DEEP, DATA-DRIVEN MODELING OF ROOM ACOUSTICS: LITERATURE REVIEW AND RESEARCH PERSPECTIVES

Toon van Waterschoot
Department of Electrical Engineering (ESAT-STADIUS), KU Leuven, Leuven, Belgium

ABSTRACT

Our everyday auditory experience is shaped by the acoustics of the indoor environments in which we live. Room acoustics modeling is aimed at establishing mathematical representations of acoustic wave propagation in such environments. These representations are relevant to a variety of problems ranging from echo-aided auditory indoor navigation to restoring speech understanding in cocktail party scenarios. Many disciplines in science and engineering have recently witnessed a paradigm shift powered by deep learning (DL), and room acoustics research is no exception. The majority of deep, data-driven room acoustics models are inspired by DL-based speech and image processing, and hence lack the intrinsic space-time structure of acoustic wave propagation. More recently, DL-based models for room acoustics that include either geometric or wave-based information have delivered promising results, primarily for the problem of sound field reconstruction. In this review paper, we will provide an extensive and structured literature review on deep, data-driven modeling in room acoustics. Moreover, we position these models in a framework that allows for a conceptual comparison with traditional physical and data-driven models. Finally, we identify strengths and shortcomings of deep, data-driven room acoustics models and outline the main challenges for further research.

Keywords: *room acoustics, deep learning, data-driven modeling, literature review*

1. INTRODUCTION

People spend about 90 % of their time indoors [1], hence our auditory system has been trained to perceive and process sound only after it has been "shaped" by the acoustics of our inside living environments. Whereas room acoustics is potentially beneficial for perceptual tasks and experiences including human indoor navigation by means of echolocation [2, 3] and audience engagement in concert halls [4], it may also hamper speech understanding [5] and contribute to the cocktail party effect [6], in particular for hard-of-hearing people [7]. In addition to this distinction of room acoustics being desired or undesired, problems involving room acoustics can further be classified into forward and inverse problems. Forward problems are aimed at predicting the sound field (or sound signal at one position) in a room, when information about the sound sources (*i.e.*, location, directivity, pressure signal) and the room (*i.e.*, geometry, boundary properties) are given. Inverse problems instead focus on the retrieval of source or room parameters from sound field measurements.

Various perspectives on room acoustics have emerged over the past century. The physical perspective considers room acoustics as a space-time or space-frequency physical phenomenon that can be modeled as an interior boundary value problem, and has been applied primarily to forward problems [8, 9]. The architectural perspective, propelled by the seminal work of Sabine [10], features compact, empirical or statistical descriptors of room acoustics such as reverberation time (T60), early decay time (EDT), and clarity index (C50) [11]. These descriptors capture the temporal, spectral and/or spatial acoustic behavior of the room at a macroscopic level. Their estimation from sound field measurements is at the core of many inverse problems [12, 13]. In the signals and systems perspective, point-to-point acoustic responses within a room are modeled as linear time-invariant (LTI) systems, which can be represented as linear filters [14, 15] or state-space models [16]. This perspective is used for both forward and inverse problems considered above, and often involves the processing of acoustic data acquired from microphone signal measurements. Finally, since 2015 a data science perspective on room acoustics has materialized. In this perspective, forward and inverse problems are tackled as predictive learning problems from simulated and/or measured data. As opposed to LTI system modeling, deep learning models involving nonlinear operations are commonly used [17–21].

**Corresponding author*: *tvanwate@esat.kuleuven.be.*

2. LITERATURE REVIEW

2.1 Concise overview of traditional room acoustics models

Traditional room acoustics models (*i.e.*, those developed before the deep learning era) can be categorized along two dimensions: the first dimension represents the degree to which models are based on room measurements ("data-driven models") or on physics first principles ("physical models"), while the second dimension indicates whether the model assumes a geometric or wave-based sound behavior. The leftmost part of Figure 1 illustrates this categorization and guides the reader through the below concise overview of traditional models.

The architectural and signals and systems perspectives on room acoustics mostly rely on impulse response models, *i.e.*, the room impulse response (RIR) [22] and spatial RIR (SRIR) [23]. Other data-driven models are modal response models, *e.g.*, common-acoustical-pole and zero (CAPZ) [24], orthonormal basis function (OBF) [25], and parallel filter (PF) models [26]. Data-driven models are more effective for room acoustics applications when a physical prior is included in the model structure [27]. A geometric prior is used in RIR decomposition models such as the spatial decomposition method (SDM) [28], whereas wave-based priors are used in wave decomposition models, *e.g.*, wave field analysis (WFA) [29], plane-wave decomposition (PWD) [30], and spherical-harmonic decomposition (SHD) [31]. Data-driven models incorporating a physical prior based on the boundary integral equation (BIE) [32] include the equivalent source model (ESM) [33–35] and the boundary integral operator state-space (BIOSS) model [36]. Even though the BIE is a wave-based prior, it asymptotically admits a geometric solution [37], hence these models are capable of representing both wave-based and geometric sound behavior. Purely physical models have mainly been used in virtual acoustics and include reflection path models [38] (*e.g.*, image source models (ISM) [39], ray tracing (RT) [40], and beam tracing (BT) [41]), delay networks (*e.g.*, feedback delay networks (FDN) [42], digital waveguide networks (DWN) [43], and scattering delay networks (SDN) [44]), and discretized partial differential equation (PDE) models (*e.g.*, boundary element (BEM) [45], finite element (FEM) [46], finite difference (FDM) [9, 47], and finite volume models (FVM) [8]).

2.2 Purely data-driven deep learning models

Over the past decade, many scientific disciplines have witnessed a paradigm shift driven by deep learning (DL). This paradigm shift has also occurred in room acoustics research, and as a consequence, the state of the art has meanwhile fundamentally changed, as illustrated in the rightmost part of Figure 1. Not surprisingly, research on deep, data-driven room acoustics modeling has started by considering inverse problems. In these problems, the input is typically a reverberant sound signal from which information on the underlying room acoustics or sound source is to be inferred. Most of the relevant literature considers speech source signals, hence allowing to use DL model structures that have previously shown their merit in speech analysis problems such as automatic speech recognition [48, 49], speaker identification [50], or speech emotion recognition [51, 52].

The first research efforts in deep, data-driven room acoustics modeling have been focused on inverse problems in which a high-dimensional input (*e.g.*, a reverberant speech signal captured by one or more microphones) is transformed into a low-dimensional output, considering model structures that are often adopted from other DL application areas such as image processing and computer vision. The most widely studied problem in this context is the estimation of room acoustic parameters from reverberant speech [13]: reverberation time (T60) [53–66], clarity index (C50) [58, 59, 63–65, 67, 68], direct-to-reverberant ratio (DRR) [56, 57, 63–65], early decay time (EDT) [58], definition (D50) [58], center time (Ts) [58], speech transmission index (STI) [59, 64], speech intelligibility index (SII) [59], room volume [60, 66, 69], and source distance [70]. In a similar context, the classification of rooms from reverberant speech has been investigated [71, 72]. Even if initially multi-layer perceptron (MLP) model structures were used [53, 57], it was soon realised that long-term temporal characteristics of room reverberation are highly relevant, motivating the use of recurrent neural networks (RNNs) with long-short term memory (LSTM) layers [54] or their bidirectional version (BLSTM) [67]. Alternatively, convolutional neural networks (CNNs) have been considered [55–60], and in particular their combination with recurrent structures into convolutional recurrent neural networks (CRNNs) [61–64, 68], possibly including an attention mechanism [70, 71], has become an established model structure for the room acoustic parameter estimation task. More recently, the popular Transformer model structure [73] has also been used for this task [65, 66, 69].

A similar evolution from MLP-based over C(R)NN-based to Transformer-based models has occurred for other inverse problems in which a low-dimensional output is sought. In room inference problems, the estimation of room boundary characteristics [74] (*e.g.*, reflection coefficients or impedance values) has been tackled with MLPs [75, 76], CNNs [75], and CRNNs [77], whereas the room geometry can be inferred by means of CNNs [76, 78, 79] or CRNNs [80, 81], possibly with an attention mechanism [82]. For the room geometry inference task, also a combination of a residual network (ResNet) with an autoencoder (AE) has been proposed [83], while Transformer-based models were observed to perform below expectations [84]. Finally, sound source localization from multi-microphone signal observations has been achieved with MLPs [85], CNNs [85] (in particular its 3D-CNN extension [86]), CRNNs [85] (possibly combined with a classical MUSIC localization method [87]), AEs [85], and Transformers [85].

In acoustic signal enhancement problems, the DL model input (*i.e.*, speech corrupted by artefacts such as echo, reverberation, or interfering speech) and output (*i.e.*, enhanced speech) have similar dimensions. As the focus of these problems is on the (speech) source signal, the room acoustics modeling aspect is often implicit, hence model structures originally proposed for speech analysis problems have also found their way here. However, in order to produce a high-dimensional output, the models typically lack pooling layers or, more recently, adopt an

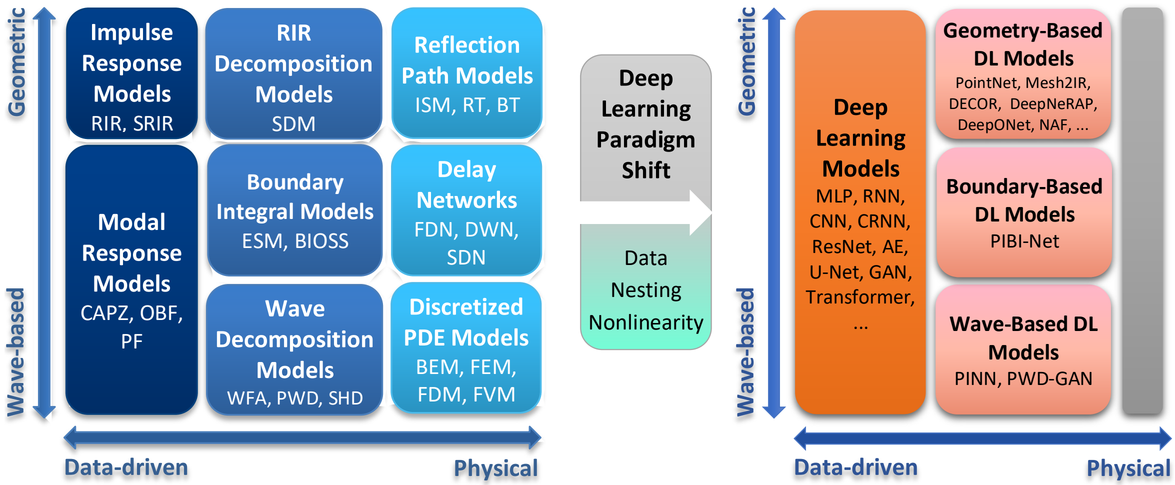

Figure 1. Classification of room acoustics models before (left) and after (right) the deep learning paradigm shift. Acronyms are defined in the text.

encoder-decoder structure reminiscent of U-Net, variational AE, and Transformer architectures. Deep, data-driven approaches to acoustic echo cancellation, until recently an archetypical example of linear estimation theory [88], have been extensively studied, resulting in algorithms based on RNNs [89–92], possibly with an attention mechanism [93, 94], CRNNs [95–101], possibly combined with a Kalman filter [102]. In many of these algorithms, the use of gated recurrent units (GRUs) was found beneficial [92, 94, 100–102]. In acoustic howling suppression, narrowband interferences due to closed-loop instability need to removed [103], which has been achieved with CRNNs [104], possibly involving GRUs and an attention mechanism [105], and with a hybrid RNN-Kalman filter [106]. For dereverberation [107], traditionally considered one of the most challenging inverse problems in room acoustics, various model structures have been considered. In the single-microphone case, MLPs [108], RNNs [109], possibly conditioned on the room's energy decay curve (EDC) [110], CNNs [111], possibly with an attention mechanism [112], generative adversarial networks (GANs) [113], and a combination of U-Net and Transformer models [114] have been used. Also the single-channel sound source separation problem [115] has been addressed jointly with the dereverberation problem using RNNs [116] and U-Nets [117]. Whereas the multi-microphone dereverberation problem seems somewhat underexplored, with the exception of the T60-conditioned MLP model proposed in [118], multi-channel source separation has received more attention with the development of CNN [119], CRNN [120], and U-Net-based models [121, 122], often designed to work with a spatially preprocessed input such as interaural level/phase differences [119], ambisonic signal components [121], and direction-dependent [121] or location-dependent [122] features. The state-of-the-art SpatialNet approach achieves joint multi-channel speech separation and enhancement by clustering of acoustic transfer functions in a combined Transformer and CNN model structure [123]. For the problem of active noise control [124], MLPs [125], CRNNs [126, 127], and their combination [128] have been considered. Finally, the problem of acoustic matching (*i.e.*, transforming a reverberant sound signal such that it perceptually matches a different room than the one where it was recorded), is typically addressed by means of generative models such as WaveNets [129] or GANs [130].

Yet another category of problems are those where the desired DL model output is a RIR, a set of RIRs, or a sound pressure field. These problems are fundamentally different from those discussed above, and often require the use of a generative DL model structure. In blind acoustic system identification [131], the aim is to estimate RIRs from reverberant speech observations, and an encoder-decoder structure is generally preferred: the encoder serves to estimate a latent room acoustics representation from the reverberant signal (using CNNs [132, 133] or ResNets [134]), while the decoder adheres to a GAN structure to generate RIRs from the latent embedding [132–135]. In the context of room equalization [136], a U-Net model structure has been proposed to generate a spatially averaged room transfer function (RTF) from a set of measured RTFs [137]. In sound field reconstruction [138], one aims to use a set of sound pressure (or RIR) measurements to predict the sound pressure (or RIR) at receiver positions where no measurements have been made. This problem has been addressed with U-Nets [139, 140], conditionally invertible neural networks (CINNs) [141], and Transformers [142]. In artificial reverberation synthesis [15], RIRs are generated for given source and receiver positions and room specifications. Deep, data-driven approaches to this problem include the estimation of FDN parameters from measured RIRs using CNNs [143] and the generation of binaural RIRs for moving receivers with VAEs [144]. Note that even though artificial reverberation synthesis is traditionally a different problem than SFR, the distinction between both problems becomes somewhat ambiguous in a data-driven setting, as the input to both problems consists of room acoustics measurements. Finally, the problem of upmixing RIRs to SRIRs [145] has been tackled with VAEs [19] and GANs [146].

2.3 Deep learning models with physical priors

Despite their excellent performance for parameter or signal estimation problems, the models discussed above are not suitable (or not optimal) for problems in which RIRs or space-time sounds pressure fields need to be computed, due to the fact that the intrinsic structure of acoustic wave propagation (*e.g.*, representation of time delays, preservation of space-time relations) is not efficiently represented in the model structure. This has given rise to the development of deep, data-driven models that are partially physics-informed, either via geometric or wave-based priors.

Geometry-based DL models (not to be confused with "Geometric DL" which refers to a particular instance of neural networks [147]) have been developed for artificial reverberation synthesis by (1) explicitly encoding the scene geometry with MLPs (PointNet [148]) or GANs with graph convolution layers (Mesh2IR [149, 150]), (2) conditioning the DL model on the source/receiver positions and room geometry with GANs (Fast-RIR [151]) or (V)AEs [152, 153], (3) conditioning the DL model on early reflections with GRU-based CNNs [154] or AEs (DECOR [155]), or (4) using geometric information only with MLPs (DeepNeRAP [156]) or neural operators replacing the PDE (DeepONet [157]). The strategy of conditioning the DL model on the receiver position has also been used in the context of sound field reconstruction with U-Net [158] and dynamic kernel [159] model structures. A more prominent geometry-based room acoustics model for sound field reconstruction is the Neural Acoustic Field (NAF) [160], which represents the continuous mapping from source/receiver pairs to RIRs by means of MLPs, conditioned on local geometric information present at the source and receiver locations. Extensions of NAF involve the inclusion of boundary information, *i.e.*, boundary geometry (INRAS [161]) and material properties (NACF [162]), and joint audio-visual scene generation (Few-ShotRIR [163], NeRAF [164], AV-NeRF [165], SOAF [166]). A similar concept named Novel-View Acoustic Synthesis (NVAS) has also been applied to video-aided sound field reconstruction [167] and to the inverse problem of joint source localization and recovery [168]. Finally, geometry-based DL has also resulted in novel sound source localization methods by (1) integrating the ray-space transform into CNNs [169], (2) exploiting shift-equivariance [170] and rotation-equivariance [171] in CNNs, (3) enforcing geometric proximity in the embedding space [172], and (4) combining pairwise networks conditioned on microphone pair positions (Neural-SRP [173]).

Wave-based DL models are primarily based on physics-informed neural networks (PINNs). The vanilla PINN consists of an MLP-based deep neural network (DNN) which is trained on a loss function that includes a regularization term imposing the acoustic wave equation, and has been successfully applied to artificial reverberation synthesis [174] and sound field reconstruction [175, 176]. PINN variations for sound field reconstruction involve the use of trigonometric activation functions (SIREN [177–179]), regularization with the Helmholtz equation [180], and a deep kernel method regularized by the wave equation [181]. Finally, another wave-based DL model for sound field reconstruction consists in the estimation of PWD coefficients by means of a GAN (PWD-GAN [182, 183]).

3. RESEARCH PERSPECTIVES

From the above literature study, it is clear that deep learning holds great potential for room acoustics modeling. We end this review paper by formulating three prominent perspectives for future research.

Firstly, data availability remains the first and foremost requirement in the development of deep, data-driven models. Over the past five years, we have witnessed a strong rise in the availability of high-quality, large-scale, and open-access datasets of room acoustics measurements in diverse conditions and measurement setups, *e.g.*, [139, 184–198]. The main challenge in developing additional datasets is to reconcile the conflicting requirements of designing setups that correspond to realistic audio scenes (*e.g.*, with moving and directional sources and microphones) while allowing for accurate data labeling (*e.g.*, in terms of source and microphone positions and orientations).

Secondly, a more fundamental understanding is required of why deep, data-driven models appear to be highly suitable for room acoustics modeling. Two key elements of deep, data-driven models are their nested model structure consisting of composed functions referred to as layers, and the nonlinear activation functions used in these layers. It is not well understood how nesting and nonlinearity contribute to the accurate and efficient modeling of room acoustics, in particular as these properties seem to contradict the widely accepted premise that room acoustics can be modeled as a linear, time-invariant process.

Thirdly, upon comparing our classification of traditional models with deep, data-driven models for room acoustics in Figure 1, there is an apparent gap between geometry-based and wave-based deep, data-driven room acoustics models. Given the asymptotic geometric interpretation of the wave-based BIE discussed above [37], the key to combining geometric and wave-based information may lie in the inclusion of boundary information both in the model structure and in the model training strategy of DL models. A first and promising result in this direction consists in the use of physics-informed boundary integral networks (PIBI-Nets) for sound field reconstruction [199].

4. ACKNOWLEDGEMENTS

This research work was carried out at the ESAT Laboratory of KU Leuven, in the frame of KU Leuven internal funds C14/21/075 and C3/23/056, FWO projects G0A0424N and S005525N, and the AI Research Program of the Flemish Government. The scientific responsibility is assumed by its authors.

5. REFERENCES

[1] N. E. Klepeis *et al.*, "The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants," *J. Expo. Sci. Environ. Epidemiol.*, vol. 11, pp. 231–252, 2001.

[2] A. J. Kolarik *et al.*, "A summary of research investigating echolocation abilities of blind and sighted humans," *Hearing Res.*, vol. 310, pp. 60–68, 2014.

[3] D. Pelegrín-García *et al.*, "Localization of a virtual wall by means of active echolocation by untrained sighted persons," *Applied Acoustics*, vol. 139, pp. 82–92, 2018.

[4] T. Lokki *et al.*, "Engaging concert hall acoustics is made up of temporal envelope preserving reflections," *J. Acoust. Soc. Am.*, vol. 129, no. 6, pp. EL223–EL228, 2011.

[5] T. Houtgast and H. J. M. Steeneken, "The modulation transfer function in room acoustics as a predictor of speech intelligibility," *Acta Acustica united with Acustica*, vol. 28, no. 1, pp. 66–73, 1973.

[6] A. W. Bronkhorst, "The cocktail party phenomenon: a review of research on speech intelligibility in multiple-talker conditions," *Acta Acustica united with Acustica*, vol. 86, no. 1, pp. 117–128, 2000.

[7] B. Kollmeier and J. Kiessling, "Functionality of hearing aids: state-of-the-art and future model-based solutions," *Int. J. Audiology*, vol. 57, p. S3–S28, 2016.

[8] S. Bilbao *et al.*, "Finite volume time domain room acoustics simulation under general impedance boundary conditions," *IEEE/ACM Trans. Audio Speech Language Process.*, vol. 24, no. 1, pp. 161–173, 2016.

[9] B. Hamilton and S. Bilbao, "FDTD methods for 3-D room acoustics simulation with high-order accuracy in space and time," *IEEE/ACM Trans. Audio Speech Language Process.*, vol. 25, no. 11, pp. 2112–2124, 2017.

[10] W. C. Sabine, *Collected Papers on Acoustics*. Cambridge, MA, USA: Harvard University, 1922.

[11] J. S. Bradley, "Review of objective room acoustics measures and future needs," *Applied Acoustics*, vol. 72, no. 10, pp. 713–720, 2011.

[12] P. Kendrick *et al.*, "Monaural room acoustic parameters from music and speech," *J. Acoust. Soc. Am.*, vol. 124, no. 1, pp. 278–287, 2008.

[13] J. Eaton *et al.*, "Estimation of room acoustic parameters: The ACE challenge," *IEEE/ACM Trans. Audio Speech Language Process.*, vol. 24, no. 10, pp. 1681–1693, 2016.

[14] J. Mourjopoulos and M. A. Paraskevas, "Pole and zero modeling of room transfer functions," *J. Sound Vib.*, vol. 146, no. 2, pp. 281–302, 1991.

[15] V. Välimäki *et al.*, "Fifty years of artificial reverberation," *IEEE Trans. Audio Speech Language Process.*, vol. 20, no. 5, pp. 1421–1448, 2012.

[16] K. MacWilliam *et al.*, "State-space estimation of spatially dynamic room impulse responses using a room acoustic model-based prior," *Frontiers in Signal Process.*, vol. 4, 2024.

[17] Z. H. Michalopoulou *et al.*, "Introduction to the special issue on machine learning in acoustics," *J. Acoust. Soc. Am.*, vol. 150, no. 4, pp. 3204–3210, 2021.

[18] M. Cobos *et al.*, "An overview of machine learning and other data-based methods for spatial audio capture, processing, and reproduction," *EURASIP J. Audio Speech Music Process.*, vol. 2022, 2022. Article No. 10.

[19] W. Yu, *The estimation of acoustic parameters and representations based on room impulse responses*. PhD thesis, Delft University of Technology, The Netherlands, 2024.

[20] G. Götz, *Data-driven room-acoustic modelling*. PhD thesis, Aalto University, Finland, 2024.

[21] X. Karakonstantis, *Data-driven methods for large-scale sound field acquisition and analysis*. PhD thesis, Technical University of Denmark, Denmark, 2024.

[22] H. Kuttruff, *Room Acoustics*. Spon Press, 5 ed., 2009.

[23] J. Merimaa and V. Pulkki, "Spatial impulse response rendering I: Analysis and synthesis," *J. Audio Eng. Soc.*, vol. 53, no. 12, pp. 1115–1127, 2005.

[24] Y. Haneda, S. Makino, and Y. Kaneda, "Common acoustical pole and zero modeling of room transfer functions," *IEEE Trans. Speech Audio Process.*, vol. 2, no. 2, pp. 320–328, 1994.

[25] G. Vairetti *et al.*, "A scalable algorithm for physically motivated and sparse approximation of room impulse responses with orthonormal basis functions," *IEEE/ACM Trans. Audio Speech Language Process.*, vol. 25, no. 7, pp. 1547–1561, 2017.

[26] J. Rämö, V. Välimäki, , and B. Bank, "High-precision parallel graphic equalizer," *IEEE/ACM Trans. Audio Speech Language Process.*, vol. 22, no. 12, pp. 1894–1904, 2014.

[27] T. van Waterschoot, G. Rombouts, and M. Moonen, "Optimally regularized adaptive filtering algorithms for room acoustic signal enhancement," *Signal Processing*, vol. 88, no. 3, pp. 594–611, 2008.

[28] S. Tervo *et al.*, "Spatial decomposition method for room impulse responses," *J. Audio Eng. Soc.*, vol. 61, no. 1/2, pp. 17–28, 2013.

[29] A. J. Berkhout, D. de Vries, and J. J. Sonke, "Array technology for acoustic wave field analysis in enclosures," *J. Acoust. Soc. Am.*, vol. 102, no. 5, pp. 2757–2770, 1997.

[30] F. Pinto and M. Vetterli, "Space-time-frequency processing of acoustic wave fields: Theory, algorithms, and applications," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4608–4620, 2010.

[31] D. P. Jarrett, E. A. P. Habets, and P. A. Naylor, *Theory and Applications of Spherical Microphone Array Processing*. Springer, 2017.

[32] S. Marburg, "A unified approach to finite and boundary element discretization in linear time-harmonic acoustics," in *Computational Acoustics of Noise Propagation in Fluids - Finite and Boundary Element Methods* (S. Marburg and B. Nolte, eds.), Springer, 2008.

[33] S. Lee, "The use of equivalent source method in computational acoustics," *J. Comput. Acoust.*, vol. 25, no. 1, 2017. Article No. 1630001.

[34] N. Antonello *et al.*, "Room impulse response interpolation using a sparse spatio-temporal representation of the sound field," *IEEE/ACM Trans. Audio Speech Language Process.*, vol. 25, no. 10, pp. 1929–1941, 2017.

[35] N. Antonello *et al.*, "Joint acoustic localization and dereverberation through plane wave decomposition and sparse regularization," *IEEE/ACM Trans. Audio Speech Language Process.*, vol. 27, no. 12, pp. 1893–1905, 2019.

[36] R. Ali *et al.*, "A state-space framework for the boundary integral equation," tech. rep., KU Leuven, Leuven, Belgium, 2025.

[37] R. Ali *et al.*, "Relating wave-based and geometric acoustics using a stationary phase approximation," in *Proc. Forum Acusticum 2023*, (Turin, Italy), 2023.

[38] L. Savioja and P. Svensson, "Overview of geometrical room acoustic modeling techniques," *J. Acoust. Soc. Am.*, vol. 138, no. 2, pp. 708–730, 2015.

[39] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, 1979.

[40] A. Krokstad, S. Strøm, and S. Sørsdal, "Calculating the acoustical room response by the use of a ray tracing technique," *J. Sound Vib.*, vol. 8, no. 1, pp. 118–125, 1968.

[41] T. Funkhouser *et al.*, "A beam tracing method for interactive architectural acoustics," *J. Acoust. Soc. Am.*, vol. 115, no. 2, pp. 739–756, 2004.

[42] J.-M. Jot and A. Chaigne, "Digital delay networks for designing artificial reverberators," in *Preprints AES 90th Conv.*, (Paris, France), 1991. AES Preprint 3030.

[43] D. Rocchesso and J. O. Smith, "Circulant and elliptic feedback delay networks for artificial reverberation," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 1, pp. 51–63, 1997.

[44] E. D. Sena *et al.*, "Efficient synthesis of room acoustics via scattering delay networks," *IEEE/ACM Trans. Audio Speech Language Process.*, vol. 23, no. 9, pp. 1478–1492, 2015.

[45] M. R. Bai, "Application of BEM (boundary element method)-based acoustic holography to radiation analysis of sound sources with arbitrarily shaped geometries," *J. Acoust. Soc. Am.*, vol. 92, no. 1, pp. 533–549, 1992.

[46] T. Shuku and K. Ishihara, "The analysis of the acoustic field in irregularly shaped rooms by the finite element method," *J. Sound Vib.*, vol. 29, no. 1, pp. 67–76, 1973.

[47] D. Botteldooren, "Finite-difference time-domain simulation of low-frequency room acoustic problems," *J. Acoust. Soc. Am.*, vol. 98, no. 6, pp. 3302–3308, 1995.

[48] D. Yu and L. Deng, *Automatic speech recognition: A deep learning approach*. Springer, 2015.

[49] P. P. Parada *et al.*, "Reverberant speech recognition exploiting clarity index estimation," *EURASIP J. Adv. Signal Process.*, vol. 2015, 2015. Article No. 54.

[50] Z. Bai and X.-L. Zhang, "Speaker recognition based on deep learning: An overview," *Neural Networks*, vol. 140, pp. 65–99, 2021.

[51] P. Tzirakis, J. Zhang, and B. W. Schuller, "End-to-end speech emotion recognition using deep neural networks," in *Proc. 2018 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '18)*, (Calgary, AB, Canada), pp. 5089–5093, 2018.

[52] D. Tang *et al.*, "End-to-end speech emotion recognition using a novel context-stacking dilated convolution neural network," *EURASIP J. Audio, Speech, Music Process.*, vol. 2021, 2021. Article No. 18.

[53] X. Xiao *et al.*, "Learning to estimate reverberation time in noisy and reverberant rooms," *in Proc. INTERSPEECH 2015, Dresden, Germany*, pp. 3431–3435, 2015.

[54] J. F. Santos and T. H. Falk, "Blind room acoustics characterization using recurrent neural networks and modulation spectrum dynamics," in *Proc. AES 60th Conf.*, (Leuven, Belgium), 2016.

[55] Z. Tang *et al.*, "Scene-aware audio rendering via deep acoustic analysis," *IEEE Trans. Vis. Comput. Graphics*, vol. 26, no. 5, pp. 1991–2001, 2020.

[56] N. J. Bryan, "Impulse response data augmentation and deep neural networks for blind room acoustic parameter estimation," in *in Proc. 2020 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '20)*, 2020.

[57] C. Kehling, "Evaluation of data augmentation techniques of room impulse responses for improved AI-based estimations," in *Proc. DAGA 2024*, (Hannover, Germany), pp. 1285–1288, 2024.

[58] S. Duangpummet *et al.*, "Blind estimation of speech transmission index and room acoustic parameters based on the extended model of room impulse response," *Applied Acoustics*, vol. 185, 2022. Article No. 108372.

[59] J. Lopez-Ballester *et al.*, "AI-IoT platform for blind estimation of room acoustic parameters based on deep neural networks," *IEEE Internet of Things Journal*, vol. 10, no. 1, pp. 855–866, 2022.

[60] C. Ick, A. Mehrabi, and W. Jin, "Blind acoustic room parameter estimation using phase features," in *Proc. 2023 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '23)*, (Rhodes, Greece), 2023.

[61] S. Deng, W. Mack, and E. A. P. Habets, "Online blind reverberation time estimation using CRNNs," in *Proc. INTERSPEECH 2020*, (Shanghai, China), pp. 5061–5065, 2020.

[62] P. Götz *et al.*, "Blind reverberation time estimation in dynamic acoustic conditions," in *Proc. 2022 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '22)*, (Singapore), pp. 581–585, 2022.

[63] P. Callens and M. Cernak, "Joint blind room acoustic characterization from speech and music signals using convolutional recurrent neural networks," *arXiv preprint arXiv:2010.11167*, 2020.

[64] P. S. López, P. Callens, and M. Cernak, "A universal deep room acoustics estimator," in *Proc. 2021 IEEE Workshop Appls. Signal Process. Audio Acoust. (WASPAA '21)*, (New Paltz, NY, USA), pp. 356–360, 2021.

[65] B. Yang and X. Li, "Self-supervised learning of spatial acoustic representation with cross-channel signal reconstruction and multi-channel conformer," *IEEE/ACM Trans. Audio Speech Language Process.*, vol. 32, pp. 4211–4225, 2024.

[66] C. Wang *et al.*, "Exploring the power of pure attention mechanisms in blind room parameter estimation," *EURASIP J. Audio Speech Music Process.*, vol. 2024, 2024. Article No. 23.

[67] P. P. Parada *et al.*, "A single-channel non-intrusive C50 estimator correlated with speech recognition performance," *IEEE/ACM Trans. Audio Speech Language Process.*, vol. 24, no. 4, pp. 719–732, 2016.

[68] M. Lavechin *et al.*, "Brouhaha: multi-task training for voice activity detection, speech-to-noise ratio, and C50 room acoustics estimation," in *Proc. 2023 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU '23)*, (Taipei, Taiwan), 2023.

[69] C. Wang *et al.*, "Attention is all you need for blind room volume estimation," in *Proc. 2024 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '24)*, (Seoul, Korea), pp. 1341–1345, 2024.

[70] M. Neri *et al.*, "Speaker distance estimation in enclosures from single-channel audio," *IEEE/ACM Trans. Audio Speech Language Process.*, vol. 32, pp. 2242–2254, 2024.

[71] C. Papayiannis, C. Evers, and P. A. Naylor, "End-to-end classification of reverberant rooms using DNNs," *IEEE/ACM Trans. Audio Speech Language Process.*, vol. 28, pp. 3010–3017, 2020.

[72] C. Papayiannis, C. Evers, and P. A. Naylor, "Data augmentation of room classifiers using generative adversarial networks," *arXiv preprint arXiv:1901.03257*, 2019.

[73] A. Vaswani *et al.*, "Attention is all you need," in *Proc. 31st Conf. Neural Inf. Process. Syst. (NIPS '17)*, (Long Beach, CA, USA), 2017.

[74] N. Antonello *et al.*, "Evaluation of a numerical method for identifying surface acoustic impedances in a reverberant room," in *Proc. 10th European Congress & Exposition Noise Control Eng. (EURONOISE '15)*, (Maastricht, The Netherlands), 2015.

[75] C. Foy, A. Deleforge, and D. D. Carlo, "Mean absorption estimation from room impulse responses using virtually supervised learning," *J. Acoust. Soc. Am.*, vol. 150, no. 2, pp. 1286–1299, 2021.

[76] W. Yu and W. B. Kleijn, "Room acoustical parameter estimation from room impulse responses using deep neural networks," *IEEE/ACM Trans. Audio Speech Language Process.*, vol. 29, pp. 436–447, 2020.

[77] C. Papayiannis, C. Evers, and P. A. Naylor, "Detecting sound-absorbing materials in a room from a single impulse response using a CRNN," *arXiv preprint arXiv:1901.05852*, 2019.

[78] G. Bologni, R. Heusdens, and J. Martinez, "Acoustic reflectors localization from stereo recordings using neural networks," in *Proc. 2021 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '21)*, (Toronto, ON, Canada), 2021.

[79] B. S. Liang *et al.*, "Reconstructing room scales with a single sound for augmented reality displays," *J. Inf. Display*, vol. 24, no. 1, pp. 1–12, 2023.

[80] N. Poschadel *et al.*, "Room geometry estimation from higher-order ambisonics signals using convolutional recurrent neural networks," in *Preprints AES 150th Conv.*, 2021.

[81] C. Tuna *et al.*, "Data-driven 3D room geometry inference with a linear loudspeaker array and a single microphone," in *Proc. Forum Acusticum 2023*, (Turin, Italy), 2023.

[82] H. N. Bicer *et al.*, "Data-driven joint detection and localization of acoustic reflectors," in *Proc. 2024 IEEE Int. Conf. Acoust., Speech, Signal Process. Workshops (ICASSPW '24)*, (Seoul, Korea), pp. 745–749, 2024.

[83] I. Yeon *et al.*, "EchoScan: scanning complex room geometries via acoustic echoes," *IEEE/ACM Trans. Audio Speech Language Process.*, vol. 32, pp. 4768–4782, 2024.

[84] D. Schindler, F. Schultz, and S. Spors, "Towards a data-driven plane wave decomposition from multichannel room impulse responses," in *Proc. DAGA 2023*, (Hamburg, Germany), pp. 1671–1674, 2023.

[85] P. A. Grumiaux *et al.*, "A survey of sound source localization with deep learning methods," *J. Acoust. Soc. Am.*, vol. 152, no. 1, pp. 107–151, 2022.

[86] D. Diaz-Guerra, A. Miguel, and J. R. Beltran, "Robust sound source tracking using SRP-PHAT and 3D convolutional neural networks," *IEEE/ACM Trans. Audio Speech Language Process.*, vol. 29, pp. 300–311, 2020.

[87] H. Li, W. Zhang, and L. Zhang, "DoA estimation of room reflections using NN-based MUSIC algorithm," in *in Proc. 2023 Asia Pacific Signal Inf. Process. Assoc. Annual Summit Conf. (APSIPA ASC '23)*, (Taipei, Taiwan), 2023.

[88] T. van Waterschoot *et al.*, "Double-talk-robust prediction error identification algorithms for acoustic echo cancellation," *IEEE Trans. Signal Process.*, vol. 55, no. 3, pp. 846–858, 2007.

[89] H. Zhang and D. Wang, "Deep learning for acoustic echo cancellation in noisy and double-talk scenarios," in *Proc. INTERSPEECH 2018*, (Hyderabad, India), pp. 3239–3243, 2018.

[90] C. Zhang and X. Zhang, "A robust and cascaded acoustic echo cancellation based on deep learning," in *Proc. INTERSPEECH 2020*, (Shanghai, China), pp. 5061–5065, 2020.

[91] N. L. Westhausen and B. T. Meyer, "Acoustic echo cancellation with the dual-signal transformation LSTM network," in *Proc. 2021 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '21)*, (Toronto, ON, Canada), 2021.

[92] A. Fazel, M. El-Khamy, and J. Lee, "Deep multitask acoustic echo cancellation," in *Proc. INTERSPEECH 2019*, (Graz, Austria), pp. 4250–4254, 2019.

[93] H. Zhang, M. Yu, and D. Yu, "Deep learning for joint acoustic echo and acoustic howling suppression in hybrid meetings," in *Proc. 2023 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU '23)*, (Taipei, Taiwan), 2023.

[94] A. Fazel, M. El-Khamy, and J. Lee, "CAD-AEC: context-aware deep acoustic echo cancellation," in *Proc. 2020 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '20)*, 2020.

[95] C. Zhang, J. Liu, and X. Zhang, "A complex spectral mapping with inplace convolution recurrent neural networks for acoustic echo cancellation," in *Proc. 2022 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '22)*, (Singapore), pp. 751–755, 2022.

[96] H. Zhang, K. Tan, and D. Wang, "Deep learning for joint acoustic echo and noise cancellation with nonlinear distortions," in *Proc. INTERSPEECH 2019*, (Graz, Austria), pp. 4255–4259, 2019.

[97] L. Cheng *et al.*, "Deep learning-based stereophonic acoustic echo suppression without decorrelation," *J. Acoust. Soc. Am.*, vol. 150, no. 2, pp. 816–829, 2021.

[98] H. Zhang and D. Wang, "A deep learning approach to multi-channel and multi-microphone acoustic echo cancellation," in *Proc. INTERSPEECH 2021*, (Brno, Czech Republic), pp. 1139–1143, 2021.

[99] H. Zhang and D. Wang, "Multi-channel and multi-microphone acoustic echo cancellation using a deep learning based approach," *arXiv preprint arXiv:2103.02552*, 2021.

[100] H. Zhao *et al.*, "A deep hierarchical fusion network for fullband acoustic echo cancellation," in *Proc. 2022 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '22)*, (Singapore), pp. 9112–9116, 2022.

[101] G. Li *et al.*, "Deep learning-based acoustic echo cancellation for surround sound systems," *Appl. Sci.*, vol. 13, no. 3, 2023. Article No. 1266.

[102] Y. Liu *et al.*, "A deep hybrid model for stereophonic acoustic echo control," *Circuits Syst. Signal Process.*, 2024.

[103] T. van Waterschoot and M. Moonen, "Fifty years of acoustic feedback control: state of the art and future challenges," *Proc. IEEE*, vol. 99, no. 2, pp. 288–327, 2011.

[104] C. Zheng *et al.*, "A deep learning solution to the marginal stability problems of acoustic feedback systems for hearing aids," *J. Acoust. Soc. Am.*, vol. 152, no. 6, pp. 3616–3634, 2022.

[105] H. Zhang, M. Yu, and D. Yu, "Deep AHS: A deep learning approach to acoustic howling suppression," in *Proc. 2023 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '23)*, (Rhodes, Greece), 2023.

[106] H. Zhang *et al.*, "Enhanced acoustic howling suppression via hybrid Kalman filter and deep learning models," *IEEE/ACM Trans. Audio Speech Language Process.*, vol. 32, pp. 2828–2840, 2024.

[107] P. A. Naylor and N. D. Gaubitch, eds., *Speech Dereverberation*. Springer, 2010.

[108] Y. Zhao, Z.-Q. Wang, and D. Wang, "Two-stage deep learning for noisy-reverberant speech enhancement," *IEEE/ACM Trans. Audio Speech Language Process.*, vol. 27, no. 1, pp. 53–62, 2018.

[109] W. Mack *et al.*, "Single-channel dereverberation using direct MMSE optimization and bidirectional LSTM networks," in *Proc. INTERSPEECH 2018*, (Hyderabad, India), pp. 1314–1318, 2018.

[110] L. Bahrman *et al.*, "Speech dereverberation constrained on room impulse response characteristics," in *Proc. INTERSPEECH 2024*, (Kos Island, Greece), pp. 622–626, 2024.

[111] X. Luo, Y. Ke, X. Li, and C. Zheng, "On phase recovery and preserving early reflections for deep-learning speech dereverberation," *J. Acoust. Soc. Am.*, vol. 155, no. 1, pp. 436–451, 2024.

[112] Y. Zhao *et al.*, "Monaural speech dereverberation using temporal convolutional networks with self attention," *IEEE/ACM Trans. Audio Speech Language Process.*, vol. 28, pp. 1598–1607, 2020.

[113] V. Kothapally and J. H. L. Hansen, "SkipConvGAN: Monaural speech dereverberation using generative adversarial networks via complex time-frequency masking," *IEEE/ACM Trans. Audio Speech Language Process.*, vol. 30, pp. 1600–1613, 2022.

[114] J. Donley and P. Calamia, "DARE-Net: Speech dereverberation and room impulse response estimation," tech. rep., Stanford University, Stanford, CA, USA, 2022.

[115] E. Vincent, T. Virtanen, and S. Gannot, eds., *Audio source separation and speech enhancement*. Wiley, 2018.

[116] M. Delfarah and D. Wang, "Deep learning for talker-dependent reverberant speaker separation: An empirical study," *IEEE/ACM Trans. Audio Speech Language Process.*, vol. 27, no. 11, pp. 1839–1848, 2019.

[117] E. W. Healy *et al.*, "A causal and talker-independent speaker separation/dereverberation deep learning algorithm: Cost associated with conversion to real-time capable operation," *J. Acoust. Soc. Am.*, vol. 150, no. 5, pp. 3976–3986, 2021.

[118] B. Wu *et al.*, "An end-to-end deep learning approach to simultaneous speech dereverberation and acoustic modeling for robust speech recognition," *IEEE J. Select. Topics Signal Process.*, vol. 11, no. 8, pp. 1289–1300, 2017.

[119] A. Zermini, *Deep learning for speech separation*. PhD thesis, University of Surrey, UK, 2020.

[120] A. Aroudi and S. Braun, "DBnet: DOA-driven beamforming network for end-to-end reverberant sound source separation," in *Proc. 2021 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '21)*, (Toronto, ON, Canada), 2021.

[121] F. Lluís *et al.*, "Direction specific ambisonics source separation with end-to-end deep learning," *Acta Acustica*, vol. 7, 2023. Article No. 29.

[122] A. Bohlender *et al.*, "Spatially selective speaker separation using a DNN with a location dependent feature extraction," *IEEE/ACM Trans. Audio Speech Language Process.*, vol. 32, pp. 930–945, 2024.

[123] C. Quan and X. Li, "SpatialNet: Extensively learning spatial information for multichannel joint speech separation, denoising and dereverberation," *IEEE/ACM Trans. Audio Speech Language Process.*, vol. 32, pp. 1310–1323, 2024.

[124] S. M. Kuo and D. R. Morgan, "Active noise control: a tutorial review," *Proc. IEEE*, vol. 87, no. 6, pp. 943–973, 1999.

[125] S. Im *et al.*, "Deep learning-assisted active noise control in a time-varying environment," *J. Mech. Sci. Technol.*, vol. 37, no. 3, pp. 1189–1196, 2023.

[126] H. Zhang and D. Wang, "Deep ANC: A deep learning approach to active noise control," *Neural Networks*, vol. 141, pp. 1–10, 2021.

[127] H. Zhang and D. Wang, "Deep MCANC: A deep learning approach to multi-channel active noise control," *Neural Networks*, vol. 158, 318-327, 2023.

[128] Y.-J. Cha, A. Mostafavi, and S. S. Benipal, "DNoiseNet: Deep learning-based feedback active noise control in various noisy environments," *Eng. Appl. Artificial Intell.*, vol. 121, Article No. 105971, 2023.

[129] J. Su, Z. Jin, and A. Finkelstein, "Acoustic matching by embedding impulse responses," in *Proc. 2020 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '20)*, 2020.

[130] C. Chen *et al.*, "Visual acoustic matching," in *Proc. 2022 IEEE/CVF Conf. Comput. Vision Pattern Recognition (CVPR '22)*, (New Orleans, LA, USA), pp. 18836–18846, 2022.

[131] N. D. Gaubitch, *Blind identification of acoustic systems and enhancement of reverberant speech*. PhD thesis, Imperial College London, UK, 2007.

[132] C. J. Steinmetz, V. K. Ithapu, and P. Calamia, "Filtered noise shaping for time domain room impulse response estimation from reverberant speech," in *Proc. 2021 IEEE Workshop Appls. Signal Process. Audio Acoust. (WASPAA '21)*, (New Paltz, NY, USA), pp. 221–225, 2021.

[133] A. Ratnarajah *et al.*, "AV-RIR: Audio-visual room impulse response estimation," in *Proc. 2024 IEEE/CVF Conf. Comput. Vision Pattern Recognition (CVPR '24)*, (Seattle, WA, USA), pp. 27164–27175, 2024.

[134] Z. Liao *et al.*, "Blind estimation of room impulse response from monaural reverberant speech with segmental generative neural network," in *Proc. INTERSPEECH 2023*, (Dublin, Ireland), pp. 2723–2727, 2023.

[135] A. Ratnarajah *et al.*, "Towards improved room impulse response estimation for speech recognition," in *Proc. 2023 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '23)*, (Rhodes, Greece), 2023.

[136] S. Cecchi, A. Carini, and S. Spors, "Room response equalization – a review," *Appl. Sci.*, vol. 8, no. 1, 2017. Article No. 16.

[137] C. Tuna *et al.*, "Data-driven local average room transfer function estimation for multi-point equalization," *J. Acoust. Soc. Am.*, vol. 152, no. 6, pp. 3635–3647, 2022.

[138] S. A. Verburg and E. Fernandez-Grande, "Reconstruction of the sound field in a room using compressive sensing," *J. Acoust. Soc. Am.*, vol. 143, no. 6, pp. 3770–3779, 2018.

[139] M. S. Kristoffersen *et al.*, "Deep sound field reconstruction in real rooms: introducing the isobel sound field dataset," *arXiv preprint arXiv:2102.06455*, 2021.

[140] M. Pezzoli *et al.*, "Deep prior approach for room impulse response reconstruction," *Sensors*, vol. 22, no. 7, 2022. Article No. 2710.

[141] X. Karakonstantis, E. Fernandez-Grande, and P. Gerstoft, "Efficient sound field reconstruction with conditional invertible neural networks," *arXiv preprint arXiv:2404.06928*, 2024.

[142] Z. Qiu *et al.*, "Transformer-based virtual microphone estimator," in *Proc. 2024 Workshop Hands-Free Speech Commun. Microphone Arrays (HSCMA '24)*, (Seoul, Korea), 2024.

[143] S. V. Lyster and C. Erkut, "A differentiable neural network approach to parameter estimation of reverberation," in *Proc. 19th Sound Music Comput. Conf. (SMC '22)*, (Saint-Étienne, France), pp. 358–364, 2022.

[144] D. A. Sanaguano-Moreno *et al.*, "Real-time impulse response: a methodology based on machine learning approaches for a rapid impulse response generation for real-time acoustic virtual reality systems," *Intell. Syst. Appl.*, vol. 21, 2024. Article No. 200306.

[145] C. Avendano and J. M. Jot, "A frequency-domain approach to multichannel upmix," *J. Audio Eng. Soc.*, vol. 52, no. 7/8, pp. 740–749, 2004.

[146] J. Xia and W. Zhang, "Upmix B-Format Ambisonic room impulse responses using a generative model," *Appl. Sci.*, vol. 13, no. 21, 2023. Article No. 11810.

[147] J. E. Gerken *et al.*, "Geometric deep learning and equivariant neural networks," *Artif. Intell. Rev.*, vol. 56, pp. 14605–14662, 2023.

[148] Z. Tang, H.-Y. Meng, and D. Manocha, "Learning acoustic scattering fields for dynamic interactive sound propagation," in *Proc. 2021 IEEE Virtual Reality 3D User Interfaces (VR '21)*, (Lisbon, Portugal), pp. 835–844, 2021.

[149] A. Ratnarajah *et al.*, "MESH2IR: Neural acoustic impulse response generator for complex 3D scenes," in *Proc. 30th ACM Int. Conf. Multimedia (MM '22)*, (Lisbon, Portugal), pp. 924–933, 2022.

[150] L. Kelley *et al.*, "RIR-in-a-Box: Estimating room acoustics from 3D mesh data through shoebox approximation," in *Proc. INTERSPEECH 2024*, (Kos Island, Greece), 2024.

[151] A. Ratnarajah *et al.*, "FAST-RIR: Fast neural diffuse room impulse response generator," in *Proc. 2022 IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '22)*, (Singapore), pp. 571–575, 2022.

[152] I. M. Salinas, J. A. B. Rodríguez, and G. P. Sipán, "Synthesis of room impulse responses by means of deep learning," in *in Proc. 53o Congreso Español de Acústica and XII Congreso Ibérico de Acústica (Tecniacústica '22)*, (Elche, Spain), 2022.

[153] I. Martin *et al.*, "Predicting room impulse responses through encoder-decoder convolutional neural networks," in *in Proc. 33rd IEEE Int. Workshop Machine Learning Signal Process. (MLSP '23)*, (Rome, Italy), 2023.

[154] S. Kim, J. h. Yoo, and J.-W. Choi, "Echo-aware room impulse response generation," *J. Acoust. Soc. Am.*, vol. 156, no. 1, pp. 623–637, 2024.

[155] J. Lin, G. Götz, and S. J. Schlecht, "Deep room impulse response completion," *arXiv preprint arXiv:2402.00859*, 2024.

[156] Y. He *et al.*, "Deep neural room acoustics primitive," in *Proc. 41st Int. Conf. Machine Learning (ICML '24)*, (Vienna, Austria), 2024.

[157] N. Borrel-Jensen *et al.*, "Sound propagation in realistic interactive 3D scenes with parameterized sources using deep neural operators," *Proc. Natl. Acad. Sci.*, vol. 121, no. 2, 2024. Article No. e2312159120.

[158] F. Ronchini *et al.*, "Room transfer function reconstruction using complex-valued neural networks and irregularly distributed microphones," in *Proc. 32nd European Signal Process. Conf. (EUSIPCO '24)*, (Lyon, France), pp. 441–445, 2024.

[159] Z. Liang, W. Zhang, and T. D. Abhayapala, "Sound field reconstruction using neural processes with dynamic kernels," *EURASIP J. Audio Speech Music Process.*, vol. 2024, 2024. Article No. 13.

[160] A. Luo *et al.*, "Learning neural acoustic fields," in *Adv. Neural Inf. Process. Syst. 35 (NeurIPS '22)*, pp. 3165–3177, 2022.

[161] K. Su, M. Chen, and E. Shlizerman, "INRAS: Implicit neural representation for audio scenes," in *Adv. Neural Inf. Process. Syst. 35 (NeurIPS '22)*, pp. 8144–8158, 2022.

[162] S. Liang *et al.*, "Neural acoustic context field: Rendering realistic room impulse response with neural fields," *arXiv preprint arXiv:2309.15977*, 2023.

[163] S. Majumder *et al.*, "Few-shot audio-visual learning of environment acoustics," in *Adv. Neural Inf. Process. Syst. 35 (NeurIPS '22)*, pp. 2522–2536, 2022.

[164] A. Brunetto, S. Hornauer, and F. Moutarde, "NeRAF: 3D scene infused neural radiance and acoustic fields," in *Proc. 13th Int. Conf. Learning Representations (ICLR '25)*, (Singapore), 2025, to appear.

[165] S. Liang *et al.*, "AV-NeRF: Learning neural fields for real-world audio-visual scene synthesis," in *Adv. Neural Inf. Process. Syst. 36 (NeurIPS '23)*, 2023.

[166] H. Gao *et al.*, "SOAF: Scene occlusion-aware neural acoustic field," *arXiv preprint arXiv:2407.02264*, 2024.

[167] C. Chen *et al.*, "Novel-view acoustic synthesis," in *Proc. 2023 IEEE/CVF Conf. Comput. Vision Pattern Recognition (CVPR '23)*, (Vancouver, BC, Canada), pp. 6409–6419, 2023.

[168] B. Ahn *et al.*, "Novel-view acoustic synthesis from 3D reconstructed rooms," in *Proc. INTERSPEECH 2024*, (Kos Island, Greece), pp. 3260–3264, 2024.

[169] L. Comanducci *et al.*, "Source localization using distributed microphones in reverberant environments based on deep learning and ray space transform," *IEEE/ACM Trans. Audio Speech Language Process.*, vol. 28, pp. 2238–2251, 2020.

[170] A. Berg *et al.*, "Extending GCC-PHAT using shift equivariant neural networks," in *Proc. INTERSPEECH 2022*, (Incheon, Korea), pp. 1791–1795, 2022.

[171] D. D.-G. Aparicio, *A geometric deep learning approach to sound source localization and tracking*. PhD thesis, Universidad de Zaragoza, Spain, 2023.

[172] D. Tang, M. Taseska, and T. van Waterschoot, "Toward learning robust contrastive embeddings for binaural sound source localization," *Front. Neuroinform.*, vol. 16, 2022. Article No. 942978.

[173] E. Grinstein *et al.*, "The Neural-SRP method for universal robust multi-source tracking," *IEEE Open J. Signal Process.*, vol. 5, pp. 19–28, 2023.

[174] N. Borrel-Jensen, A. P. Engsig-Karup, and C.-H. Jeong, "Machine learning-based room acoustics using flow maps and physics-informed neural networks," *J. Acoust. Soc. Am.*, vol. 151, no. 4, pp. A232–A233, 2022.

[175] K. Niebler *et al.*, "Towards reconstruction of acoustic fields via physics-informed neural networks," in *Proc. 51st Int. Congress & Exposition Noise Control Eng. (INTER-NOISE '22)*, vol. 265, (Glasgow, Scotland, UK), 2022.

[176] X. Karakonstantis *et al.*, "Room impulse response reconstruction with physics-informed deep learning," *J. Acoust. Soc. Am.*, vol. 155, no. 2, pp. 1048–1059, 2024.

[177] M. Pezzoli, F. Antonacci, and A. Sarti, "Implicit neural representation with physics-informed neural networks for the reconstruction of the early part of room impulse responses," in *Proc. Forum Acusticum 2023*, (Turin, Italy), 2023.

[178] I. Tsunokini *et al.*, "Spatial extrapolation of early room impulse responses with noise-robust physics-informed neural network," *IEICE Trans. Fundam. Electron. Commun. Comput. Sci., Article No. 2024EAL2015*, 2024.

[179] M. Olivieri *et al.*, "Physics-informed neural network for volumetric sound field reconstruction of speech signals," *EURASIP J. Audio Speech Music Process.*, vol. 2024, 2024. Article No. 42.

[180] F. Ma, S. Zhao, and I. S. Burnett, "Sound field reconstruction using a compact acoustics-informed neural network," *J. Acoust. Soc. Am.*, vol. 156, no. 3, pp. 2009–2021, 2024.

[181] D. Sundström, S. Koyama, and A. Jakobsson, "Sound field estimation using deep kernel learning regularized by the wave equation," in *Proc. 2024 Int. Workshop Acoustic Signal Enhancement (IWAENC '24)*, (Aalborg, Denmark), pp. 319–323, 2024.

[182] X. Karakonstantis and E. Fernandez-Grande, "Generative adversarial networks with physical sound field priors," *J. Acoust. Soc. Am.*, vol. 154, no. 2, pp. 1226–1238, 2023.

[183] E. Fernandez-Grande *et al.*, "Generative models for sound field reconstruction," *J. Acoust. Soc. Am.*, vol. 153, no. 2, pp. 1179–1190, 2023.

[184] I. Szoke *et al.*, "Building and evaluation of a real room impulse response dataset," *IEEE J. Select. Topics Signal Process.*, vol. 13, no. 4, pp. 863–876, 2019.

[185] J. Čmejla *et al.*, "MIRaGe: Multichannel database of room impulse responses measured on high-resolution cube-shaped grid," in *Proc. 28th European Signal Process. Conf. (EUSIPCO '20)*, (Amsterdam, The Netherlands), pp. 56–60, 2020.

[186] D. D. Carlo *et al.*, "dEchorate: a calibrated room impulse response dataset for echo-aware signal processing," *EURASIP J. Audio Speech Music Process.*, vol. 2021, 2021. Article No. 39.

[187] S. Koyama *et al.*, "MeshRIR: A dataset of room impulse responses on meshed grid points for evaluating sound field analysis and synthesis methods," in *Proc. 2021 IEEE Workshop Appls. Signal Process. Audio Acoust. (WASPAA '21)*, (New Paltz, NY, USA), pp. 1–5, 2021.

[188] S. Zhao *et al.*, "A room impulse response database for multizone sound field reproduction," *J. Acoust. Soc. Am.*, vol. 152, no. 4, pp. 2505–2512, 2022.

[189] T. Dietzen *et al.*, "MYRiAD: a multi-array room acoustic database," *EURASIP J. Audio Speech Music Process.*, vol. 2023, 2023. Article No. 17.

[190] D. Fejgin, W. Middelberg, and S. Doclo, "BRUDEX database: Binaural room impulse responses with uniformly distributed external microphones," in *Proc. 15th ITG Conf. Speech Commun.*, (Aachen, Germany), pp. 126–130, 2023.

[191] Z. Chen *et al.*, "Real acoustic fields: An audio-visual room acoustics dataset and benchmark," in *Proc. 2024 IEEE/CVF Conf. Comput. Vision Pattern Recognition (CVPR '24)*, (Seattle, WA, USA), pp. 21886–21896, 2024.

[192] G. Chesworth, A. Bastine, and T. Abhayapala, "Room impulse response dataset of a recording studio with variable wall paneling measured using a 32-channel spherical microphone array and a B-Format microphone array," *Appl. Sci.*, vol. 14, no. 5, 2024. Article No. 2095.

[193] B. Yang *et al.*, "RealMAN: A real-recorded and annotated microphone array dataset for dynamic speech enhancement and localization," in *Adv. Neural Inf. Process. Syst. 37 (NeurIPS '24)*, pp. 105997–106019, 2024.

[194] A. Kujawski, A. J. Pelling, and E. Sarradj, "MIRACLE – a microphone array impulse response dataset for acoustic learning," *EURASIP J. Audio Speech Music Process.*, vol. 2024, 2024. Article No. 32.

[195] F. Miotello *et al.*, "HOMULA-RIR: A room impulse response dataset for teleconferencing and spatial audio applications acquired through higher-order microphones and uniform linear microphone arrays," in *Proc. 2024 IEEE Int. Conf. Acoust., Speech, Signal Process. Workshops (ICASSPW '24)*, (Seoul, Korea), pp. 795–799, 2024.

[196] S. Fragner *et al.*, "Dataset of directional room impulse responses for realistic speech data," *Data in Brief*, vol. 53, 2024. Article No. 110229.

[197] G. Stolz *et al.*, "Spatial room impulse response dataset: a robot's journey through coupled rooms of a reverberant university building," in *Proc. DAGA 2024*, (Hannover, Germany), pp. 245–247, 2024.

[198] S. Damiano, K. MacWilliam, *et al.*, "The tRIRjectory database: room acoustic recordings along a trajectory of moving microphones," tech. rep., KU Leuven, Leuven, Belgium, 2025.

[199] S. Damiano and T. van Waterschoot, "Sound field reconstruction using physics-informed boundary integral networks," in *Proc. 33rd European Signal Process. Conf. (EUSIPCO '25)*, (Palermo, Sicily, Italy), 2025, submitted for publication.