DEEP LEARNING-BASED PREDICTION OF ENERGY DECAY CURVES FROM ROOM GEOMETRY AND MATERIAL PROPERTIES

Imran Muhammad, Gerald Schuller

Technische Universität Ilmenau, Germany

ABSTRACT

Accurate prediction of energy decay curves (EDCs) enables robust analysis of room acoustics and reliable estimation of key parameters. We present a deep learning framework that predicts EDCs directly from room geometry and surface absorption. A dataset of 6000 shoebox rooms with realistic dimensions, source–receiver placements, and frequency-dependent wall absorptions was synthesized. For each configuration we simulate room impulse responses (RIRs) using Pyroomacoustics and compute target EDCs. Normalized room features are provided to a long short-term memory (LSTM) network that maps configuration to EDC. Performance is evaluated with mean absolute error (MAE) and root mean square error (RMSE) over time. We further derive early decay time (EDT), reverberation time (T20), and clarity index (C50) from predicted and target EDCs; close agreement is observed (e.g., EDT MAE 0.017 s, T20 MAE 0.021 s). The approach generalizes across diverse rooms and supports efficient room-acoustics modeling for early-stage design and real-time applications.

Index Terms— Room Acoustics, Neural Network, Energy Decay Curves, LSTM

1. INTRODUCTION

Room acoustics plays a central role in shaping the auditory experience in enclosed environments such as concert halls, classrooms, offices, and virtual spaces. Key room acoustic parameters, such as reverberation times (T20), early decay time (EDT) and clarity (C50) provide quantitative insight into how sound behaves in a space and are critical for evaluating acoustic quality of the closed space. These parameters are often derived from room impulse responses (RIRs) or energy decay curves (EDCs), both of which describe how sound energy evolves over time after an acoustic excitation [1]. Traditionally, accurate estimation of such acoustic descriptors requires either detailed physical modeling (e.g., Ray Tracing, Image Source Method (ISM), and wave-based simulations) [2] or time-consuming on-site measurements. Simplified empirical approaches, such as Sabine's [3] or Eyring's [4] formulas, are also used, though they rely on idealized assumptions and may be inaccurate in complex rooms. Fast geometrical acoustics software such as pysound [5], ITA Geometrical Acoustics [6], RAVEN (Room Acoustics Simulation for Virtual Environments) [7], openRay [8], EVERTims [9], GSoundSIR [10] and Pyroomacoustics [11] provide an efficient alternative to physical measurements. However, these methods are based on approximations and struggle to capture wave-based phenomena accurately, particularly at low frequencies. Wave-based solvers offer greater accuracy but are often computationally intensive and unsuitable for real-time applications. This trade-off between speed and accuracy presents an opportunity for deep learning or data-driven approaches. Crucially, such models should be trained on high-quality data, ideally derived from physical measurements or wave-based simulations, to ensure reliability. Recently, data-driven approaches utilizing deep learning have shown significant promise in modeling complex acoustic environments by learning representations directly from structured input data [12]. Deep neural networks have been increasingly employed in room acoustics research, including the prediction of key parameters such as reverberation time and clarity indices [13, 14, 15, 16], the estimation of room impulse responses from incomplete or noisy measurements [17], and the modeling of energy decay functions through specialized neural network architectures. These advancements highlight the effectiveness of deep learning in capturing the intricate temporal and spatial characteristics of room acoustics, providing efficient and scalable alternatives to conventional physics-based modeling techniques.

In this study, we introduce a new approach that predicts EDCs from room features as an intermediate step, instead of directly predicting full room impulse responses. EDCs provide a more compact and stable representation of a room's temporal energy decay and reverberation characteristics [1], which makes them better suited for learning-based methods. In contrast, RIRs are high-dimensional signals with complex temporal and spectral details, making direct prediction more challenging and error-prone. Since EDCs also serve as the basis for computing important acoustic parameters [2], this strategy simplifies the modeling task while preserving the essential information required for an accurate room acoustics analysis. The synthesis of RIRs from predicted EDCs is a promising direction for our future studies, but it is not within the scope of this study.

We propose a deep learning-based framework that uses a Long Short-Term Memory (LSTM) neural network to predict EDCs directly from the geometric features of the room and the material absorption characteristics. LSTM networks are particularly well suited for time series prediction tasks due to their ability to capture long-term dependencies and temporal patterns, overcoming the limitations of standard recurrent neural networks (RNNs) such as the problem of vanishing gradients [18, 19]. Compared to convolutional neural networks (CNNs), which are more effective in spatial pattern recognition, LSTMs provide improved performance on sequential data, which is crucial for modeling the temporal decay nature of EDCs [20]. Although alternative architectures such as Transformers offer competitive performance in sequence modeling, their higher computational complexity and data requirements make LSTMs a more efficient choice for this application [21, 22]. This makes LSTM a robust and computationally feasible solution for EDCs predictions from room configuration parameters.

Our key contributions are as follows. We present a deep learning model capable of predicting EDCs directly from realistic room geometries and room material absorption characteristics, reducing the need for full physical-based simulations or impulse response measurements. Second, we introduce a framework in which the key acoustic parameters derived from the predicted EDCs are validated against those obtained from the simulated data, thereby demonstrating the precision and practical applicability of the proposed method. This proposal offers a foundation for rapid and efficient estimation of room acoustic behavior and has potential applications in early-stage acoustic design, real-time auralization.

2. METHODOLOGY

We developed an LSTM-based deep learning framework to predict the EDCs of shoebox rooms using its geometric parameters and absorption properties as input features. The model is trained on a dataset generated through realistic combinations of room dimensions, source-receiver placements, and surface material absorptions. These features are normalized and structured before being fed into the LSTM network, which is designed to learn the mapping of room features to the temporal profile of the EDCs. The training process uses the mean squared error (MSE) as the loss function.

2.1. Dataset and Room Simulation Parameters

A data-set comprising 6000 distinct room configurations was used as input room features, with each room defined by its length (L), width (W), height (H), source and receiver positions in three dimensions (i.e. X, Y, and Z), and frequency-dependent absorption coefficients averaged for all walls. These parameters were varied over a range of realistic values (see Table 1) to simulate various acoustic environments. For input data generation, we used the Pyroomacoustics library [23] to simulate a diverse set of shoebox-shaped rooms. Pyroomacoustics applies geometrical acoustics techniques, namely, the Image Source Method (ISM) and ray tracing, suitable for modeling high-frequency sound propagation. However, these methods do not account for wave-based phenomena. The dimensions of the rooms and the positions of the source and receiver were systematically varied and the surface absorption properties were sampled from a measured material database [11]. These absorption coefficients are used to approximate the surface behavior, though they primarily account for energy loss and not phase-related effects. RIRs and the corresponding EDCs are synthesized using a hybrid simulation approach (ISM and ray tracing) within the Pyroomacoustics framework and saved as time series data at sampling rate of $44.1\,\mathrm{kHz}$. While efficient, these methods are based on geometrical acoustics and are primarily valid at high frequencies, where wave-based effects are less significant.

Table 1. Range of simulated room acoustic features used as input parameters for LSTM model training

Parameter	**Range / Values**
Dimensions ($L \times W \times H$)	$3-6\,\mathrm{m} \times 3-6\,\mathrm{m} \times 2.5-4\,\mathrm{m}$
Source-Receiver Distances	$1-4\,\mathrm{m}$
Wall Absorptions Range	$0.14-0.65$ (125 - 8000 Hz)
Total Room Configurations	6000

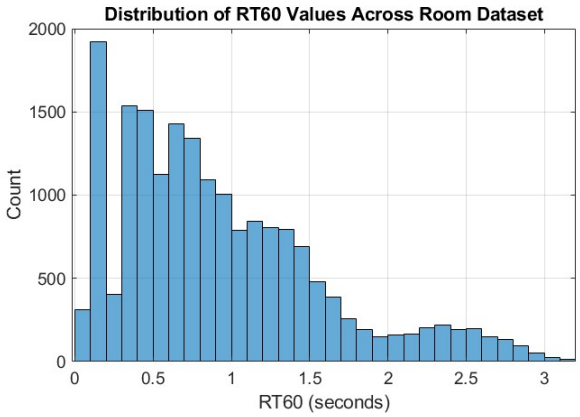

Fig. 1. Histogram showing the distribution of $T60$ values across the room dataset

To capture a wide range of acoustic environments, the reverberation time ($T60$) was computed from the RIR of each simulated room configuration. This ensures the dataset spans from nearly anechoic to highly reverberant conditions, covering both short and long $T60$ values. Figure 1 shows the histogram of $T60$ across all configurations, highlighting the variability and coverage of reverberation conditions. Such diversity is crucial for training and evaluating models that must generalize across different room types.

2.1.1. *Input Data Preprocessing and Normalization*

The input features, representing room parameters, were scaled using *MinMax()* normalization to the $[0, 1]$ range. Though, the normalized energy decay curves dataset inherently fall within this interval, *MinMax* normalization was still applied as a standard preprocessing step to maintain consistency with best practices in deep learning and to promote stable training dynamics. Following normalization, the data were reshaped into a three-dimensional format of $[N, 1, 16]$, where N denotes the number of rooms and 16 corresponds to the number of features per room. While the input features are not temporally sequential, the Long Short-Term Memory (LSTM) architecture requires a three-dimensional input of the form (batch_size, sequence_length, feature_dimension). To satisfy this requirement, each room sample was represented as a sequence of length one containing 16 feature values. This transformation enables compatibility with the LSTM structure while leveraging its gating mechanisms to model potentially complex inter-dependencies among the input features. Although a two-dimensional input of shape $[N, 16]$ would be sufficient for a fully connected neural network, the LSTM architecture mandates the inclusion of a sequence dimension, even if it represents only a single step. For model evaluation, the dataset was partitioned into 60% training, 20% validation, and 20% test subsets.

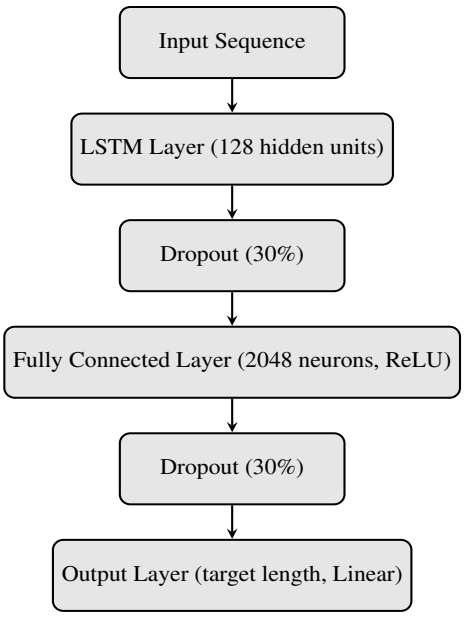

Fig. 2. Architecture of the LSTM model used for EDC prediction.

2.2. LSTM Model Architecture

As Figure 2 shows, the LSTM model was implemented in PyTorch with a single LSTM layer of 128 hidden units, followed by a 30 % dropout. A fully connected layer with 2048 ReLU-activated neurons and another dropout layer precede the output layer, which has a size equal to the target EDC sequence and uses a linear activation for regression. The model was trained with the Adam optimizer ($LR = 0.001$) using mean squared error loss, with early stopping after 10 epochs of no validation improvement. We compute MSE loss for EDCs on the linear scale instead of the dB scale. The dB transform exaggerates small variations in the late decay and downplays early decay differences, biasing the model. Linear loss better reflects true energy differences and yields a more stable, physically meaningful objective.

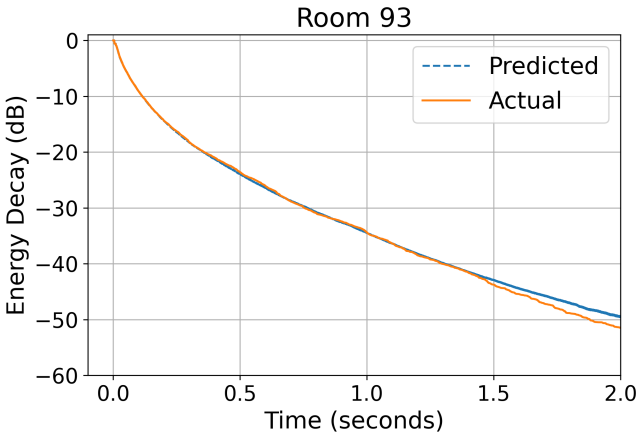

Fig. 3. Target and predicted EDCs for randomly selected room

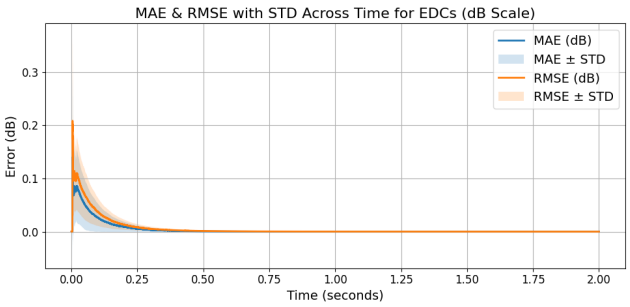

Fig. 4. **MAE** and **RMSE** of the predicted EDCs, averaged over time across all room configurations

3. RESULTS AND DISCUSSION

After training, the model shows strong predictive accuracy on unseen rooms. Predicted EDCs are evaluated and used to derive room acoustic parameters (mentioned in Section 1). These are compared with those from target EDCs to assess reliability.

Table 2. Performance of LSTM Model: **MAE**, **RMSE**, and coefficient of determination (R^2) computed between predicted and target EDCs

Metric	MAE	RMSE	R^2
EDT (s)	0.017	0.023	0.987
T20 (s)	0.021	0.029	0.978
C50 (dB)	0.917	2.023	0.888

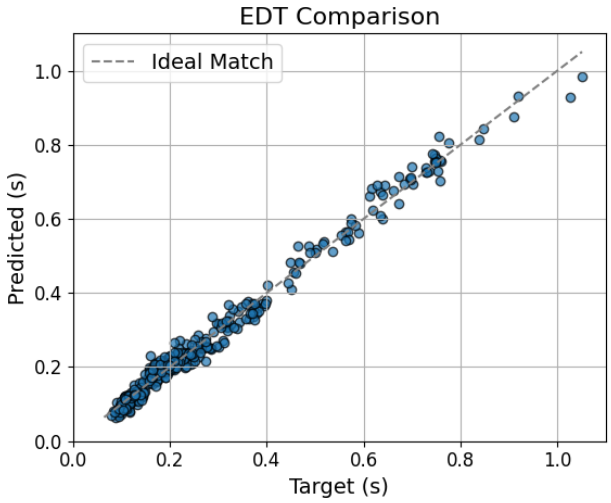

Fig. 5. Comparison of predicted and target EDT values

3.1. Prediction of Energy Decay Curves

To assess the model's predictive performance, qualitative comparisons were conducted between the target (i.e. simulated) and predicted EDCs for a randomly selected room as shown in Figure 3. The graph reveal a strong agreement between the predicted and target EDC curves, suggesting that the model effectively captures the temporal decay characteristics of rooms. The predicted EDCs not only replicate the overall decay envelope but also capture fine-grained fluctuations in the decay profile. To further evaluate the accuracy of the predicted EDCs, we reported the MAE and RMSE computed across all predicted EDCs, averaged over time. Additionally, we included the standard deviation to reflect variability across the dataset as shown in Figure 4 to provide a more comprehensive evaluation of model behavior.

3.2. Acoustic Parameter Estimation Accuracy

Additional analyses such as key room acoustic parameters (as discussed in Section 1) were derived from both the target (simulated) and predicted EDCs. The predictive performance was quantified using MAE, RMSE, and R^2 (coefficient of determination) for each parameter across all room configurations in the test dataset. Table 2 summarizes the error metrics.

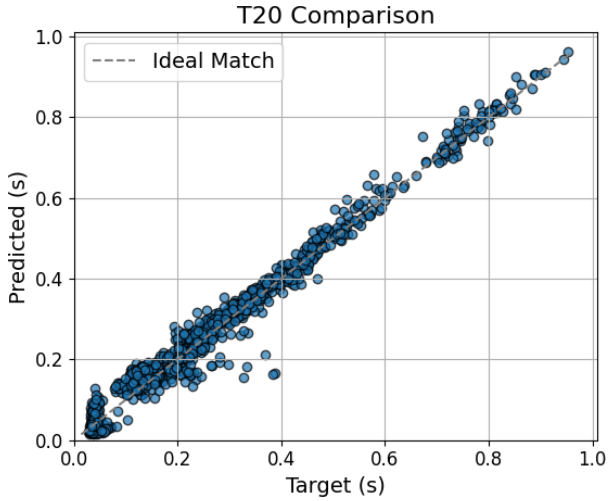

Fig. 6. Comparison of predicted and target T20 values

3.2.1. EDT and Reverberation Times (EDT, T20)

Figures 5 and 6 compare the predicted vs. target values for each parameter. Across these graphs, a close alignment along the identity line ($y = x$) indicates strong agreement with the model. Particularly, the clustering around the ideal match line was the most prominent suggests that the model captured early reverberant characteristics effectively. The model shows a good performance in estimating EDT and T20, with a low MAE of $0.017\,\mathrm{s}$ and an RMSE of $0.023\,\mathrm{s}$ and MAE of $0.021\,\mathrm{s}$ and an RMSE of $0.029\,\mathrm{s}$ respectively. The high R^2 value of 0.987 and 0.978 for EDT and T20 respectively indicates that the predictions closely follow the trend of target values, confirming the model's ability to learn the early decay slope of the EDCs.

4. SUMMARY AND OUTLOOK

The LSTM-based model demonstrated good performance in predicting EDCs and estimating key acoustic parameters with good accuracy. The model achieved low error values and consistently high R^2 scores. Our future work is on extending the model's generalization to real-world measured data and synthesis of full RIRs conditioned on these EDCs using methods such as stochastic modeling, inverse filtering, or generative neural approaches. We believe this approach enables accurate and plausible RIR synthesis for applications such as real-time room acoustics applications, such as intelligent room tuning, auralization engines, or speech enhancement systems, where fast and accurate acoustic parameter estimation is critical. The source code, preprocessed dataset, and trained model used in this study are available at [24].

5. REFERENCES

[1] Heinrich Kuttruff, *Room Acoustics*, CRC Press, 5 edition, 2009.

[2] Michael Vorländer, *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*, Springer, Cham, 2 edition, 2020.

[3] G. Millington, "A modified formula for reverberation," *The Journal of the Acoustical Society of America*, vol. 4, pp. 69–82, 1932.

[4] W. H. Sette, "A new reverberation time formula," *The Journal of the Acoustical Society of America*, vol. 4, pp. 193–210, 1933.

[5] Carl Schissler and Dinesh Manocha, "Gsound: Interactive sound propagation for games," in *Audio Engineering Society Conference: 41st International Conference: Audio for Games*. Audio Engineering Society, 2011.

[6] Virtual Acoustics, "Itageometricalacoustics," 2025.

[7] Dirk Schröder and Michael Vorländer, "Raven: A real-time framework for the auralization of interactive virtual environments," in *Forum acusticum*. Aalborg Denmark, 2011, pp. 1541–1546.

[8] EmergentMind, "openray," 2025.

[9] S. Laine, S. Siltanen, T. Lokki, and L. Savioja, "Accelerated beam tracing algorithm," *Applied Acoustics*, vol. 70, no. 1, pp. 172–181, 2009.

[10] Yongyi Zang and Qiuqiang Kong, "Gsound-sir: A spatial impulse response ray-tracing and high-order ambisonic auralization python toolkit," 2025.

[11] Robin Scheibler, Eric Bezzam, and Ivan Dokmanić, "Pyroomacoustics: A python package for audio room simulation and array processing algorithms," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 351–355.

[12] Wangyang Yu and W. Bastiaan Kleijn, "Room geometry estimation from room impulse responses using convolutional neural networks," 2019.

[13] Chen Meng, Noam Shabtai, and Boaz Rafaely, "Predicting room acoustic parameters from room geometry using deep learning," *The Journal of the Acoustical Society of America*, vol. 154, no. 4, pp. 2452–2461, 2023.

[14] Karolina Prawda, Sebastian J. Schlecht, and Vesa Välimäki, "Calibrating the sabine and eyring formulas," *The Journal of the Acoustical Society of America*, vol. 152, no. 2, pp. 1158–1169, 08 2022.

[15] Georg Götz, Ricardo Falcón Pérez, Sebastian J. Schlecht, and Ville Pulkki, "Neural network for multi-exponential sound energy decay analysis," *The Journal of the Acoustical Society of America*, vol. 152, no. 2, pp. 942–953, 08 2022.

[16] Wangyang Yu and W. Bastiaan Kleijn, "Room acoustical parameter estimation from room impulse responses using deep neural networks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 436–447, 2021.

[17] Jihoon Kim and Youngmoo E Yang, "Generative adversarial neural network for room impulse response synthesis," *arXiv preprint arXiv:2311.02581*, 2023.

[18] Sepp Hochreiter and Jürgen Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[19] Yoshua Bengio, Patrice Simard, and Paolo Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE transactions on neural networks*, vol. 5, no. 2, pp. 157–166, 1994.

[20] Fazle Karim, Somshubra Majumdar, Hamid Darabi, and Shun Chen, "Lstm fully convolutional networks for time series classification," *IEEE Access*, vol. 6, pp. 1662–1669, 2018.

[21] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008.

[22] Shaojie Bai, J Zico Kolter, and Vladlen Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *arXiv preprint arXiv:1803.01271*, 2018.

[23] Robin Scheibler, Elie Bezzam, and Ivan Dokmanić, "Pyroomacoustics: A python package for audio room simulation and array processing algorithms," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 351–355.

[24] Gerald Schuller and Imran Muhammad, "Lstm-model-energy-decay-curves: Github repository (2025) link:," `https://github.com/TUIlmenauAMS/LSTM-Model-Energy-Decay-Curves`.