 points out that node 20 in the corridor above source N3 shows a higher DRR than node 65 in the living room, although the latter is closer to the source. This is explained by the fact that node 20 still receives direct sound from the source while much of the reverberation is confined to the living room. A similar observation has been made in real measurements for a transition of receivers between coupled rooms [17]. In the small room next to the bedroom, only sound that travels through both doors can be picked up by node 0, with direct sound being a significant portion of it. We conclude that the RIRs presented in this work show realistic behavior and constitute a suitable data set for various tasks related to room acoustics and ASN.

B. An Exemplary Application: Node Clustering

In order to evaluate the utility of the database, we simulate a complex acoustic environment with four simultaneously active sound sources and 41 microphone nodes. In this environment we perform clustering as in [2], [13], [19], [20] for 200 scenarios where, for each scenario, all microphone nodes and sources are randomly chosen. For this purpose, each microphone node is equipped with a pre-trained neural autoencoder. These autoencoders are then re-trained using a spectral representation of local recordings, and cluster-membership is determined by the cosine-similarities of the respective node's weight-updates. Due to the fact that the quality of the clustering is directly dependent on the RIRs, we regard the cluster-to-source distances (CTS) [20] as a means to evaluate the provided database. The CTS-distances can be understood as the euclidean distances between each cluster centroid and source location, divided by the average of all unique source-to-source distances. Averaged CTS-distances for 200 simulations are presented in Table II. In general, the utilized clustering mechanism generates a variable amount of clusters per scenario, the maximum number being 14. Predicated on this notion, we sort the clusters c_k in the CTS-matrix depending on their distances to the first four sound sources s_z. A large CTS-distance implies that the corresponding cluster is relatively far away from a source, while a small CTS-distance indicates proximity. Specifically, the first four clusters exhibit both properties: while being close to one source, such a cluster is further away to most other sources. Thus, performing clustering using our data set generates source-dominated clusters that show good potential for subsequent ASN-based tasks.

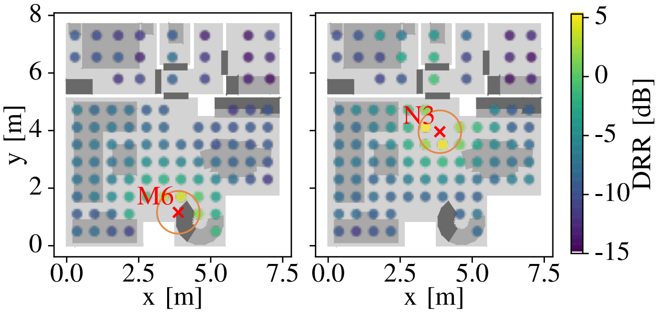


Fig. 5: Direct-to-reverberant ration from sources M6 and N3 (red cross) with critical distance (orange circle) to all microphones (colored dots). The color of the microphone dots represents the DRR value for each microphone in dB.

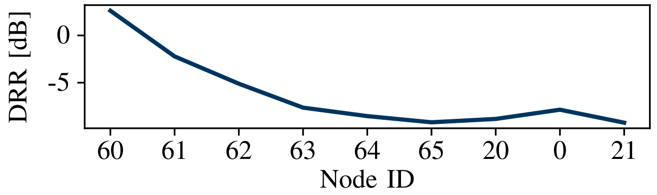


Fig. 6: DRR from source M6 to nodes 60 - 21.

TABLE II: Normalized cluster-to-source distance from cluster c_k to source s_z, averaged over 200 simulated scenarios. The average cluster size is 6.25. The first four clusters c_1 - c_4 are grouped around sources s_1 - s_4, respectively, while the remaining nodes are assigned to clusters with centroids farther away from any source. Clusters c_9 - c_{14} are not shown.

	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8
s_1	**0.35**	1.1	1.14	1.17	1.01	1.39	1.41	1.49
s_2	0.93	**0.51**	1.21	1.25	1.28	1.11	1.46	1.54
s_3	0.93	1.09	**0.61**	1.23	1.25	1.38	1.08	1.52
s_4	0.85	0.97	1.08	**0.77**	1.33	1.28	1.34	1.27

VI. Conclusions

In this paper, we have proposed a new data set with simulated room impulse responses in a complex multi-source acoustic environment with several coupled rooms. The apartment model emulates the geometry and acoustic properties of the apartment used for recording the SINS data set [8]. Starting with blind reverberation time estimations, we further fine-tuned the room model in informal listening experiments. The simulation was performed using a combination of the deterministic image-source method and specular and diffuse cone tracing methods. A grid of 202 sources and 98 nodes with four omnidirectional microphones each result in 7918 RIRs available in the data set.

Numerical analysis shows a consistent decrease of the direct-to-reverberant ratio with distance. Clustering experiments have demonstrated the use in an acoustic sensor network based application. Additionally, RIRs from this data set have already been used in [2], [19], [20] underpinning its utility.

In the future, we will release additional versions of the data set that will include the presence of humans and closed doors.

VII. Data Set Download

The data set including example code is available under a GNU General Public License v3.0 at https://github.com/Jearde/asn-database, downloadable in either WAV, binary, or MAT format. Supplementary data includes 3D models of room geometries in various file formats, the simulation source files for CATT-Acoustic, as well as a description of the data set.

Audio samples for the scenario described in Section V-A are provided as well.

References

[1] S. Adavanne, A. Politis, and T. Virtanen, "A multi-room reverberant dataset for sound event localization and detection," *arXiv preprint arXiv:1905.08546*, 2019.

[2] L. Becker, A. Nelus, R. Glitza, and R. Martin, "Accelerated unsupervised clustering in acoustic sensor networks using federated learning and a variational autoencoder," in *2022 Int. Workshop on Acoustic Signal Enhancement (IWAENC)*, 2022, pp. 1–5.

[3] A. Bertrand, "Applications and trends in wireless acoustic sensor networks: A signal processing perspective," in *2011 18th IEEE symp. on communications and vehicular technology in the Benelux*, 2011, pp. 1–6.

[4] J. Blauert and N. Xiang, *Acoustics for Engineers*. Berlin, Heidelberg: Springer, 2009.

[5] J. Borish, "Extension of the image model to arbitrary polyhedra," *The Journal of the Acoustical Society of America*, vol. 75, no. 6, pp. 1827–1836, 1984.

[6] M. Cartwright, A. E. M. Mendez, J. Cramer, V. Lostanlen, H.-H. Wu, J. Salamon, and J. Bello. (2019) DCASE 2019: Urban Sound Tagging. Accessed February, 2020. [Online]. Available: http://dcase.community/challenge2019/task-urban-sound-tagging.

[7] B.-I. Dalenbäck, *TUCT v2.0e:1*, CATT, Mariagatan 16A, SE-41471 Gothenburg, Sweden, 2019. [Online]. Available: http://www.catt.se

[8] G. Dekkers, S. Lauwereins, B. Thoen, M. W. Adhana, H. Brouckxon, T. van Waterschoot, B. Vanrumste, M. Verhelst, and P. Karsmakers, "The SINS database for detection of daily activities in a home environment using an acoustic sensor network," in *Proc. Workshop on Detection and Classification of Acoustic Scenes and Events*, 2017, pp. 32–36.

[9] G. Dekkers, L. Vuegen, T. van Waterschoot, B. Vanrumste, and P. Karsmakers, "DCASE 2018 Challenge - Task 5: Monitoring of domestic activities based on multi-channel acoustics," KU Leuven, Tech. Rep., 2018.

[10] J. Ebbers, L. Drude, R. Haeb-Umbach, A. Brendel, and W. Kellermann, "Weakly supervised sound activity detection and event classification in acoustic sensor networks," in *2019 IEEE 8th Int. Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 2019, pp. 301–305.

[11] E. A. Habets, "Room impulse response generator," *Technische Universiteit Eindhoven, Tech. Rep*, vol. 2, no. 2.4, p. 1, 2006.

[12] J. Hu, Y. Cao, M. Wu, Q. Kong, F. Yang, M. D. Plumbley, and J. Yang, "Sound event localization and detection for real spatial sound scenes: Event-independent network and data augmentation chains," *arXiv preprint arXiv:2209.01802*, 2022.

[13] T. Koppelmann, L. Becker, A. Nelus, R. Glitza, L. Schönherr, and R. Martin, "Clustering-based Wake Word Detection in Privacy-aware Acoustic Sensor Networks," in *Proc. Interspeech 2022*, 2022, pp. 719–723.

[14] Y. Koyama, K. Shigemi, M. Takahashi, K. Shimada, N. Takahashi, E. Tsunoo, S. Takahashi, and Y. Mitsufuji, "Spatial data augmentation with simulated room impulse responses for sound event localization and detection," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 8872–8876.

[15] H. Löllmann, E. Yilmaz, and P. Vary, "An Improved Algorithm for Blind Reverberation Time Estimation," in *Proceedings of Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, 2010, pp. 1–4.

[16] S. Markovich-Golan, A. Bertrand, M. Moonen, and S. Gannot, "Optimal distributed minimum-variance beamforming approaches for speech enhancement in wireless acoustic sensor networks," *Signal Processing*, vol. 107, pp. 4–20, 2015.

[17] T. McKenzie, S. J. Schlecht, and V. Pulkki, "Acoustic analysis and dataset of transitions between coupled rooms," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 481–485.

[18] G. F. Miller, A. Brendel, W. Kellermann, and S. Gannot, "Misalignment recognition in acoustic sensor networks using a semi-supervised source estimation method and markov random fields," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 766–770.

[19] A. Nelus, R. Glitza, and R. Martin, "Estimation of microphone clusters in acoustic sensor networks using unsupervised federated learning," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 761–765.

[20] A. Nelus, R. Glitza, and R. Martin, "Unsupervised clustered federated learning in complex multi-source acoustic environments," in *29th European Signal Processing Conference (EUSIPCO)*, 2021, pp. 1115–1119.

[21] V. Panayotov, G. Chen, D. Povey, and S. Khudanpur, "Librispeech: An ASR corpus based on public domain audio books," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2015, pp. 5206–5210.

[22] S. Pasha, C. Ritz, and J. Lundgren, "A survey on ad hoc signal processing: Applications, challenges and state-of-the-art techniques," in *2019 IEEE Int. Symposium on Signal Processing and Information Technology (ISSPIT)*. IEEE, 2019, pp. 1–6.

[23] L. Savioja and U. P. Svensson, "Overview of geometrical room acoustic modeling techniques," *The Journal of the Acoustical Society of America*, vol. 138, no. 2, pp. 708–730, 2015.

[24] R. Scheibler, E. Bezzam, and I. Dokmanić, "Pyroomacoustics: A python package for audio room simulation and array processing algorithms," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 351–355.

[25] S. Siltanen, T. Lokki, and L. Savioja, "Rays or waves? understanding the strengths and weaknesses of computational room acoustics modeling techniques," in *Proceedings of the International Symposium on Room Acoustics, ISRA*, 2010, pp. 29–31.

[26] K. Sridhar, R. Cutler, A. Saabas, T. Parnamaa, M. Loide, H. Gamper, S. Braun, R. Aichner, and S. Srinivasan, "ICASSP 2021 acoustic echo cancellation challenge: Datasets, testing framework, and results," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 151–155.

[27] U. P. Svensson, R. I. Fred, and J. Vanderkooy, "An analytic secondary source model of edge diffraction impulse responses," *The Journal of the Acoustical Society of America*, vol. 106, no. 5, pp. 2331–2344, 1999.

[28] M. Vorländer, *Auralization*. Berlin, Heidelberg: Springer, 2020.

[29] J. Zhao and C. Ritz, "Co-prime circular microphone arrays and their application to direction of arrival estimation of speech sources," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 800–804.