HIGHER-ORDER SCATTERING DELAY NETWORKS FOR ARTIFICIAL REVERBERATION

Matteo Scerbo , Orchisama Das , Patrick Friend and Enzo De Sena

University of Surrey
Department of Music and Media
Guildford, United Kingdom
m.scerbo@surrey.ac.uk | o.das@surrey.ac.uk | e.desena@surrey.ac.uk

ABSTRACT

Computer simulations of room acoustics suffer from an efficiency vs accuracy trade-off, with highly accurate wave-based models being highly computationally expensive, and delay-network-based models lacking in physical accuracy. The Scattering Delay Network (SDN) is a highly efficient recursive structure that renders first order reflections exactly while approximating higher order ones. With the purpose of improving the accuracy of SDNs, in this paper, several variations on SDNs are investigated, including appropriate node placement for exact modeling of higher order reflections, redesigned scattering matrices for physically-motivated scattering, and pruned network connections for reduced computational complexity. The results of these variations are compared to state-of-the-art geometric acoustic models for different shoebox room simulations. Objective measures (Normalized Echo Densities (NEDs) and Energy Decay Curves (EDCs)) showed a close match between the proposed methods and the references. A formal listening test was carried out to evaluate differences in perceived *naturalness* of the synthesized Room Impulse Responses. Results show that increasing SDNs' order and adding directional scattering in a fully-connected network improves perceived naturalness, and higher-order pruned networks give similar performance at a much lower computational cost.

1. INTRODUCTION

Models that simulate the acoustical behaviour of environments are employed in several fields, e.g. in the entertainment industry, where they can apply artificial reverberation to sound effects to make them match a movie or video game's scenery, in music productions to add depth and warmth, or in Virtual Reality (VR) / Augmented Reality (AR) applications to increase immersion [1, 2]. Acoustic models are also employed by architects and civil engineers to predict the acoustic responses of the designed spaces and to ensure, for instance, a pleasing reverberation in listening spaces, intelligible speech in learning or public spaces, or reduced noise levels in busy environments.

Starting with the Schroeder reverberator [3, 4], a wide variety of approaches for room acoustics simulation have been introduced over the past six decades [5, 6]. Currently, there are several prominent classes of such acoustic simulators, each rooted in profoundly different concepts. Each approach has advantages and drawbacks, in particular there tends to be a trade-off between performance and computational complexity. In general, said trade-off is relatively fixed for each approach, with some approaches being much more suitable for fast-running but physically inaccurate simulations, while others are exclusively capable of highly accurate simulations with long running times.

One of the more recent propositions is known as Scattering Delay Networks (SDNs) [7], which are derived from Digital Waveguide Networks (DWNs) [8]. The goal of these systems is to combine the high efficiency of Digital Waveguide Networks (DWNs) [8] – which are very lightweight but not physically accurate – with the basic concepts of geometric models such as the Image Method (IM) or Ray Tracing – which can reach high degrees of physical accuracy but are far more complex than DWNs, and cannot be rendered in real-time. SDNs are designed to simulate 1st order reflections exactly, while higher order reflections are approximated more and more coarsely. Djordjević *et al.* [9] compared SDNs' perceptual qualities and found that they perform better than Feedback Delay Networks (FDNs) and two geometric methods (IM and Ray Tracing). With their low computational requirements and good perceptual accuracy, SDNs have proven advantageous for several applications, such as real-time binaural rendering [10, 11] and AR audio applications [12]. Several works have also proposed extensions or adaptations of SDNs to allow more flexibility with relation to the simulated room geometry [13, 14] or to provide more physically-accurate scattering [15]. The purpose of the present work is to propose a design framework which would allow gradually increasing SDNs' geometrical accuracy, paving the way towards more physically accurate extensions. This was tackled by increasing the number of accurately modelled reflections to include higher orders, and rendering the recursive higher order reflections more accurately. Two different approaches for this were designed and tested.

The paper is organised as follows. The fundamental concepts of SDNs will be explained in Sec. 2, along with some recent advancements in the field. Sec. 3 will introduce the basic idea behind the introduction of more accurate reflections, the issues that arise from it (incorrectly duplicated echoes, degenerating scattering matrices), and two possible ways of circumventing those issues. Finally, Sec. 4 will present the results of both objective measurements and subjective experiments conducted on the proposed methods, and Sec. 5 will draw some conclusions on said results and suggest possible directions for future work.

2. BACKGROUND

An SDN is composed of delay lines that connect the sound source to the virtual microphone, and to walls via scattering nodes. The

scattering nodes are all interconnected with delay lines and scattering matrices. There are attenuating filters at the end of delay lines to model wall-absorption. The source and listener positions, and room geometry can be adapted in real-time to simulate moving sources and listeners. The structure of the network comprises a "source node" connected to every other node by uni-directional outgoing delay lines, a "receiver node" connected to every other node by uni-directional incoming delay lines, and a number of "scattering nodes", which are connected to each other by bidirectional delay lines.

The fundamental operation sees the input signal being injected into the network by the source node, and the output signal being read at the receiver node. Each scattering node's internal operation consists of reading the signals from incoming delay lines, multiplying them by some scattering coefficients to obtain the outgoing signals, and filtering those with the attenuation filters before writing them to the outgoing delay lines. With the sets of incoming and outgoing signals being grouped into vectors, the scattering coefficients form a matrix, such that the matrix multiplication of the incoming signal vector by the scattering matrix gives the outgoing signal vector. The scattering matrices are designed to be unitary [16], which means that the scattering operation itself does not introduce or remove energy from the system [17]. Energy is only introduced by the input signal, and only removed by the attenuation filters, which guarantees stability. The original SDN employs the Digital Waveguide Mesh (DWM) scattering matrix, defined as

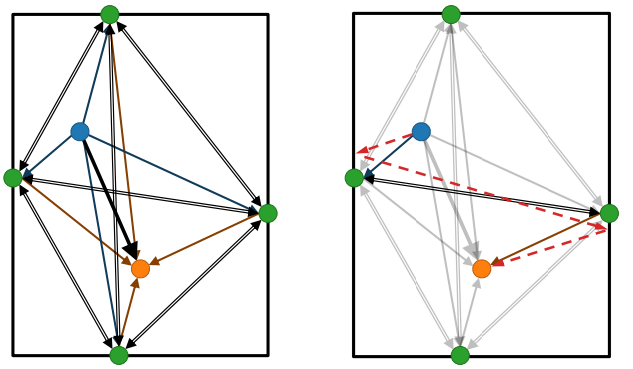
$$\mathbf{S} = \frac{2}{N-1}\mathbf{1}_{(N-1)\times(N-1)} - \mathbf{I}, \tag{1}$$

where N is the number of scattering nodes, $\mathbf{1}$ is the unit matrix, and $\mathbf{I}$ is the identity matrix. Note that this is the same for every scattering node, and the scattering coefficients do not depend in any way on the relative directions of the incoming/outgoing signal pairs, hence it is referred to as *isotropic* scattering. Other possible choices for unitary scattering matrices will be explored in the next sections.

Of particular note is the fact that the SDNs' parameters are based on physical characteristics of the environments one wants to simulate. Each node corresponds to a position in space, and the lengths of the delay lines are based on the sound propagation time between those points. The source node and receiver node are placed in the desired positions of the sound source and microphone respectively, while the scattering nodes are placed on the points of 1st order reflections, found geometrically as in the IM. Fig. 1 shows the placement of nodes in space, for the case of a 2-dimensional room[1]. The gain that each scattering node applies to the linear combination of its inputs is based on the absorbing properties of the wall material.

Since both the delay and the attenuation correspond to those 1st order reflections, the SDN accurately models said reflections perfectly, by propagating the signal along the delay lines from the source to each scattering node and from each scattering node to the microphone. The delay lines between scattering nodes, on the other hand, introduce the recursion (feedback) necessary to model higher order reflections. In many cases, such as the simple geometry of "shoebox" rooms, the propagation along these lines offers a convincing approximation of 2nd order reflections (see Fig. 1b), a coarse approximation of 3rd order reflections, and so on. This progressive approximation of higher order reflections is acceptable,

[1]This and other figures are 2-D for illustrations purposes, while all presented results are for 3-D simulations.



(a) *A 1st order SDN. Double lines mark bi-directional connections.*

(b) *Approximation of a 2nd order reflection: the dashed line shows the correct (specular) path.*

Figure 1: *Standard SDNs for a 2-D rectangular room. Blue - source node, green - scattering nodes, orange - receiver node.*

since the perceptual importance of reflections' details diminishes with their order [18].

Since their introduction, SDNs have been extended in a number of ways to improve their flexibility and accuracy. For example, Pekçetin [13] adapted SDNs for non-shoebox room geometries, by considering visibility and edge diffraction. Stevens *et al.* [15] went further by considering outdoors and sparsely reflecting environments. Signals are scattered in a frequency dependent manner by use of filters, the energy loss due to open spaces is modeled by fully absorbent "sky-nodes", and multiple adaptations provide accurate modeling of 2nd order reflections. The latter point was achieved by adding scattering nodes in positions of 2nd order reflections. This concept will be expanded upon in the next section.

As mentioned earlier, unitary scattering matrices are important for the energy stability of the system. Schlecht and Habets [19] introduced a method for creating unitary, direction-dependent scattering matrices based on the Bidirectional Reflectance Distribution Function (BRDF) [20]. With these matrices, scattering between nodes depends on the directions of incoming and outgoing lines, specifically on the similarity between the outgoing direction and the correct specular direction. The scattering coefficient between the i^{th} incoming node (represented by direction vector $\boldsymbol{v_i}$) and the j^{th} outgoing node (represented by direction vector $\boldsymbol{v_j}$) is given by

$$s_{i,j} = \frac{1}{\angle \boldsymbol{v_r}, \boldsymbol{v_j} + 0.1}, \tag{2}$$

where $\boldsymbol{v_i}$ is the incoming direction, $\boldsymbol{v_j}$ is the outgoing direction, and $\boldsymbol{v_r} = \boldsymbol{v_i} - 2(\boldsymbol{v_i}^\top \hat{\boldsymbol{n}})\hat{\boldsymbol{n}}$ is the reflected direction at the wall surface normal $\hat{\boldsymbol{n}}$. The matrices constructed in this way are then converted to their closest sign-agnostic unitary form as explained in [19].

3. PROPOSED HIGH-ORDER SDNS

The objective of this work is to increase the number of accurately modelled reflections, in a way that allows flexible scaling of the reflection order. With directional scattering in mind, for the following extensions it was chosen to add scattering nodes in the positions of higher order reflection points, as shown in Fig. 2.

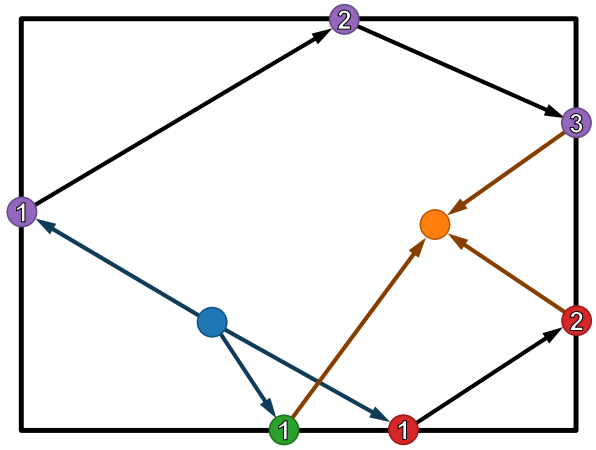


Figure 2: *Three possible specular reflection paths, and scattering nodes' roles in the respective paths, in a 2-D rectangular room. Green represents 1st order reflections, red 2nd, purple 3rd. Numbers on the nodes denote the order in which they are traversed to model the specular reflections. Note that each reflection involves a number of scattering nodes equal to its order.*

With this strategy, a reflection of order R will add R nodes; in a generic room, the total number of scattering nodes N is then $N = \sum_{r=1}^{R} r \cdot S_r$, where R is the maximum reflection order considered and S_r is the number of (visible, distinct) image sources of order r. While S_r has the potential to grow exponentially with the reflection order in generic rooms, it is known to grow cubically in shoebox rooms [21].

3.1. Naive approach

Let us consider a case in which a standard SDN is extended by simply adding scattering nodes in the positions of high-order reflections (see Fig. 2), and naively treating the new nodes in the same manner as 1st order nodes in the standard SDN. The scattering nodes in this *naive SDN* will still present isotropic scattering matrices, and will still be connected to every other node – source, receiver, and all scattering nodes (except the ones laying on the same wall, assuming locally-reactive wall surfaces [22]).

The total number of connections C between nodes is given by $C = N \cdot (N-1) + 2N + 1$, since each of the N scattering nodes is connected bi-directionally to the remaining $N-1$ scattering nodes as well as uni-directionally to both the source and receiver nodes, and the source and receiver nodes are connected uni-directionally to each other. This equation ignores the fact that nodes on the same wall are not connected. The actual number of connections will be somewhat smaller, but still depends on N quadratically.

Apart from the apparent increase in complexity, taking this approach introduces several issues. First of all, adding scattering nodes to the network without any further precautions leads to duplicated reflections. For example, Fig. 3 shows some reflection paths involving a cluster of 1st and 2nd order scattering nodes on the same wall. Since the scattering operation performed by the nodes is not directional, the resulting modeled reflections will all be subject to the same attenuation, leading to a tight cluster of identical-amplitude echoes. If the intention is to model directional scattering, each echo should be attenuated differently; if the intention is to model strictly specular reflections, only one echo should be present at all – the one involving the single 1st order node. The same duplication issue is present in higher order reflections. Take for example the 2nd order reflection in Fig. 1b: the reflection as approximated by the 1st order scattering nodes is now present *in addition* to the one correctly modeled by 2nd order scattering nodes.

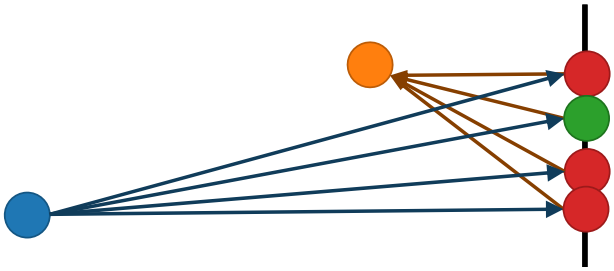


Figure 3: *1st order reflections modeled by a cluster of scattering nodes. The green circle denotes a 1st order nodes, while red circles denote 2nd order nodes.*

Another issue emerges as a consequence of the scattering matrices being isotropic. As the number of connections at each scattering node increases, the isotropic scattering matrices used in the standard SDNs converge to identity matrices (see Eq. 1). In practice, this means that when a signal reaches a scattering node over an incoming delay line, most energy is reflected back in the incoming direction, with nearly no energy being transmitted to other scattering nodes. As will be shown by test results in the next section, this matrix degeneration aggravates the abrupt early energy decay which was already observed in SDNs [23] and leads to irregular echo density behaviour.

3.2. Directional matrix approach

One possible approach with relation to the limitations of the discussed *naive SDN* approach is to use direction-dependent scattering matrices such as those seen in Eq. 2. While this does not alleviate the increase in complexity, it does tackle the issues of repeating echoes and isotropic matrix degeneration. The duplicated echoes are attenuated based on their direction: different reflections will be subject to different attenuations based on their directions of incidence, with most energy being directed towards the specular path. For the same reason, the backwards-reflecting behaviour of the matrices approaching identity is prevented, the early energy decay is improved, and the echo density behaviour regularized.

3.3. Connection-cutting approach

Computational complexity can be drastically reduced with little perceptual loss by pruning connections between nodes. This is in line with the rationale behind standard SDNs, which use a minimal topology, where "minimal" signifies that the removal of any of the nodes or paths would make the system unable to model a significant number of specular reflections. Connection cutting can furthermore be exploited to tackle the *naive SDN*'s issues, by removing paths that cause the discussed duplicated reflections, favouring strictly specular reflections. For example, the issue related to duplicated direct reflections (see Fig. 3) can be overcome by cutting some of the source-to-node and node-to-receiver connections.

Recall Fig. 2, and note how – when it comes to strictly specular reflections – the source and receiver nodes are not connected to all scattering nodes. Specifically, only the subset of scattering nodes which are encountered first on their respective paths are directly connected to the source, and only those encountered last to the receiver. Employing such a connection arrangement for source and receiver nodes removes the erroneous 1st order reflections that involve higher order scattering nodes, but does not affect the duplication of higher order reflections, which is due to node-to-node connections. This network structure will be referred to as *reduced* SDN.

The same idea can, however, be extended to the extreme. Let us start by considering the absolute minimum network structure that is capable of simulating all specular reflections up to a certain order, and no more – its behaviour would be that of an FIR filter. The first step in building such a network is to make note of the order in which reflection points are encountered while moving from the source to the receiver (recall again Fig. 2). This can be done differently depending on the chosen method for computing reflection point positions. Like in the *reduced* SDN, only the subset of scattering nodes which are encountered first on their respective paths shall be directly connected to the source, and only those encountered last to the receiver. In addition, the nodes along the path shall only be directly connected to their predecessor and successor, uni-directionally. The solid lines in Fig. 4 show a network view of such a system: note again that there is no recursion or feedback present.

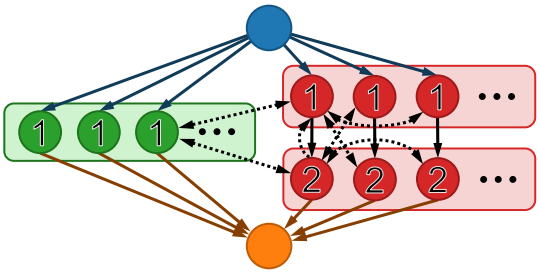


Figure 4: *A network view of the minimal connection scheme. Scattering nodes are colored by order (Green - 1st order, red - 2nd order) and grouped by their place in the reflection sequence. Unbroken lines show minimum connections to model specular reflections, dotted lines show a few possible added connections that would introduce recursion.*

The number of connections in such bare-bones network is drastically reduced to $C = \sum_{r=1}^{R}(1+r)\cdot S_r$. This is due to the fact that, for each specular reflection path, there is a connection from the source node to the first scattering node, a connection from the last scattering node to the receiver, and $r-1$ connections between the scattering nodes. Since this network is, by design, only capable of producing a finite number of reflections, some connections must be added to introduce the recursive properties characteristic of SDNs, as shown by the dotted lines in Fig. 4. For the sake of testing the importance (or lack thereof) of the strategy for picking these additional connections, for the experiments that will be presented in the next section the recursive connections were chosen pseudo-randomly. More specifically, two bi-directional connections (meaning four delay lines) were added from every scattering node in the system to two scattering nodes chosen pseudo-randomly. On top of that, the connections between scattering nodes on each reflection path were also made to be bi-directional, so that all scattering matrices were square and their diagonals always represented the returning attenuation for the incoming direction. Considering the above, plus the connections from the source and to the receiver, the number of delay lines in the network is $C = \sum_{r=1}^{R}(2+2(r-1)+4r)\cdot S_r = \sum_{r=1}^{R} 6\cdot r\cdot S_r$. This grows linearly with the number of nodes, rather than quadratically, leading to a substantial reduction of the computational complexity.

With this design, like in the *reduced* SDN, the duplicated 1st order echoes are removed entirely; the vast majority of higher order duplicates are also removed because of the extensive pruning. This strategy also prevents the scattering matrix from degenerating to the identity matrix, since the matrices are restricted to relatively small sizes by the limited number of connections at each node. This network structure will be referred to as *pruned* SDN.

4. EVALUATION

The set of Room Impulse Responses (RIRs) to be compared using both objective and subjective tests was constructed as follows. Eight variations of SDNs were considered, with characteristics as listed in Tab. 1. The *Standard* and *Identity* SDNs were included for reference, while the remaining six are the extensions proposed in the previous section. All SDN RIRs were produced through a purpose-made Python implementation of the structures. In addition to these SDNs, the experiments included an Image Method simulation with PyRoomAcoustics [24] and a Ray Tracing simulation using the state-of-the-art CATT-acoustic model v9.1 [25]. Sweeping echoes were removed from the IM case by adding a slight randomization (8 cm range) to the image sources' positions [26]. The CATT-acoustic simulation was run using the *closed room* algorithm, and advanced air absorption settings. For each of the mentioned methods, three simulations were conducted in different environments. All were windowless shoebox rooms, using the parameters reported in Tab. 2. In all cases, both the wall absorption and the scattering were assumed to be frequency-independent, while frequency-dependent air absorption was modeled.

Table 1: *Considered SDNs. "Reduced" refers to removing some source-node and node-receiver connections as discussed at the start of Sec. 3.3, preserving all node-node connections. "Pruned" refers to removing node-node connections as detailed later in the same section.*

Descriptor	Nodes' order	Connections	Matrices
Standard SDN	1st	full	isotropic
Identity SDN	1st	full	identity
BRDF SDN, 2nd	1st, 2nd	full	BRDF
BRDF SDN, 3rd	1st, 2nd, 3rd	full	BRDF
Reduced SDN, 2nd	1st, 2nd	reduced	isotropic
Reduced SDN, 3rd	1st, 2nd, 3rd	reduced	isotropic
Pruned SDN, 2nd	1st, 2nd	pruned	isotropic
Pruned SDN, 3rd	1st, 2nd, 3rd	pruned	isotropic

4.1. Objective measures

Energy Decay Curves and Normalized Echo Densities were employed as objective measures of the proposed methods' perfor-

Table 2: *Parameters used for the room simulations.*

Descriptor		Small	Medium	Large
Size	W	3.3 m	6.3 m	10.3 m
	D	5.3 m	9.3 m	20.3 m
	H	4.3 m	4.3 m	7.3 m
Source pos.	W	1.5 m	1.5 m	5.5 m
	D	1.5 m	1.5 m	3.5 m
	H	1.5 m	1.5 m	1.5 m
Mic pos.	W	3.7 m	5.7 m	7.7 m
	D	3.7 m	1.7 m	10.7 m
	H	2.7 m	2.7 m	1.7 m
Wall absorption		0.2	0.1	0.07
Humidity		70 %	50 %	30 %
Temperature		25 °C	20 °C	15 °C
Atm. pressure		100 kPa	100 kPa	100 kPa

mance. While both measures were evaluated in all rooms detailed in Tab. 2, only the *medium room* case's results are shown here. The following remarks on early energy decay and erratic echo density apply to all tests.

4.1.1. Energy Decay Curves

The EDC is defined as the total energy remaining in the RIR at time t. In most commonly-occurring scenarios, the EDC decays linearly[2] – on a dB scale – and the moment it reaches $-60\,\mathrm{dB}$ is defined as the reverberation time, T_{60}. The T_{60} values achieved by the different methods are presented in Tab. 3. In all rooms but the largest, the IM seems to achieve longer reverberation times than Ray Tracing, with different SDNs falling in between[3].

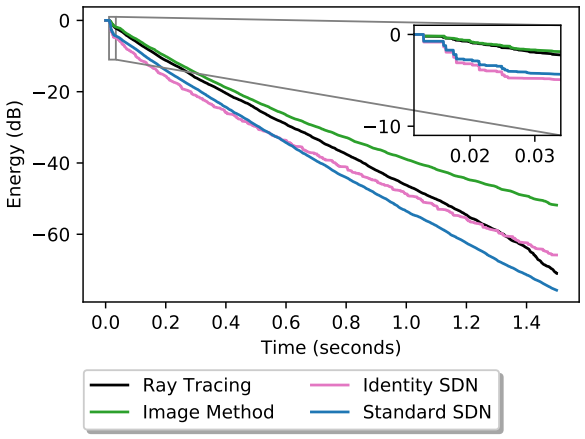
Table 3: T_{60} *values achieved by the compared methods.*

Method	T_{60} (seconds)		
	Small	Medium	Large
Standard SDN	0.46	1.15	5.29
BRDF SDN, 2nd order	0.50	1.20	5.43
BRDF SDN, 3rd order	0.52	1.22	5.55
Reduced SDN, 2nd order	0.48	1.21	5.68
Reduced SDN, 3rd order	0.48	1.19	5.81
Pruned SDN, 2nd order	0.52	1.20	5.37
Pruned SDN, 3rd order	0.56	1.27	5.61
Identity SDN	0.51	1.32	5.91
Image Method	0.57	1.88	5.13
Ray Tracing	0.45	1.32	3.31
Sabine estimate	0.55	1.55	3.69

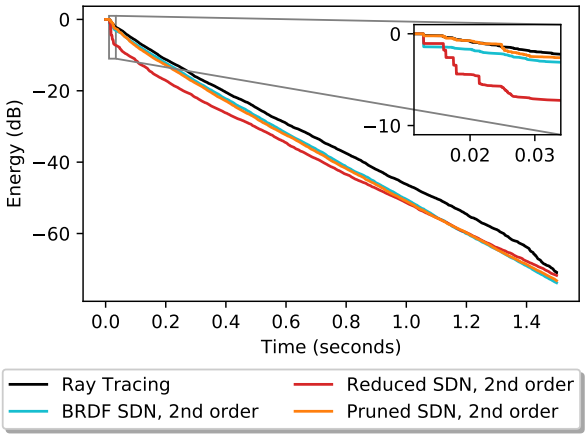
SDNs, especially those that use isotropic scattering matrices, tend to display an abrupt energy decay at the very start of the RIR [23], as can be seen in Fig. 5a. The energy lost in this abrupt decay increases if isotropic matrices are allowed to degenerate as discussed in Sec. 3.1, which is apparent in the *reduced SDN* plots in Figs. 5b and 5c. The same figures also show that both the *BRDF SDNs* and *pruned SDN* methods are able to reduce this effect.

[2]Non-linear decay behaviours can be observed in environments such as coupled volumes, see [14] for an SDN extension regarding such a case.

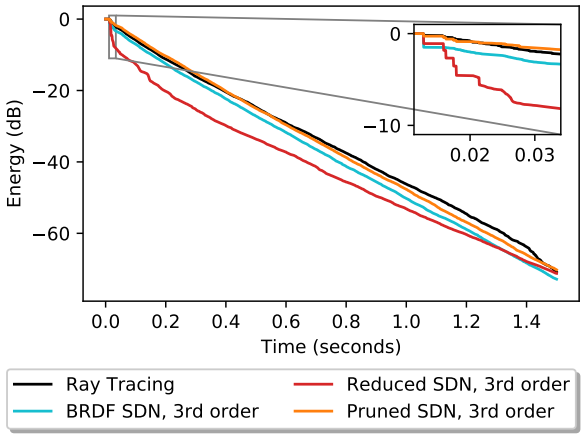
[3]The employed IM algorithm modeled air absorption much more coarsely than the other methods, retrieving attenuation values from a look-up table as opposed to computing them based on the given parameters.



(a) *References*

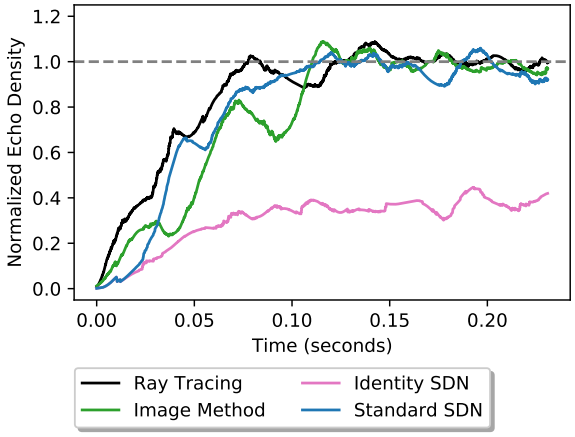


(b) *2nd order SDNs*

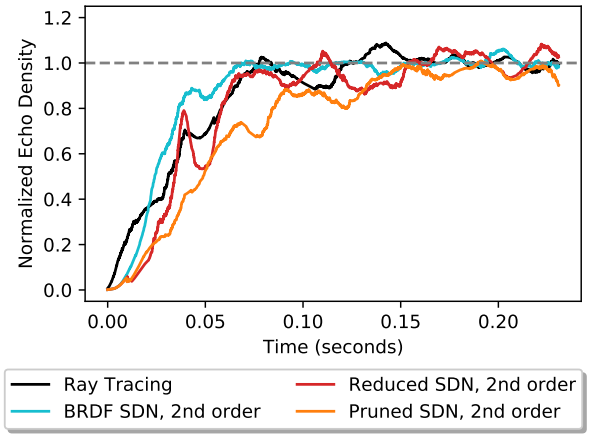


(c) *3rd order SDNs*

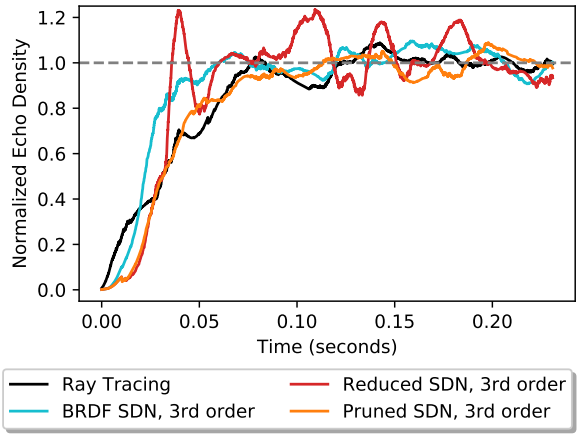
Figure 5: *Energy Decay Curves of the compared methods. An inset magnification of the early section highlights the abrupt early decay of standard and reduced SDNs, and how BRDF and pruned variants mitigate it. All plots pertain the medium-sized room detailed in Tab. 2.*



(a) *References*



(b) *2nd order SDNs*



(c) *3rd order SDNs*

Figure 6: *Early section of the Normalized Echo Densities for the compared methods. Note the Identity SDN's low final value and the Reduced SDNs' erratic growth. All plots pertain the medium-sized room detailed in Tab. 2.*

4.1.2. Normalized Echo Density

The Normalized Echo Density (NED) of a RIR is defined as the ratio, in a given time window, between the number of samples lying more than a standard deviation away from the mean in the RIR and that which is expected for Gaussian noise. This measure has been found to be strongly correlated to the perception of *texture* in RIRs [27]. In realistic scenarios, it is expected to rise steadily from 0 to 1, settling at the latter value; the earliest moment at which the NED reaches a value of 0.9 is conventionally interpreted as the so-called *mixing time*. If the mixing time occurs too early, it is a sign that the reverberator reached a state of diffuseness too quickly. If the NED fails to reach a value of 1, the reverberator does not yield sufficient echo density.

It has been shown [7] that using identity scattering matrices – or permutations of the identity matrix – in an SDN leads to insufficient echo density. This behaviour can be seen in Fig. 6a, for the *identity SDN* case.

Figs. 6b and 6c show the NEDs of the proposed methods with 2nd and 3rd order scattering nodes respectively. The *BRDF SDNs* and *pruned SDNs* cases show comparable behaviours to that of Ray Tracing – with only slightly different mixing times – suggesting good perceptual accuracy. On the other hand, the erratic behaviour of the *reduced SDN* cases indicates that merely removing the duplicated 1st order reflections is not enough, and higher-order duplicates (combined with the degenerate isotropic matrices) are also influential on the echo density's behaviour.

4.2. Perceptual tests

In order to evaluate and compare the finer perceptual qualities of the considered methods, a subjective experiment was carried out.

4.2.1. Experiment setup

Using a full factorial experiment design with the ten methods listed at the start of this section, the three rooms from Tab. 2, and three items of programme material – a gunshot, a music sample played on acoustic guitar, and a fragment of male speech – resulted in a total of 90 stimuli[4]. The experimental methodology was that of the Multiple Stimulus with Hidden Reference and Anchor (MUSHRA) method, specified in the ITU-R BS.1534-3 recommendation [28]. Sennheiser HD600 headphones were used for playback, and the stimuli were loudness-normalised to -36 LUFS. During a familiarisation phase, participants were instructed to set the headphone volume to a comfortable level, which was then kept constant. After the familiarization phase, which allowed subjects to listen to the entire range of possible stimuli, they were asked to rate how closely each sample corresponded to the given reference – the *Ray Tracing* method. The *Ray Tracing* method was chosen as reference as it is the state-of-the-art in geometric room acoustic models, and as previously stated the aim of the present work is to build a bridge between SDNs and the more accurate geometric methods. No low anchor was provided. Participants were instructed to give the maximum score to at least one sample on each comparison page, being informed that each page contained a hidden reference to be located. All of the 15 total participants were between 20 and 25 years of age and acquainted with audio processing; 12 were male, 3 female, none of them reported

[4]These stimuli, and the discussed RIRs, are available at the URL github.com/SCReAM-Surrey/DAFx2022-RIRs-and-Stimuli

any hearing impairments. Three participants failed to consistently give at least one high rating on each page, and were removed from the results. The remaining 12 participants followed the instructions correctly, each of them successfully locating the hidden reference within the tolerance defined by the MUSHRA standard (>85% of pages) [28]. Subjects were asked to assess how close individual stimuli were to the reference in terms of *naturalness*[5], which was defined as "the degree to which the stimuli conform to your experience of a sound in a room" [9].

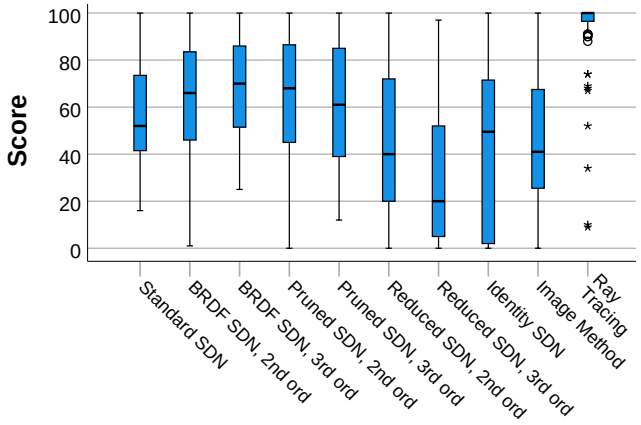


Figure 7: *Box plot of the naturalness scores for the compared methods, for all rooms and programme material. Medians shown by ticks, 25-75% quartiles by boxes, maxima and minima by whiskers. Outliers are present for Ray Tracing, these are cases in which subjects failed to rate the hidden reference with the top score.*

4.2.2. Experiment results

The results of Shapiro-Wilk tests showed that the scores' distributions were not normal, with significance values < 0.05 for all considered methods. For this reason, non-parametric tests were conducted. A Kruskal-Wallis H-test confirmed statistically significant difference between the different methods ($H(9) = 338.9$, $p < .001$).

Table 4: *Mean values and standard deviations for the naturalness ratings of the compared methods. Means report the 95% confidence interval.*

Method	Mean (95% conf.)	Std. Dev.
Standard SDN	58.06 ± 4.07	23.059
BRDF SDN, 2nd order	64.50 ± 4.26	22.917
BRDF SDN, 3rd order	68.56 ± 3.94	21.625
Reduced SDN, 2nd order	45.81 ± 5.37	28.354
Reduced SDN, 3rd order	30.17 ± 5.62	28.313
Pruned SDN, 2nd order	64.42 ± 4.82	26.561
Pruned SDN, 3rd order	62.25 ± 4.84	26.107
Identity SDN	42.32 ± 6.38	32.262
Image Method	46.43 ± 5.36	27.805
Ray Tracing	94.64 ± 3.04	18.461

[5]Ratings of *texture* were also elicited, but results were highly correlated with the *naturalness* attribute and are not presented for space reasons.

Table 4 shows the means and standard deviations of the scores for the compared methods, for all rooms and programme material. Figure 7 shows their box plot. All proposed extensions, except for the *reduced SDNs*, achieved higher mean ratings than the *standard SDN*. The *reduced SDNs* had lower ratings, with the 3rd order case scoring even lower than the *identity SDN*. The IM scored only marginally higher than the *identity SDN*, despite the absence of sweeping echoes. It should be noted that the relatively poor performance of IM may be due to the longer reverberation time mentioned earlier in Sec. 4.1.1 for the medium sized room.

Table 5: *Results of Mann-Whitney tests for several pairs of methods. The number of samples is* 108 *for all cases, meaning the Mann-Whitney U statistic should be interpreted as $U(108, 108)$. z is its standardized value.*

Methods	U	z	p
Standard SDN BRDF SDN 2nd order	338.9	-2.226	0.026
Standard SDN BRDF SDN 3rd order	4161.5	-3.638	< 0.001
BRDF SDN 2nd order BRDF SDN 3rd order	5272.5	-1.219	0.223
BRDF SDN 2nd order Pruned SDN 2nd order	5782.5	-0.108	0.914
BRDF SDN 3rd order Pruned SDN 3rd order	4987.0	-1.841	0.066

Mann-Whitney U-tests were conducted on some selected pairs of methods to gauge the statistical significance between them. The results of these tests are reported in Tab. 5.

Statistically significant differences are seen between the ratings of *standard SDN* and *BRDF SDNs* of both 2nd and 3rd order, confirming that increasing the number of accurate reflections has an impact on the perceived sound. The two *BRDF SDNs* fail to show a statistically significant difference between their ratings, posing a doubt on the usefulness of increasing the order above the 2nd.

Very high similarity can be seen from the significance result between the *2nd order pruned SDN* and the *2nd order BRDF SDN*, suggesting that the two methods are perceptually comparable. This is remarkable, considering that the latter presented approximately 6 times the number of connections with relation to the former – and thus had much higher computational complexity. The 3rd order counterparts of the same two methods fail to show the same similarity.

5. CONCLUSIONS

This paper investigated several extensions of SDN with the intent of improving its accuracy. Two higher-order extensions of SDNs were introduced; one that models directional scattering (BRDF SDN) and one that greatly reduces computational costs (Pruned SDN).

Objective and subjective measures indicated that (a) increasing the number of accurately modeled reflections brings a perceivable improvement in reverb naturalness; (b) the presented extensions raise the Standard SDN's accuracy by alleviating its issues with abrupt early energy decay; (c) the two presented extensions

were found to be perceptually comparable, despite one being much more lightweight than the other.

Future work will look further into possible design choices for directional scattering matrices (including frequency-dependent scattering), connection pruning criteria, and possible combination of the two methods.

6. ACKNOWLEDGMENTS

This work was supported by the EPSRC under the "SCalable Room Acoustics Modeling (SCReAM)" grant EP/V002554/1.

7. REFERENCES

[1] J. Catic, S. Santurette, and T. Dau, "The role of reverberation-related binaural cues in the externalization of speech," *J. Acoust. Soc. Am.*, vol. 138, no. 2, pp. 1154–1167, 2015.

[2] J. G. Apostolopoulos, P. A. Chou, B. Culbertson, T. Kalker, M. D. Trott, and S. Wee, "The road to immersive communication," *Proc. of the IEEE*, vol. 100, no. 4, pp. 974–990, 2012.

[3] M. R. Schroeder and B. F. Logan, ""colorless" artificial reverberation," *IRE Trans. on Audio*, , no. 6, pp. 209–214, 1961.

[4] M. R. Schroeder, "Natural sounding artificial reverberation," *J. Audio Eng. Soc.*, vol. 10, no. 3, pp. 219–223, 1961.

[5] V. Välimäki, J. D. Parker, L. Savioja, J. O. Smith, and J. S. Abel, "Fifty years of artificial reverberation," *IEEE Trans. on Audio, Speech, and Lang. Proc.*, vol. 20, no. 5, pp. 1421–1448, 2012.

[6] V. Välimäki, J. Parker, L. Savioja, J. O. Smith, and J. Abel, "More than 50 years of artificial reverberation," in *Proc. 60th Int. Conf. of the Audio Eng. Soc.*, 2016.

[7] E. De Sena, H. Hacıhabiboğlu, Z. Cvetković, and J. O. Smith, "Efficient synthesis of room acoustics via scattering delay networks," *IEEE/ACM Trans. on Audio, Speech, and Lang. Proc.*, vol. 23, no. 9, pp. 1478–1492, 2015.

[8] J. O. Smith, "A new approach to digital reverberation using closed waveguide networks," in *Int. Comp. Music Conf.*, 1985.

[9] S. Djordjevic, H. Hacıhabiboğlu, Z. Cvetkovic, and E. De Sena, "Evaluation of the perceived naturalness of artificial reverberation algorithms," in *148th Conv. of the Audio Eng. Soc.*, 2020.

[10] M. Geronazzo, J. Y. Tissieres, and S. Serafin, "A minimal personalization of dynamic binaural synthesis with mixed structural modeling and scattering delay networks," in *IEEE Int. Conf. on Acoust., Speech and Sig. Proc. (ICASSP-20)*, 2020.

[11] C. Yeoward, R. Shukla, R. Stewart, M. Sandler, and J. D. Reiss, "Real-time binaural room modelling for augmented reality applications," *J. Audio Eng. Soc.*, vol. 69, no. 11, pp. 818–833, 2021.

[12] A. Baldwin, S. Serafin, and C. Erkut, "Towards the design and evaluation of delay-based modeling of acoustic scenes in mobile augmented reality," in *IEEE 4th VR Workshop on Sonic Interactions for Virtual Environments (SIVE-18)*, 2018.

[13] S. Pekçetin, "Extended scattering delay networks incorporating arbitrarily shaped rooms and edge diffraction," M.S. thesis, Middle East Technical University, 2016.

[14] T. B. Atalay, Z. Sü Gül, E. De Sena, Z. Cvetkovic, and H. Hacıhabiboğlu, "Simulation of coupled volume acoustics with coupled volume scattering delay network models," *J. Acoust. Soc. Am.*, vol. 149, no. 4, pp. A117–A117, 2021.

[15] F. Stevens, D. T. Murphy, L. Savioja, and V. Välimäki, "Modeling sparsely reflecting outdoor acoustic scenes using the waveguide web," *IEEE/ACM Trans. on Audio, Speech, and Lang. Proc.*, vol. 25, no. 8, pp. 1566–1578, 2017.

[16] M. A. Gerzon, "Unitary (energy-preserving) multichannel networks with feedback," *Electronics Letters*, vol. 12, no. 11, pp. 278–279, 1976.

[17] S. J. Schlecht and E. Habets, "On lossless feedback delay networks," *IEEE Trans. on Sig. Proc.*, vol. 65, no. 6, pp. 1554–1564, 2016.

[18] Huan Mi, Gavin Kearney, and Helena Daffern, "Impact thresholds of parameters of binaural room impulse responses (brirs) on perceptual reverberation," *Appl. Sciences*, vol. 12, no. 6, pp. 2823, 2022.

[19] S. J. Schlecht and E. Habets, "Sign-agnostic matrix design for spatial artificial reverberation with feedback delay networks," in *133rd Conv. of the Audio Eng. Soc.*, 2018.

[20] D. Rocchesso, "The ball within the box: a sound-processing metaphor," *Comp. Music J.*, vol. 19, no. 4, pp. 47–57, 1995.

[21] H. Kuttruff, *Room acoustics*, Taylor & Francis Group, 2016.

[22] J. T. Beale, "Acoustic scattering from locally reacting surfaces," *Indiana University Mathematics Journal*, vol. 26, no. 2, pp. 199–222, 1977.

[23] H. Hacıhabiboğlu, E. De Sena, and Z. Cvetkovic, "Frequency-domain scattering delay networks for simulating room acoustics in virtual environments," in *Proc. of the Seventh IEEE Int. Conf. on Signal Image Tech. & Internet-Based Sys.*, 2011.

[24] R. Scheibler, E. Bezzam, and I. Dokmanić, "Pyroomacoustics: A python package for audio room simulation and array processing algorithms," in *IEEE Int. Conf. on Acoust., Speech and Sig. Proc. (ICASSP-18)*, 2018.

[25] Bengt-Inge Dalenbäck, "Acoustic / the fireverb suite / reflphinder / gratisvolver pro," Accessed: 15-Mar-2022.

[26] E. De Sena, N. Antonello, M. Moonen, and T. Van Waterschoot, "On the modeling of rectangular geometries in room acoustic simulations," *IEEE/ACM Trans. on Audio, Speech, and Lang. Proc.*, vol. 23, no. 4, pp. 774–786, 2015.

[27] J. S. Abel and P. Huang, "A simple, robust measure of reverberation echo density," in *121st Conv. of the Audio Eng. Soc.*, 2006.

[28] B. Series, "Method for the subjective assessment of intermediate quality level of audio systems," *Int. Telecom. Union Radiocomm. Assembly*, 2014.

[29] P. Huang, J. S. Abel, H. Terasawa, and J. Berger, "Reverberation echo density psychoacoustics," in *125th Conv. of the Audio Eng. Soc.*, 2008.

[30] J. Jot and A. Chaigne, "Digital delay networks for designing artificial reverberators," in *90th Conv. of the Audio Eng. Soc.*, 1991.