Audio Engineering Society

Convention Paper 10218

Presented at the AES 158th Convention
2025 May 22–24, Warsaw, Poland

An Artificial Reverberator Informed by Room Geometry and Visual Appearance

Ruari Molyneux[1] and Joshua D Reiss[1,2]

[1]*Queen Mary University*
[2]*Centre for Digital Music*

Correspondence should be addressed to Ruari Molyneux (ruari.molyneux@gmail.com)

ABSTRACT

Without relying on audio data as a reference, artificial reverberation models often struggle to accurately simulate the acoustics of real rooms. To address this, we propose a hybrid reverberator derived from a room's physical properties. Room geometry is extracted via Light Detection and Ranging mapping, enabling the calculation of acoustic reflection paths via the Image Source Method. Frequency-dependent absorption is found by classifying room surface materials with a multi-modal Large Language Model and referencing a database of absorption coefficients. The extracted information is used to parametrise a hybrid reverberator, divided into two components: early reflections, using a tapped delay line, and late reverberation, using a Scattering Feedback Delay Network. Our listening test results show that participants often rate the proposed system as the most natural simulation of a small hallway room. Additionally, we compare the reverberation metrics of the hybrid reverberator and similar state-of-the-art models to those of the small hallway.

1 Introduction

Traditional methods for recording Room Impulse Response (RIR) involve significant overhead, including the use of multiple microphones, heavy speakers, and the need for a completely silent environment. This makes the process resource-intensive and challenging. Without these prerequisites, the RIR becomes unusable even after extensive post-processing.

To overcome these difficulties, we propose a system that artificially models a room's reverberation using only its geometric and visual properties. We envision a portable approach to extracting these properties, emphasising consumer-grade hardware like phone cameras and LiDAR scanners. We break the task down into the following steps, also shown in Fig. 1.

- Extracting the desired room geometry via LiDAR scanning,
- Identifying the primary material of each surface from images of the room,
- Finding reflection paths based on room geometry using the Image Source Method,
- Retrieving absorption coefficients from a dataset based on each wall's primary material,
- Defining a hybrid reverberator based on the reflection paths and absorption characteristics.

The aim is to model the reverberation of a real room using scalable techniques that prioritise high sound

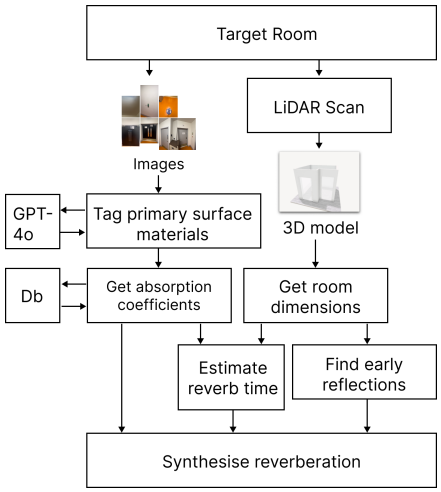

Fig. 1: Geometric and visual reverb system overview.

quality for audio production rather than exact acoustic replication. This approach emphasises creating an idealised room response that accurately conveys the space's perceived size and absorption characteristics. The idealised response should deliver an aesthetically pleasing version of a room rather than replicating the natural imperfections.

2 Related Work

2.1 Artificial Reverberation

The Image Source Method (ISM) [1] is a geometric acoustic technique that simulates room acoustics with high accuracy. The method generates virtual sources by reflecting the original sound source across the room's boundaries. The source is reflected for each axis, and the resulting virtual image sources are then reflected again, creating higher-order reflections. This process is repeated until the desired reflection order is achieved.

Introduced by Gerzon [2] [3] and implemented by Puckette [4] the Feedback Delay Network (FDN) is a recursive reverberation structure containing a set of delays lines, a mixing matrix that controls the distribution between delay lines, and absorption filters which set the reverb decay time. In [5], short velvet noise Finite Impulse Response (FIR) filters are added to each mixing matrix element to model scattering on material surfaces, increasing reverb density.

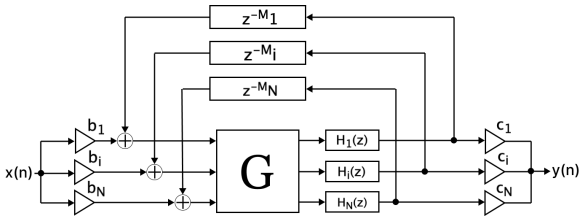

Fig. 2: Transposed FDN, where the output is taken from the mixing matrix. z^{-M_i} is the feedback delay time, G is the mixing matrix, $H_i(z)$ is the absorption filter of the ith delay line and b, c are the input and output gains, respectively.

The Scattering Delay Network (SDN) [6] combines the efficiency of the FDN with the accuracy of the Digital Waveguide Mesh [7] by using a sparse network of waveguides, with scattering nodes placed at first-order reflection points identified by the ISM. This ensures that first-order reflections are rendered accurately while the recursive structure approximates late reflections. Frequency-dependent surface absorption is modelled using filters at each scattering node, resulting in an energy decay curve similar to the ISM.

2.2 Reverb Matching Systems

Extensive research has been done on deriving artificial reverberators based on RIRs [8, 9, 10]. These systems optimise parametric reverberators by matching features extracted from RIRs. This is demonstrated in [9], where the authors tune the FDNs gains and delay line lengths using a genetic algorithm with a Mel-Frequency Cepstral Coefficients cost function.

Deep learning methods have advanced reverb estimation by integrating visual and acoustic data [11, 12, 13]. In [11] the authors propose visual acoustic matching via cross-modal transformers to apply the acoustics of a target environment depicted in an image to audio. Furthermore, [12] combines reverberant speech samples with panoramic images of an environment to estimate RIRs. The system employs a multi-task learning framework where geometry and material properties, encoded as "Geo-Mat" features, are extracted from visual cues.

Additionally, systems that focus on extracting acoustic features without the use of audio are emerging [14] [15]. For instance, [14] employs LiDAR-based geometry reduction techniques alongside visual material classification to parametrise the ISM and generate a RIR.

3 Method

3.1 Geometry Extraction

We start the geometry extraction procedure by using Apple RoomPlan API and LiDAR-capable iPhone 14 Pro to scan the room. The result is a 3D model encapsulating the room measurements. Since our early reflection device uses the standard ISM, we extract a cuboid from the room dimension by taking the maximum length, height, and width. After the geometry extraction, we get the room dimensions of length 2.97 meters (m), width 2.61 m, and height 2.68 m. With the extracted geometry, we use the ISM to find the reflection paths in the room up to the second order.

3.2 Material Recognition

To simulate the appropriate frequency-dependent decay time of the room, we use images of each surface and find the primary materials present. We use the GPT-4o multi-modal Large Language Model (LLM) [16] to classify each image. We provide the model with a dataset, sourced from the appendix of [17], which contains a list of material names and their corresponding absorption coefficients. Then, we prompt the model to identify the primary material in the image based on the material names in the dataset. The predicted name of the primary material is then used to query the material database. We use a temperature of 0.01 to achieve a more deterministic response. Temperature is a hyperparameter that controls the randomness of the LLM output.

We test the method in a small cuboid room with minimal furnishing and simple materials, shown in Fig. 6. The matching process results are shown in Fig. 3 . Ideally, a classifier trained with surface image and absorption coefficient pairs would be used to improve performance on complex scenes.

3.3 Hybrid Reverberator

The room simulation is broken down into early and late reflection stages, using the ISM and Scattering FDN, respectively. This distinction ensures that the initial energy of the reverb matches the listener's expectations for the space. While the later reflections provide a maximally diffuse and smooth response, conveying the size and materials of the room. We use a parallel structure, as shown in Fig. 4, where early reflections and late

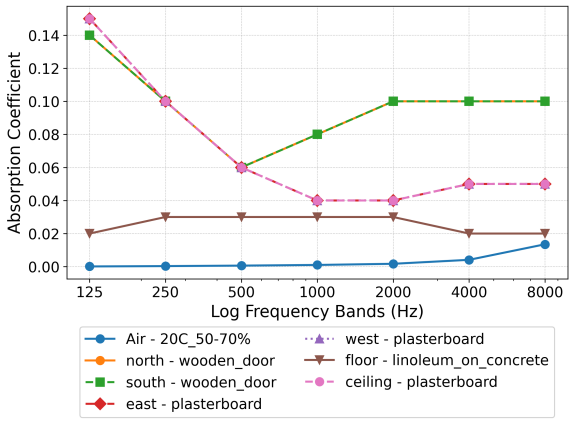

Fig. 3: Estimated surface and air absorption coefficients from material recognition process applied to the small hallway images.

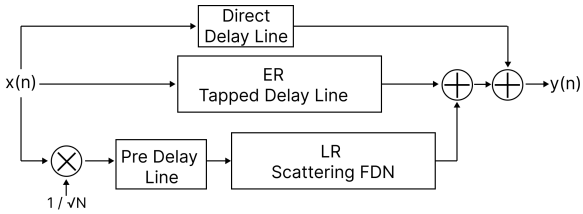

Fig. 4: Block diagram for the hybrid reverberator.

reverberation are processed separately and combined. This decision is due to a comb filtering effect when the FDN processes the output from the early reflections stage. Therefore, to ensure both stages are aligned in the time domain, the FDN input is delayed until the first reflection. Additionally, we take the output of the late reverberation from the FDN mixing matrix to create immediate density at the first reflection. The structure used to achieve this is shown in Fig. 2

The delay times for the early and late stages are determined by calculating reflection paths using the ISM. Reflections are computed up to the second order. For each reflection path, we use (1) to find the Euclidean distance and convert it to a sample delay between each image source and the listener. Where $\mathbf{x}$ is a coordinate vector, S_k denotes the image source, R is the receiver, F_s is the sampling rate, and c is the speed of sound in air.

$$D_{S_k,R} = \left\lfloor F_s \frac{\|\mathbf{x}_{S_k} - \mathbf{x}_R\|}{c} \right\rfloor \tag{1}$$

3.3.1 Early Reflections

To implement the delay times calculated with the ISM we employ a tapped delay line (TDL) where each tap represents a different reflection. Although Finite Impulse Response (FIR) filters can also be used to model the reflection paths, doing so would result in a system with an order equal to the largest delay time. In contrast, using a TDL results in a system order corresponding to the number of reflections, thereby reducing computational complexity.

$$g_{S_k,R} = \frac{1}{\|\mathbf{x}_{S_k} - \mathbf{x}_R\|} \tag{2}$$

Each tap in the TDL has an associated attenuation gain $g_{S_k,R}$ that follows the inverse distance law ($1/r$) to simulate spherical spreading. Additionally, we design a minimum-phase FIR filter to account for the frequency-dependent absorption properties of the materials involved in each reflection. This filter is based on absorption coefficients estimated from the surface image. Using the window method, we create a filter that applies a gain of $\beta_{k,f} = 1 - \alpha_{k,f}$ to each frequency band, where $\alpha_{k,f}$ is the absorption coefficient of the surface k in a frequency band f.

$$Y(z) = \sum_{k=0}^{M} H_k(z) \cdot g_{S_k,R} \cdot X(z) \cdot z^{-D_{S_k,R}} \tag{3}$$

The combined effect of all early reflection paths is captured by (3) where $X(z)$ is the input signal and $Y(z)$ represents the contributions from all reflection paths. Each term in the summation corresponds to a different reflection, where $H_k(z)$ is the transfer function of the FIR filter for the kth reflection. The result is a set of reflections that model the time and frequency-dependent characteristics of a room's early energy profile.

3.3.2 Late Reverberation

Our objective is to create a smooth late reverb that accurately mirrors the frequency-dependent decay characteristics of the target room. Therefore, we assume that the acoustic field is diffuse. We make this assumption since we are interested in reasonably reverberant rooms.

High echo and modal density correlate with smooth reverberation [18]. To achieve this while maintaining a low system order, we incorporate small velvet noise FIR filters into each element of the mixing matrix. This technique is supported by findings in [5].

The FDN delay times are determined using the values derived from (1). To further enhance the naturalness of the reverb, we address potential metallic ringing common in artificial reverb. We do this by following Schroeder's heuristic [19] that using mutually prime numbers for delay times can reduce such resonances by effectively spreading out the room modes. Therefore, for each delay $D_{S_k,R}$, we round up to the nearest prime number.

The matrix used in the FDN dictates the extent of mixing between the delay lines. Different matrix types are selected based on the system's order or computational requirements. To utilise all $D_{S_k,R}$ delays, we employ a random orthogonal matrix, which can accommodate any FDN order. However, it is not a preferred choice due to its non-deterministic nature and time complexity of $O(N^2)$[20]. An alternative is the circulant matrix, which can be arbitrarily sized and computed efficiently using the Discrete Fourier Transform of its first row [21].

The Hadamard matrix is typically utilised in FDNs due to its computational efficiency, with a $n \log n$ fast transform [20], and maximally diffuse response [22]. However, Hadamard matrices are only valid for matrix orders that are powers of 2. Given that the number of image sources for the Mth order reflections is $\sum_{m=1}^{M} N(N-1)^{m-1}$ where N represents the number of surfaces, when $N = 6$ and $M = 2$, the number of image sources becomes 36. To adapt this to a valid Hadamard matrix order, we apply k-th sampling (4) to evenly distribute the delay times, thus reducing the order to the nearest power of 2. Where K is the number of image sources and N is the target number of delay lines:

$$D_{\text{sampled}} = \left\{ D_{S_k,R} \mid k = i\frac{K}{N}, i = 0, \ldots, N-1 \right\} \tag{4}$$

3.3.3 Frequency Dependent Reverb Time

The decay envelope across different frequency bands in the late reverberation should match that of the target room. To achieve this, we use (5) to find the RT_{60} at frequency bands, where A is total absorption at frequency band f . Using this equation with the absorption coefficients for each surface, we calculate the reverberation times across octave band centre frequencies [125, 250,

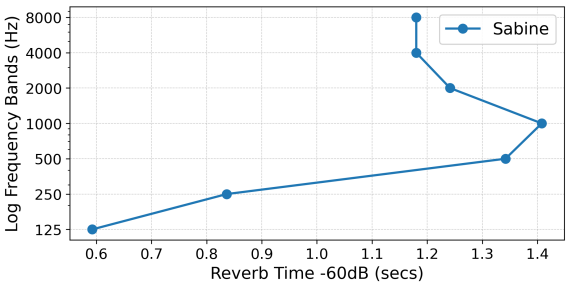

Fig. 5: Sabine RT60 at octave bands based on the extracted geometry and absorption coefficients estimated from surface images.

... 8000 Hz]. The results of these calculations for the Small Hallway room are shown in Fig. 5.

$$RT_{60,f,\,\text{Sabine}} = \frac{0.161V}{A_f} \tag{5}$$

In an FDN, the absorption filters control frequency-dependent decay time, characterised by RT_{60} at frequency bands. Each delay line is associated with a filter and parametrised based on its length and the desired reverb time. Using (6) converts the total -60 dB decay over time to the equivalent -1 dB per sample decay, scaled according to the length of the delay line. The slope is then converted from dB to magnitude and used to control the gain of the absorption filter.

$$\text{slope}_{dB} = \text{delay}\frac{-60}{RT_{60}f_s} \tag{6}$$

For computational efficiency, first-order IIR filters, which set the decay time at both $F_s/2$ and 0 Hz, are often used. Also, perceptual studies [23] have shown that reverb time should be controlled across multiple frequency bands. Alternatively, for precise control of decay across many frequency bands, FIR filters can be used where the gains are based on the $RT_{60,f}$ at octave-bands. We implement the late reverberation device using one-pole IIR and FIR filter approaches.

In the IIR approach, we employ a combination of high and low-frequency shelving filters, as well as a mid-frequency crossover band, as proposed in [24]. The decay times at $F_s/2$ and 0 Hz, are set to the $RT_{60,f,\,\text{Sabine}}$ values calculated for the 8000 Hz and 125 Hz octave bands, respectively. To determine the crossover frequency, we use (7) to apply the Schroeder formulation of the transition frequency [25]. This crossover is typically chosen as a multiple between 1 and 4 of the frequency. In our implementation, we use a multiple of 2, although this choice is often guided more by intuition than by strict calculation.

$$f_{\text{Schroeder}} = 2000\sqrt{\frac{RT_{60}}{V}}, \tag{7}$$

For the FIR approach, we calculate the required attenuation per sample at octave-bands based on the predicted $RT_{60,f,\text{Sabine}}$ and the corresponding delay length using (6) and fit a filter with the window method. To simulate the fast energy decay of sound at higher frequencies, we set the RT_{60} at Nyquist to be zero.

To summarise, we combine the ISM with the Scattering FDN in a hybrid system to implement a reverb with geometrically accurate early reflections and late reverberation matching the room's decay profile. In the evaluation below, the presented model will be referred to as the ISMFDN, followed by the FDN delay-line order N and its absorption filter type, FIR or One-pole.

4 Evaluation

To evaluate the success of the proposed system, we compare the ISMFDN One-pole and FIR methods to the SDN and ISM. Each is parametrised based on the geometry and surface absorption coefficients we extract from the room. We also include a standard one-pole absorption Hadamard FDN configured via the method presented in the late reverberation section. Additionally, a ground truth RIR was recorded in a target room, referred to as Small Hallway, shown in Fig. 6. We use the sine sweep and deconvolution method [26] to reduce the noise floor and maximise frequency coverage. The HISS [27] Max/MSP package was used to generate the sine sweep, played out of a Quested VS2205 and recorded by a SoundField SPS-442 surround sound microphone. The speaker and microphone were placed at length 0.5 m, width 1.02 m, height of 1.24 m and, length 2.2 m, width 1.02 m height 1.35 m, respectively. In all models, where possible, the source and microphone are placed at their respective locations.

4.1 Objective

In this section, we present the objective metrics used to evaluate the effectiveness of the reverberation algorithms, focusing on how closely they replicate the

Fig. 6: Recording setup for creating the small hallway RIR.

characteristics of a small hallway space Fig. 6. We evaluate the results using the metrics: RT_{60} and Modal Density at ocatve-bands and the echo density up to the mixing time of the Small Hallway (0.83 sec). To measure performance, we take the normalised root mean squared error (RMSE) between the metrics for the synthesised RIRs and Small Hallway RIR . We then invert the result so the performance scale is 0 (worst match) to 1 (best match); see Fig. 7.

4.1.1 Reverb Time

To measure $RT_{60,f}$ we use an octave-band filter bank to decompose the RIR. For each band we apply Schroeder's Energy Decay Curve method (8), where $E(t)$ is the Schroeder integral [28] and find the point at which energy decays by 60 dB.

$$RT_{60,\text{ Measured}} = 6 \cdot \frac{E(t)}{\text{Slope of } E(t)} \tag{8}$$

The resulting RMSE $RT_{60,f,Measured}$ values across all models are plotted in Fig. 7. The excellent performance of the ISM falls in line with expectations since it models absorption at each reflection. Additionally, we expect a close match from the SDN as it approximates the ISM. The moderate performance of the FIR ISMFDN is due to errors introduced by using the Sabine equation to estimate decay times. The best models (ISM, SDN and FIR ISMFDN) have absorption filters based on all octave bands of each surface's material coefficients. The worst-performing models (One-Pole ISMFDN and Hadamard FDN) have first-order absorption filters based solely on the absorption coefficients of the maximum and minimum octave bands.

4.1.2 Modal Density

Modal Density refers to the number of resonating frequencies within a given range. High density typically

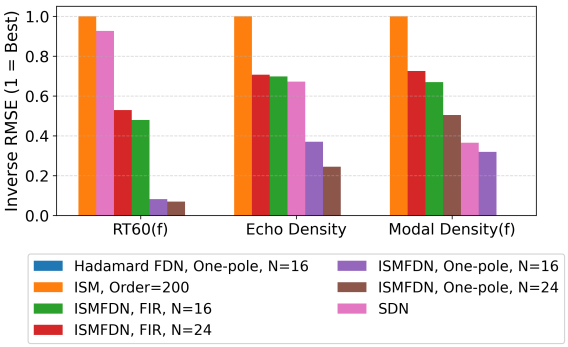

Fig. 7: Normalized performance of models for modal density and RT_{60} at octave-bands and echo density, (1=Best).

contributes to a smoother and more diffuse reverberation. In [19], the authors suggest that a minimum of 0.15 modes per Hz is sufficient for one second of reverb.

We estimate modal density by conducting spectral peak picking using parabolic interpolation [29]. This method is chosen for its robustness against noise in the frequency domain. The modal density at a frequency range is calculated as:

$$\text{Modal Density} = \frac{\text{Number of Detected Peaks}}{\text{Frequency Range}}.$$

We define modal density at a octave-band centre frequency f by counting how many detected peaks lie between $f/\sqrt{2}$ and $f\sqrt{2}$. We then divide the count by the total number of frequencies in the same range.

Fig. 7. shows the success of the ISMFDN in its FIR filter configuration. The higher-order FDN (N=24) system achieves greater density in both FIR and One-pole cases. This is due to modal density in FDNs being proportional to the summation of all delay lengths [30]. The standard Hadamard FDN struggles to generate sufficient modal density due to its simplified mixing matrix. Moreover, the ISM model performs well, consistent with its established reputation for accurately modelling acoustics in simple room geometries. Interestingly, the SDN performs poorly in this comparison. While SDNs are often praised for their natural-sounding reverberation [31], from our findings, their ability to provide realistic modal density appears limited.

4.1.3 Echo Density

Echo density is a metric that reflects how densely packed the echoes become as the sound reverberates within a space. The echo density metric, as defined in [32], measures a RIRs similarity to a Gaussian distribution over time. Closer matches to the distribution correlate with higher levels of sound diffusion.

The results in Fig.7 show the ISM achieves the closest early echo density profile to the Small Hallway RIR, significantly outperforming other models. The ISMFDN with FIR absorption filters shows the second-best performance. On the other hand, the one-pole absorption methods in the Hadamard FDN and ISMFDN demonstrate considerably lower performance.

4.2 Subjective

We conducted an in-person subjective quality rating test in the small hallway room with 5 participants, all knowledgeable in audio production. They rated the naturalness of reverberation applied to a set of sounds based on the room where the test was conducted. We used four samples to base the stimuli. Anechoic flute, drums and speech samples from the University of York's OpenAIR dataset (www.openair.hosted.york.ac.uk) and a Roland 808 drum machine clap. Each is convolved with RIRs generated by the models used in the objective analysis and the Small Hallway RIR. All stimuli were normalised, and an unprocessed sample was also included as an anchor. The test comprised four rounds, with participants rating each sample on a scale from 1 (very unnatural) to 9 (extremely natural). Each stage of the test presented different samples processed by each RIR, with the order of stages and samples randomised. A training stage was conducted first to familiarize participants with the test procedure.

The results in Fig. 8 indicate a preference for the ISMFDN methods, which generally scored higher on average across the listening test. The Hadamard FDN performs poorly, reflecting its limitations in achieving perceptually relevant reverberation. The ISM method and SDN scored moderately, demonstrating that while effective, they are preferred slightly less than the ISMFDN approaches. As expected, the real RIR achieved the highest ratings with minimal spread. The anechoic sample consistently received the lowest score, validating the participants' understanding of the test. Although the ISMFDN methods generally scored higher, the overall close distribution of scores across models suggests no obvious winner. Table 1 shows how rating varied depending on the sample type, with the same model receiving drastically different scores. This highlights the subjective nature of reverb preference and the challenges in determining the most natural-sounding model.

5 Conclusion

We present a method for extracting acoustic reflection paths and surface absorption from lidar scans and room images. We show the process applied to our hybrid artificial reverberator design and various state-of-the-art models. We evaluate the models by comparing them to a small hallway using reverberation metrics for objective analysis and a listening test for perceptual evaluation.

The promising listening test results show that the proposed reverberator outperforms similar state-of-the-art models. On the other hand, further work is needed to evaluate the model perceptually. With only five participants, the current listening test limits the ability to draw definitive conclusions about the model's viability. A more extensive MUSHRA [33] study comparing this model with the ISM across various perceptually informed metrics is essential.

The model performs well on the modal and echo density metrics, which we expect based on results in [5]. On the other hand, it loses out to the more geometrically informed ISM and SDN models on $RT_{60,f}$. This is due to the estimation of the FDN absorption filter coefficients with the Sabine equation. A possible solution to this problem could be calibrating the Sabine equation via the non-linear optimisation method shown in [34].

Acknowledgement

I want to thank Ilias Ibnyahya for his time and advice throughout the research.

Appendix

Source code related to this paper is available from `https://github.com/ruarim/geometric_and_visual_artificial_reverberation`

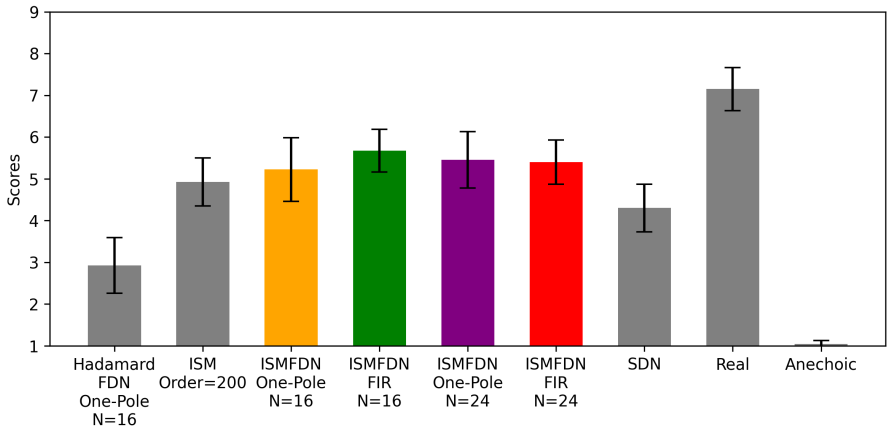

Fig. 8: Overall results of the listening test. The bars show the mean. The whiskers represent the 85% confidence interval. ISMFDN Models in, Yellow, Green, Purple and Red.

Table 1: Mean listening test results.

Model	Sample				
	Flute	**Clap**	**Drums**	**Speech**	**Overall**
ISM	5.9	4.1	4.4	5.3	4.9
SDN	5.8	3.2	4.5	3.7	4.3
Hadamard FDN	5.6	2.2	1.7	2.2	2.9
ISMFDN FIR N=24	4.7	5.8	5.3	5.8	5.4
ISMFDN FIR N=16	**6.1**	4.4	**6.9**	5.3	**5.7**
ISMFDN One-Pole N=24	4.8	**6.0**	6.2	4.8	5.5
ISMFDN One-Pole N=16	5.2	4.4	4.8	**6.5**	5.2
Small Hallway	7.2	7.0	7.6	6.8	7.2

References

[1] Allen, J. B. and Berkley, D. A., "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, 65(4), pp. 943–950, 1979.

[2] Gerzon, M. A., "Synthetic Stereo Reverberation – Part I," *Studio Sound*, 13(12), pp. 632–635, 1971.

[3] Gerzon, M. A., "Synthetic Stereo Reverberation – Part II," *Studio Sound*, 14(1), pp. 24–28, 1972.

[4] Stautner, J. and Puckette, M., "Designing Multi-Channel Reverberators," *Computer Music Journal*, 6(1), p. 52, 1982.

[5] Schlecht, S. J. and Habets, E. A., "Scattering in feedback delay networks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28, pp. 1915–1924, 2020.

[6] De Sena, E., Hacihabiboglu, H., Cvetković, Z., and Smith, J. O., "Efficient Synthesis of Room Acoustics via Scattering Delay Networks," *IEEE/ACM Trans Audio Speech Lang Process*, 23(9), pp. 1478–1492, 2015.

[7] Murphy, D., Newton, C., and Howard, D., "Digital waveguide mesh modelling of room acoustics: surround-sound, boundaries and plugin implementation DAFx: Proc," in *In Proc. Int. Conf. on Digital Audio Effects (Limerick, Ireland*, 2001.

[8] Carpentier, T., Noisternig, M., and Warusfel, O., "Hybrid reverberation processor with perceptual control," in *In Proc. 17th International Confer-*

ence on Digital Audio Effects-DAFx-14, pp. 93–100, 2014.

[9] Ibnyahya, I. and Reiss, J. D., "A method for matching room impulse responses with feedback delay networks," in *Audio Engineering Society Convention 153*, Audio Engineering Society, 2022.

[10] Murphy, D. and Stewart, R., "A hybrid artificial reverberation algorithm," in *Audio Engineering Society Convention 122*, Audio Engineering Society, 2007.

[11] Chen, C., Gao, R., Calamia, P., and Grauman, K., "Visual acoustic matching," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 18858–18868, 2022.

[12] Ratnarajah, A., Ghosh, S., Kumar, S., Chiniya, P., and Manocha, D., "Av-rir: Audio-visual room impulse response estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 27164–27175, 2024.

[13] Singh, N., Mentch, J., Ng, J., Beveridge, M., and Drori, I., "Image2reverb: Cross-modal reverb impulse response synthesis," in *In Proc of the IEEE/CVF International Conference on Computer Vision*, pp. 286–295, 2021.

[14] Colombo, M., Dolhasz, A., Hockman, J., and Harvey, C., "Acoustic rendering based on geometry reduction and acoustic material classification," in *In Proc of 2022 IEEE Conference on Games (CoG)*, pp. 409–416, IEEE, 2022.

[15] Colombo, M., Dolhasz, A., and Harvey, C., "A texture superpixel approach to semantic material classification for acoustic geometry tagging," in *In Proc. Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–7, 2021.

[16] Hurst, A., Lerer, A., Goucher, A. P., Perelman, A., Ramesh, A., Clark, A., Ostrow, A. J., Welihinda, A., Hayes, A., Radford, A., and Mądry, A., "GPT-4o system card," 2024, arXiv preprint arXiv:2410.21276.

[17] Vorländer, M., *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*, Springer-Verlag, Berlin, Germany, 2008.

[18] Santo, G. D., Prawda, K., Schlecht, S. J., and Välimäki, V., "Feedback Delay Network Optimization," *arXiv preprint arXiv:2402.11216*, 2024.

[19] Schroeder, M. R. and Logan, B. F., "" Colorless" artificial reverberation," *IRE Transactions on Audio*, (6), pp. 209–214, 1961.

[20] Schlecht, S., "FDNTB: The feedback delay network toolbox," in *In Proc. International Conference on Digital Audio Effects*, pp. 211–218, DAFx, 2020.

[21] Rocchesso, D. and Smith, J. O., "Circulant and elliptic feedback delay networks for artificial reverberation," *IEEE Trans Speech Audio Process*, 5(1), pp. 51–63, 1997.

[22] Rocchesso, D., "Maximally diffusive yet efficient feedback delay networks for artificial reverberation," *IEEE Signal Process Lett*, 4(9), pp. 252–255, 1997.

[23] Jot, J.-M. and Chaigne, A., "Digital delay networks for designing artificial reverberators," in *Audio Engineering Society Convention 90*, Audio Engineering Society, 1991.

[24] Välimäki, V. and Reiss, J., "All About Audio Equalization: Solutions and Frontiers," *Applied Sciences*, 6(5), p. 129, 2016.

[25] Schroeder, M. R., "The "Schroeder frequency" revisited," *The Journal of the Acoustical Society of America*, 99(5), pp. 3240–3241, 1996.

[26] Farina, A., "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Proc. AES 108th Conv., Paris, France*, 2000, paper no. 5093.

[27] Harker, A. and Tremblay, P. A., "The HISSTools impulse response toolbox: Convolution for the masses," in *In Proc. the international computer music conference*, pp. 148–155, The International Computer Music Association, 2012.

[28] Schroeder, M. R., "New Method of Measuring Reverberation Time," *J. Acoust. Soc. Am.*, 37, pp. 409–412, 1965.

[29] Smith, J. O., *Spectral Audio Signal Processing*, http:http://ccrma.stanford.edu/ jos/sasp//-ccrma.stanford.edu/~jos/sasp/, accessed 2024, online book, 2011 edition.

[30] Schlecht, S. J. and Habets, E. A. P., "Modal Decomposition of Feedback Delay Networks," *IEEE Trans Signal Process*, 67(20), pp. 5340–5351, 2019.

[31] Djordjevic, S., Hacihabiboglu, H., Cvetković, Z., and De Sena, E., "Evaluation of the perceived naturalness of artificial reverberation algorithms," in *Audio Engineering Society Convention 148*, Audio Engineering Society, 2020.

[32] Abel, J. S. and Huang, P., "A simple, robust measure of reverberation echo density," in *Audio Engineering Society Convention 121*, Audio Engineering Society, 2006.

[33] Series, B., "Method for the subjective assessment of intermediate quality level of audio systems," *International Telecommunication Union Radiocommunication Assembly*, 2, 2014.

[34] Prawda, K., Schlecht, S. J., and Välimäki, V., "Calibrating the Sabine and Eyring formulas," *The Journal of the Acoustical Society of America*, 152(2), pp. 1158–1169, 2022.