

Late Reverberation Synthesis: From Radiance Transfer to Feedback Delay Networks

Bai Hequn, Gael Richard, Laurent Daudet

▶ **To cite this version:**

Bai Hequn, Gael Richard, Laurent Daudet. Late Reverberation Synthesis: From Radiance Transfer to Feedback Delay Networks. 2015. hal-01142568

HAL Id: hal-01142568
https://hal.science/hal-01142568

Preprint submitted on 15 Apr 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Late Reverberation Synthesis: From Radiance Transfer to Feedback Delay Networks

Hequn Bai,[1*] Gaël Richard,[1] Laurent Daudet[2]

[1]Institut Mines-Télécom, Télécom ParisTech, CNRS-LTCI,
37-39 rue Dareau, 75014 Paris, France
[2]Institut Langevin, ESPCI ParisTech, Univ. Paris Diderot and CNRS UMR7587,
1 rue Jussieu, 75005 Paris, France

*To whom correspondence should be addressed; E-mail: hequn.bai@telecom-paristech.fr.

In room acoustic modeling, Feedback Delay Networks (FDN) are known to efficiently model late reverberation due to their capacity to generate exponentially decaying dense impulses. However, this method relies on a careful tuning of the different synthesis parameters, either estimated from a pre-recorded impulse response from the real acoustic scene, or set manually from experience. In this paper we present a new method, which still inherits the efficiency of the FDN structure, but aims at linking the parameters of the FDN directly to the geometry setting. This relation is achieved by studying the sound energy exchange between each delay line using the acoustic Radiance Transfer Method (RTM). Experimental results show that the late reverberation modeled by this method is in good agreement with the virtual geometry setting.

Introduction

The propagation of a sound within an enclosed space undergoes several reflections and diffraction at walls and obstacles. For a listener in this room, this creates a specific coloring, that depends on the position of both sound source and listener. In many application domains, there is a clear interest in

having efficient means to simulate perceptually, if not physically, this sound propagation. For example, in applications involving virtual reality (video games, serious games, movies...), it is of utmost importance to give to the listener the impression of being immersed in the acoustic scene, sometimes in a dynamic way (*1, 2*). In other applications, such as music recording in a studio, it may be desired to enrich the original acoustic space in order to simulate a larger, perceptually more pleasant stage.

The time-domain representation of such room-dependent sound propagation, between a sound source and a receiver, is classically characterized by the Room Impulse Response (RIR). Plenacoustic functions (*3–5*) extend this notion by representing the evolution of RIRs as a function of the three-dimensional spatial positions of source and receiver. Static acoustic scenes, neglecting their environmental variations such as temperature and humidity changes, can be considered as linear and time-invariant (LTI) systems, which can thus be uniquely characterized by their plenacoustic functions. Room impulse responses are classically decomposed into two parts, that provide distinct perceptual cues: the Early Reflections (ER) and the Late Reverberation (LR) (*6*). ER helps in localization and conveys spatial information about the environment, such as the shape of the room (*7*), while LR enhances immersion, contributing to the perception of the size of the environment, its amount of furnishing and type of absorptivity (*6*).

Many techniques have been proposed to compute RIRs in order to simulate the sound propagation within a room. They can be roughly divided into geometric and statistical approaches.

On the one hand, geometric approaches rely on physical principles. For example, the Image Source Method (ISM) (*8, 9*) computes specular reflection paths by considering a set of virtual sources, which are positioned at mirrored locations with respect to each wall. The ray tracing method (*10*) computes sound rays outgoing from the source through the acoustic scene, and collects them at the receiver. The beam tracing method (*11*) classifies reflection paths from a source by recursively tracing the pyramidal sound beams. A tree structure can be used to represent the beam path from the source to the full space enclosure (*12*). Due to the fact that computation time increases rapidly with the reflection order, these geometrical-based methods are mostly used to model the low-order echoes for early reflections. To

overcome this limitation, the Radiance Transfer Method (RTM), originating from the light transport algorithm, efficiently models the diffuse reflections of RIRs and the sound energy decay of the LR (*13, 14*). The RTM decouples the sound source and receiver from the sound propagation, by precomputing a linear operator that defines how a sound emitted from wall surface patches affects the radiance of other surfaces. For example, in (*15*), the so-called acoustic form factor is derived for diffuse surface reflections, in arbitrary polyhedral rooms. However, the main limitation of the RTM is its computational cost : depending on the complexity of the geometry, it may require hundreds to thousands of convolutions at run-time, which hinders real-time implementation even with alternative implementations in the frequency domain (*16–18*).

On the other hand, statistical methods (also termed perceptually-based methods) use global acoustic parameters, such as reverberation time, to model the main perceptual aspects of reverberation. In his landmark paper (*19*), Schroeder introduced a novel paradigm based on recursive comb filters and allpass filters to simulate the multiple echoes. This scheme was later improved by Moorer (*20*), and then generalized into a structure well suited for artificial reverberation called Feedback Delay Networks (FDNs)" (*21*). This structure is characterized by a set of delay lines connected in a feedback loop through a feedback matrix. Jot further studied the FDNs and developed associated techniques for designing good quality reverberators (*22*). Several extensions were proposed (*23*), including a binaural reverberator (*24*). Other approaches also exist, such as the recently proposed sparse velvet noise reverberator (*25*) and the artificial reverberator in the frequency domain to model the spectral magnitude decay (*26*). A comprehensive overview of other artificial reverberators is given in (*27*). It is however known that, given the structure of a statistical reverberator, it can be a rather tedious and empirical task to set all parameters (*28*), including the delay lengths, attenuation factors and feedback matrix. Besides, these statistical methods sometimes make use of measured impulse responses, extracting some global acoustic parameters to tune the artificial reverberator, thereby ignoring fine details of the actual room size/shape or material absorptivity. The work in (*29*) linked the physics-based digital waveguide mesh (DWM) model

with the feedback delay networks to design a scalable physics-based artificial reverberator. In a recent study (*30*), a genetic algorithm was used to automatically search an optimal setting of the parameters for a length-4 FDN, given the measured impulse responses.

The goal of this paper is to combine the two classes of approaches for the synthesis of room reverberation. First, we highlight some links between RTM and FDN before introducing a new geometry-based statistical method. Our approach inherits the computational efficiency of the FDN structure, but aims at linking the parameters of the FDN directly to the geometry setting. Such a relation is obtained by studying the sound energy exchange between each delay line, using the acoustic Radiance Transfer Method (RTM). In other words, the main principle of this hybrid approach, called FDN-RTM, is to relate a FDN statistical model to the physical room layout, without actual RIR measurements.

The paper is organized as follows. In section 1 we further describe the links between the RTM and FDN approaches and then recall the basics of the acoustic radiance transfer method and the main computation procedure. In section 2 we propose a new geometrical-based statistical modeling paradigm, and design the corresponding parameter estimation procedure. Experimental results are given in section 3. Finally, some conclusions and discussions are suggested in section 4.

1 Radiance Transfer Method

1.1 Links between RTM and FDN

It can be noticed that, in the radiance transfer method, the iteration of energy exchanges between surface patches can also be regarded as feedback loops in a feedback delay network, as illustrated on Fig. 1. Here, each delay line represents sound reflection on one surface patch, and the elements of the feedback matrix A contain both an attenuation factor a_{ij} and a time delay d_{ij}. The sound propagation between wall surface patches are represented as delay lines in feedback delay loops. It is then of practical interest to combine these two methods, using a feedback delay network as the synthesis structure, while feedback characteristics are estimated by the acoustic transfer property in the room. The proposed room geometry-

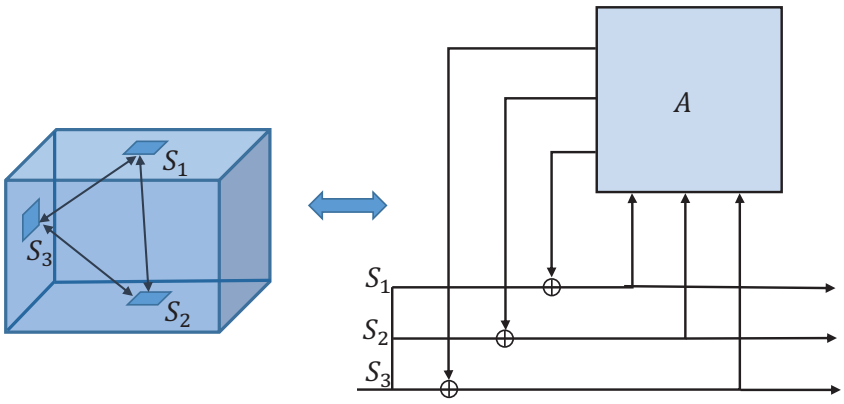


Figure 1: *Relation between radiance transfer method and feedback delay networks.*

based statistical method then uses delay lines to represent the group of diffuse reflections between room surfaces. For the FDN, the convolution of the room impulse response with the dry signal is replaced by $N^2 + kN$ multiplications and a few additions, where N is the number of delay lines. Thus, in another way, it can also be regarded as an approximated but computationally efficient structure of the radiance transfer method.

1.2 Principles of the radiance transfer method

Although some extensions to non-diffuse reflections exist (*31, 32*), the most commonly used radiance transfer method is based on the assumption that all boundaries are diffusely reflecting, governed by Lambert's law. Analytically, the radiation density of an infinitesimal wall element, x, is defined as its initial radiation density $I_0(x,t)$, plus the contribution of the radiation density from all other wall elements of the enclosure S at some earlier time $I(x', t - \frac{|x-x'|}{c})$, where c is the sound velocity in normal condition. The analytical acoustic radiance transfer model can be described as:

$$I(x,t) = I_0(x,t) + \int_S R(x,x',t) I(x', t - \frac{|x-x'|}{c}) \mathrm{d}x', \tag{1}$$

where $R(x,x',t)$ is the reflection kernel which describes how the outgoing radiation from point x' at time t influences the radiation density at point x after propagation.

To numerically simulate the radiation exchange, the room surface S is discretized into M small

planar patches. The radiation density of patch i after n diffuse reflections can be written in discretized form as:

$$I_i^{(n)}(t) = I_i^{(n-1)}(t) + \rho_i \sum_{j=1, j\neq i}^{M} F_{i,j} I_j^{(n-1)}(t - \frac{r_{i,j}}{c}). \tag{2}$$

where $r_{i,j}$ is the average distance between patch i and j, and $F_{i,j}$ is a form factor which is the integral of $f(x, x')$ over patch i and j, defining the fraction of energy leaving patch j and incident on patch i. The total radiation density from patch i can be written as:

$$I_i(t) = \sum_{n=0}^{\infty} I_i^{(n)}(t). \tag{3}$$

where $I_i^{(0)}(t)$ is the initial radiation density of patch i, distributed directly from the sound source.

The analytical expression of the form factor $F_{i,j}$ is derived by (*15*) as in Eqn. (4):

$$F_{i,j} = \frac{V_{i,j}}{A_j} \int_{S_i} \int_{S_j} \frac{\cos\theta \cos\theta'}{\pi d^2} \mathrm{d}S \mathrm{d}S', \tag{4}$$

where $V_{i,j}$ is the visibility factor between patch i and j, A_j is the area of patch j, d is the distance between the points of integration on S_i and S_j, as shown on Fig. 2. Together with the diffuse coefficient ρ and the patch-to-patch propagation delay, patch-to-patch responses are formed, which correspond to the reflection kernel $R(x, x', t)$.

At initial stage, the sound energy from the source is distributed to each surface patch, forming the so-called initial shooting matrix. Then, the sound energy is propagated within the enclosure and is exchanged between surface patches. After high-order reflections, the resulting responses are gathered from each surface patch to the listener, using the final gathering matrix. It is worth noting that the high-order reflection response is independent from the positions of the source and the listener. This property makes this method suitable for dynamic scene rendering, when both source and listener move, as only initial shooting and final gathering matrices require a real-time update.

Several studies have shown that the radiance transfer method is effective in predicting the reverberation characteristics of a given room. In (*15*), predictions were made in three spheric enclosures of

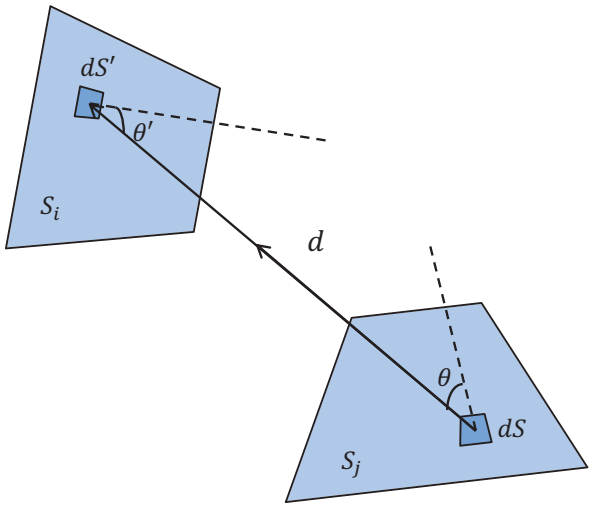


Figure 2: *Illustration of the form factor for diffuse reflections.*

varying size and absorptivity. Prediction results show that the radiance transfer method simulates impulse responses with almost the same acoustic characteristic values, including reverberation time (RT), steady-state sound-pressure and the average radiation density, compared with analytical solutions (*33*). In (*14*), further investigations were made by comparing with real measured data in a squash court, a classroom and an office. Experimental results have shown that the purely diffused radiance transfer method can predict the room sound field with good accuracy, and, in general, better than the prediction made by purely specular ray tracing methods. In the rest of this paper, we therefore take RTM as the baseline method for modeling late reverberation.

2 Geometry-Based Statistical Model

In this section, a new geometrical-based statistical model is presented, named FDN-RTM. The name of the model comes from two parts. In the one hand, it borrows the feedback delay networks (FDN) structure for its efficient implementation and flexible parameter settings. In the other hand, the order-1 patch-to-patch responses from the radiance transfer method are used to estimate the parameters of the FDN structure, in order to synthesize a realistic impulse response as expected from the specified room

geometry.

2.1 Relation between FDN and RTM

The basic idea of radiance transfer method is to divide the room surface into patches, and to model how the outgoing radiance from one patch will affect the outgoing radiance of all the other patches after order-1 diffuse reflection. High order reflections can also be achieved. This is similar to the structure of feedback delay networks where the signal from one delay line will be sent to other delay lines according to the feedback matrix, in order to increase the energy mixture between them. If we replace each patch-to-patch interaction by a delay line, and set the feedback matrix with the corresponding form factors, we can, theoretically, construct a huge FDN with hundreds to thousands of delay lines, which then would be equivalent to the radiance transfer model.

However, in most cases of late reverberation, the trend of the energy decay is sufficient to perceive the dimension and layout of the environment. Thus, we can ignore the precise model of the delay and amplitude of each pulse in the room impulse response, and only focus on modeling the global energy exchange statistics. This suggests that a simplified RTM model can be used, by grouping the patch-to-patch energy interactions, in order to reduce the number of delay lines. The parameter of each delay line is determined by the average value of the patch-to-patch delays of the corresponding group and the accumulated surface absorption factors.

The range and distribution of the delay lengths can also guide the choice of delay units in the FDN structure. To illustrate this, we have computed the distribution of occurrence and energy for different delays of the non-convex room displayed on Fig. 3 - Room 1. This room is modeled using 5000 patches. The distribution of occurrence and energy is shown in Fig. 4. Note that the distribution of energy for different delays is the product of the distribution of their occurrence with the energy transported by the corresponding delays. Although the occurrence of short delays is low, patch-to-patch sound

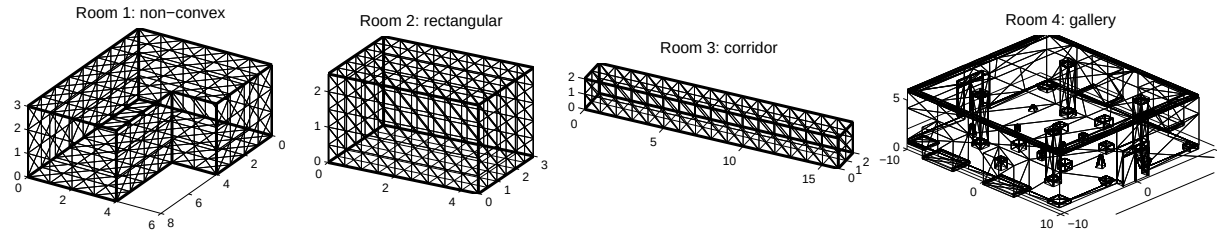


Figure 3: *The experimental geometries. Room 1: non-convex room of outside dimension* $8m \times 6m \times 3m$*. Room 2: rectangular room of dimension* $4.5m \times 3m \times 2.5m$*. Room 3: corridor of dimension* $16m \times 2m \times 2m$*. Room 4: gallery of outside dimension* $19.6m \times 17.7m \times 5.1m$*. It should be noted that the triangular patches do not illustrate the real patching density during the computation of the RTM algorithm.*

interactions with short delays convey important energy, mainly because short delays normally come from neighboring patches and transport considerably more energy than longer delays, according to Eqn. (4). A closer look on Fig. 4 reveals that both the occurrence distribution and the energy distribution have small peaks at around 390, 540, 800 and 1050 sample delays, which correspond to distances of 3m, 4m, 6m and 8m, respectively. Considering the dimensions of the room, these peaks reveal some of its acoustic modes. These distributions can be exploited to appropriately choose delay units which are related to the dimension and shape of the room. Optimal clustering algorithms (such as K-means) can also be used to find the optimal quantization of the delays based on their occurrence distribution. In some special cases the modes are even more easily observed, such as for the long corridor room displayed on Fig. 3 - Room 3. The delay distribution, as shown in Fig. 5, has a strong peak at around 250 delay samples, which corresponds to 2m, the width of the corridor. In such cases, the delays can be chosen around the strong peak values to model the acoustic modes.

2.2 System structure

The proposed reverberator is illustrated in Fig. 6. It consists of a FDN-RTM for late reverberation, as shown in the upper frame, and a series of early reflections, as shown in the lower frame.

The structure of the FDN-RTM is similar to the feedback delay network proposed in (*34*), except that

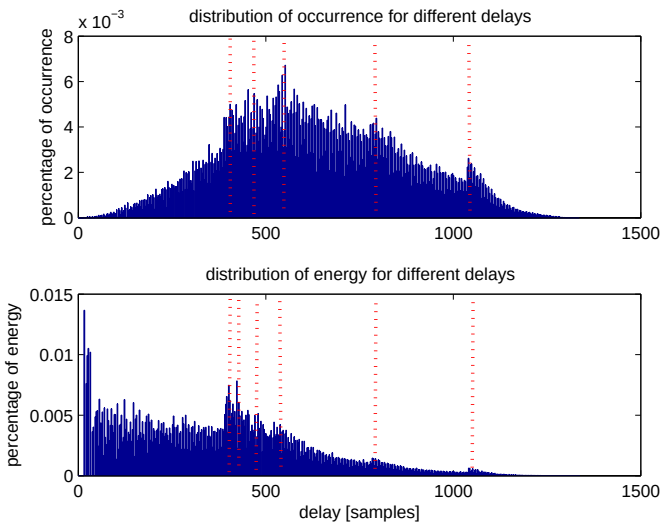


Figure 4: *Distribution of occurrence and energy for different patch-to-patch delays due to diffuse reflections for Room 1 (non-convex room). Top: histogram of occurrence. Bottom: histogram of energy. The dotted lines illustrate some peaks of the distributions. The sampling rate is* $44.1kHz$.

for each delay line, an initial delay unit and a post delay unit are added before and after the recursive circuit, respectively. Each delay line represents a group of order-1 patch-to-patch sound reflections, where the delay unit is determined by the mean patch-to-patch delay of each group. The initial delay unit is the mean delay from the sound source to the grouped patches, and the post delay unit is the mean delay from the grouped patches to the receiver. Since the dry input sound has to pass through the initial delay, the post delay and at least once through the patch-to-patch delays, the output reverberant sound actually models sound reflections starting from the second order. Thus, the first order sound reflections have to be modeled and added using the FIR filter in the lower frame.

Let $\wp_n$ denote the set of patch-to-patch sound interactions in group n. Then, if the total number of groups is N, the corresponding feedback delay network will have N delay lines as depicted on Fig. 6.

2.3 Parameter estimation

The parameters of proposed FDN-RTM are estimated as follows using the order-1 patch-to-patch interactions in the radiance transfer model.

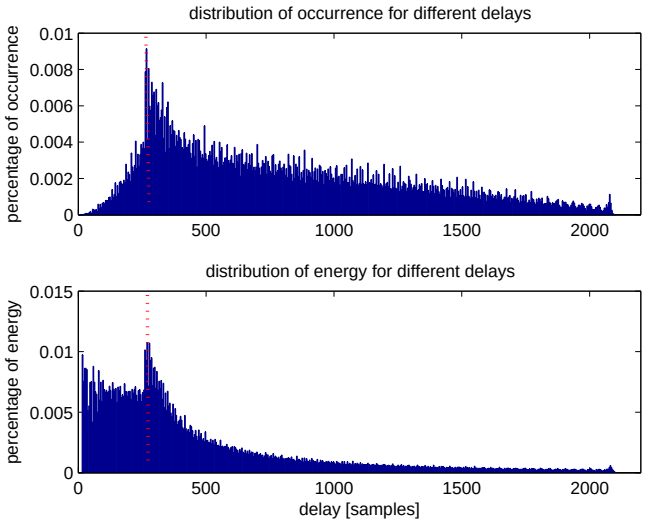


Figure 5: *Distribution of occurrence and energy for different patch-to-patch delays due to diffuse reflections for Room 3 (long corridor of dimension* $16m \times 2m \times 2m$*). Top: histogram of occurrence. Bottom: histogram of energy. The dotted lines highlight the main peak of the distributions. The sampling rate is* $44.1kHz$*.*

1. Feedback matrix

The element a_{mn} in the feedback matrix actually describes the proportion of energy which is transported by the patch-to-patch interactions within group m, that will be diffusely reflected and go to some other patch-to-patch interactions in group n, during an order-1 diffuse reflection.

Let us denote as ℓ_m the total energy transported by the patch-to-patch energy exchange in group m, and $\ell_{m,n}$ the total energy received by group n because of the diffuse reflections of group m. They can be estimated from the form factors $F_{i,j}$ in the RTM model as:

$$\ell_m = \sum_{i \to j \in \wp_m} F_{i,j} \tag{5}$$

$$\ell_{m,n} = \sum_{i \to j \in \wp_m} \sum_{j \to k \in \wp_n} F_{i,j} F_{j,k} \tag{6}$$

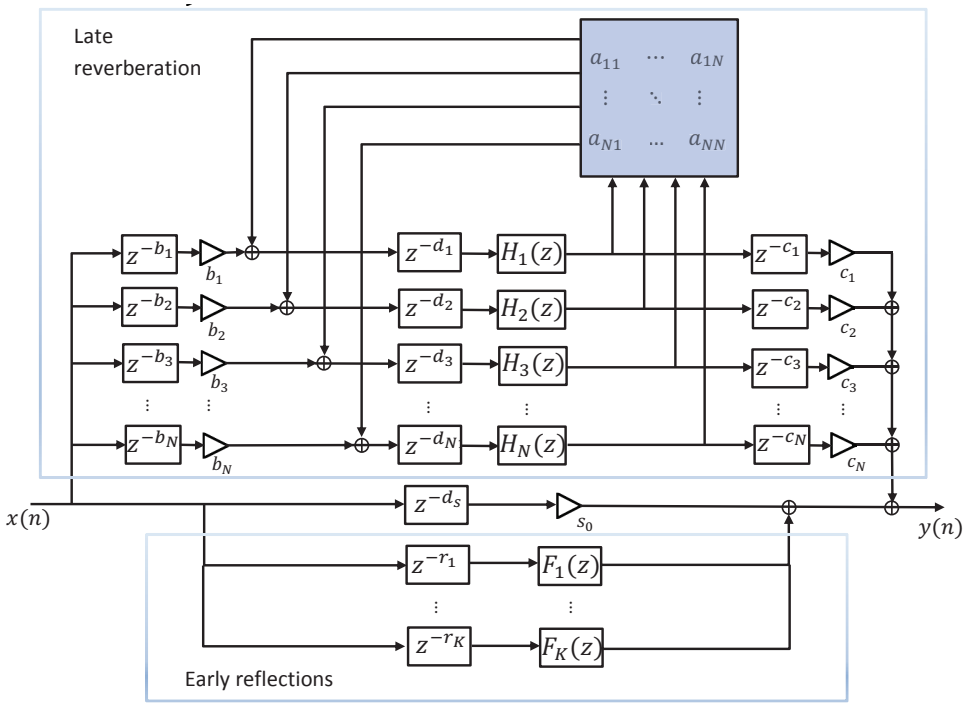


Figure 6: *The system structure of the FDN-RTM method.*

The feedback coefficient a_{mn} is defined as:

$$a_{mn} = \frac{\ell_{m,n}}{\ell_m} \tag{7}$$

$$\text{subject to } \sum_{n=1}^{N} a_{mn} = 1. \tag{8}$$

According to the energy conservation principle, Eqn. (8) will be satisfied if the form factors are computed with ideal precision. Using the form-factor calculation in (*15*), the curved surface integration is simplified to planar surface integration where the mean distance and angle are used as approximation. When using such approximated form factors, each row only approximately sums to 1.

Note that the radiance transfer method models the propagation of the sound energy within an enclosure, and thus is based on sound intensity. The feedback coefficients estimated by the radiance transfer method are also energy-based. Since the intended input of a FDN is the sound pressure of the source signal and the intended output is the sound pressure of the reverberant signal, FDN parameters must be

designed to process sound pressure. With unit outgoing energy from group $\wp_m$, the outgoing energy of group $\wp_n$ due to the radiation from $\wp_m$ is a_{mn}. In real sound propagation this amount of energy a_{mn} may contains multiple pulses. In the FDN-RTM model, this group of pulses are represented by a single pulse with sound pressure of $\sqrt{a_{mn}}$, which contains the equal total sound energy. Thus, in order to synthesize the sound pressure impulse response from the FDN-RTM structure with equivalent energy, the elements in the feedback matrix need to be set to the square-root of the energy-based values.

$$A = \begin{pmatrix} \sqrt{a_{11}} & \cdots & \sqrt{a_{1N}} \\ \vdots & \ddots & \vdots \\ \sqrt{a_{N1}} & \cdots & \sqrt{a_{NN}} \end{pmatrix}. \tag{9}$$

The signs of these entries can then be chosen according to the procedure described in section 2.5.

2. Attenuation filters

The attenuation filters $H_n(z)$ in traditional FDNs are low-pass filters, which model the frequency-dependent absorption of the surface material and air (*35*). In the case of FDN-RTM, the attenuations are energy-weighted mean reflection coefficients of all the emitting patches within the group:

$$\bar{\theta}_n^{\,s} = \sum_{i \to j \in \wp_n} F_{i,j} \theta_i^{\,s} \tag{10}$$

where $\bar{\theta}_n^{\,s}$ is the mean reflection coefficient for frequency band s of group n, $\theta_i^{\,s}$ is the material reflection coefficient of patch i for frequency band s, and $F_{i,j}$ the corresponding form factor.

In the Sabine formula the concept of absorption area is the equivalent absorption weighted by the area of each material. In Eqn. (10), the absorption area of each patch is also expressed, but is hidden in the form factor, where large patches emit more energy than small patches and thus their absorption coefficients take more importance in the summation.

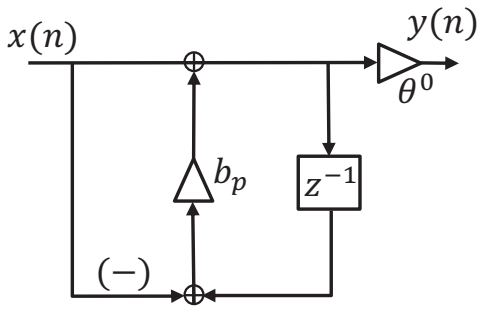


Figure 7: *Implementation of first order low-pass filter.*

For the sake of simplicity, the frequency-dependent absorption of the surface material is modeled here as a first order low-pass filter, as shown in Fig. 7. Higher order filters can be used if better quality is required.

The frequency response of the first order low-pass filter has a ratio $(1 - b_p)/(1 + b_p)$ between zero and Nyquist frequencies. Thus, the factor b_p can be estimated using the mean reflection coefficients of materials at zero and Nyquist frequencies as:

$$\frac{\bar{\theta}_n^{fs}}{\bar{\theta}_n^0} = \frac{1 - b_p}{1 + b_p}. \tag{11}$$

However, acoustic absorptions of materials are often only available from 125 Hz until 4000 Hz or 8000 Hz. To determine the low-pass filter, the frequency-dependent gain of the filter in Fig. 7 is used:

$$\|H(e^{j\omega})\|^2 = \frac{(1 - b_p)^2}{1 + bp^2 - 2b_p cos(\omega)} \tag{12}$$

Then, the absorptions coefficients at 125 Hz and 4000 Hz or 8000 Hz can be used to determine the low-pass filter coefficient b_p.

3. Delay units

In FDN-based artificial reverberators, the delay lengths are usually chosen as mutual prime numbers around the mean free path given by:

$$\bar{d} = 4V/S \tag{13}$$

where V is the total volume of the room, and S is total surface area enclosing the room. In (*36*) the delays of delay units are also suggested to be distributed over a ratio of $1 : 1.5$.

For the FDN-RTM, the mean delay length $\bar{d}_n$ for the n^{th} delay line is calculated as the energy-weighted mean delay length for all the patch-to-patch delays within that group:

$$\bar{d}_n = \frac{\sum_{i \to j \in \wp_n} F_{i,j} d_{i,j}}{\sum_{i \to j \in \wp_n} F_{i,j}}. \tag{14}$$

The delay lengths calculated using Eqn. (14) represent the true energy decay rate for each group, but may not be the best combination for sound quality. For good sound synthesis quality, it is desirable to adjust the delay lengths to be mutual prime numbers. For example, if we want to model the frequency modes shown in Fig. 4, some of the delay units can be chosen as prime numbers around the peaks of the delay distributions. Alternatively, we can choose the delay values that minimize the average error to represent the delay distribution. In order to keep the energy decaying rate unchanged, the attenuation of the delay line need to be modified accordingly when its length is adjusted:

$$\theta' = \bar{\theta}^{d'/\bar{d}}, \tag{15}$$

where $\bar{d}$ is the mean delay length for a certain delay line calculated by Eqn. (14), $\bar{\theta}$ is the mean attenuation by order-1 reflection, d' is the adjusted delay length and θ' is the attenuation factor that need to be modified accordingly.

4. Pre-mixing and post-mixing coefficients

We name $z^{-b_1}, ..., z^{-b_4}$ the pre-mixing delays, and $b_1, ..., b_4$ the gains in Fig. 6, and relate them to the initial shooting matrix used in the radiance transfer model. The sound starting from the source is first

reflected at the boundary into the designed groups. The delays $z^{-b_1}, ..., z^{-b_4}$ are chosen as the average integer number of such initial delays. The coefficients $b_1, ..., b_4$ are the amounts of energy emitted from the source which go to each group after the initial shooting.

The post-mixing delays $z^{-c_1}, ..., z^{-c_4}$ and gains $c_1, ..., c_4$ are estimated from the final gathering matrix in a similar way.

2.4 Stability of the feedback delay networks

Although the parameters estimated above satisfy the energy-conservation principle, the feedback loops in the feedback delay networks impose further constraints to keep the system stable.

Let $G = \Theta A$ where A is the feedback matrix and Θ is the attenuation coefficient matrix. Using a matrix form, the system equation of Fig. 6 becomes

$$Y = D(X + GY) \tag{16}$$

where D is the delay matrix

$$D = \begin{pmatrix} z^{-d_1} & & & 0 \\ & z^{-d_2} & & \\ & & \ddots & \\ 0 & & & z^{-d_N} \end{pmatrix}. \tag{17}$$

The iterative feedback given in Eqn. (16) can be rewritten as

$$\begin{aligned} Y &= DX + DGDX + DGDGDX + \ldots, \\ &= DX + D(GD)X + D(GD)^2X + D(GD)^3X + \ldots. \end{aligned} \tag{18}$$

If we decompose G using an eigenvalue decomposition,

$$G = P\Lambda P', \tag{19}$$

where the diagonal elements of Λ are the eigenvalues, and P is the eigenmatrix, the high order reflections

in Eqn. (18) are expressed as

$$\begin{aligned} D(GD)^n X &= D(P\Lambda P' D)^n X \\ &= \underline{DP}\ \Lambda\ \underline{P'DP}\ \Lambda\ \underline{P'DP}\ \Lambda\ \underline{P'D}\ \ldots\ X \end{aligned} \tag{20}$$

Since the product of the delay matrix D, the eigenmatrix P and its inverse (underlined in Eqn. (20))do not change the energy of the reflected sound, in order to make the system stable, it is sufficient to keep the largest eigenvalue in Λ smaller than 1. Note that the Λ gathers the eigenvalues of the product of the feedback matrix and the reflection coefficients matrix.

The non-negative matrix A, estimated using the radiance transfer method, has its dominant eigenvalue usually much larger than 1, which contains the DC component of its non-negative entries. Therefore, the square-root matrix cannot be directly used as feedback matrix. It is known that real unitary matrices, whose columns are mutually orthogonal, have their eigenvalues lying on the unit circle and thus are feasible feedback matrices for lossless feedback systems. Switching the signs of certain entries ($+1$ and -1) does not alter the energy of the system, but introduces phase inversions of the room impulse response. Therefore, by intentionally designing the grouping scheme and switching the signs of some entries as below, near orthogonal matrices are expected to be found for feedback delay networks.

2.5 Even-energy grouping scheme

The grouping scheme can be designed in different ways, with various factors under consideration. This section gives an example of a grouping scheme with even energy for each group. Such a grouping scheme gives near uniform-entry matrices which can be transformed to mutually orthogonal Hadamard matrices.

Scaled Hadamard matrices are widely used in feedback delay networks. Their entries are either $+1/\sqrt{N}$ or $-1/\sqrt{N}$, where N is the number of delay lines. The eigenvectors of such Hadamard matrices all lie on the unit circle and the absolute value of the eigenvalues are constant and equal to 1.

There are several ways to construct Hadamard matrices. If $H_{2^{k-1}}$ is a Hadamard matrix of dimension $2^{k-1} \times 2^{k-1}$, then the Hadamard matrix of dimension $2^k \times 2^k$ can be constructed from $H_{2^{k-1}}$ as:

$$H_{2^k} = \begin{pmatrix} H_{2^{k-1}} & H_{2^{k-1}} \\ \\ H_{2^{k-1}} & -H_{2^{k-1}} \end{pmatrix}. \tag{21}$$

Since the form factors contains several hidden geometry information in theirs values (for example, the patch size, distance to patches and angles), patch-to-patch interactions with similar form factors are likely to be obtained from neighboring or symmetric positions in the room. Also, because of the isotropic scattering property of each patch, form factors with neighboring or symmetric positions are likely to have similar scattering behaviors to other patches. Thus, in order to obtain a uniform-entry feedback matrix, we design a grouping scheme which assigns form factors with similar scattering behaviors to different groups. Their similarity in scattering behavior is indicated by the energy that they transport. The detailed procedure is described below. Since it is not a closed-form solution, the grouping scheme listed here is just an approach to obtain an approximated mutually orthogonal matrix.

The form factors are first sorted into descending order by their transported energy. We assume that the sorted energy is nearly linearly descending with the same decreasing step. We take the first N form factors and assign each of them to different groups, where N is the order of the desired feedback matrix. Then for the $N+1$ to $2N$ form factors, we do the same, assigning each of them to different group, but in a circular order. An example of the assigning order for $N = 4$ is shown in Fig. 8.

A grouping scheme designed in this way guarantees that both the total transmitted energy in each group and the scattered energy to the other groups are almost evenly distributed, which results in a near uniform-entry feedback matrix. Such a matrix is then square-rooted and the signs of the entries are switched according to the signs of the Hadamard matrix. The resulting matrix is an approximation of the Hadamard matrix, which makes the feedback system stable and lossless, while still related to the main physical phenomena.

2.6 FDN-RTM algorithm

The different steps of FDN-RTM algorithm are summarized in Fig. 9:

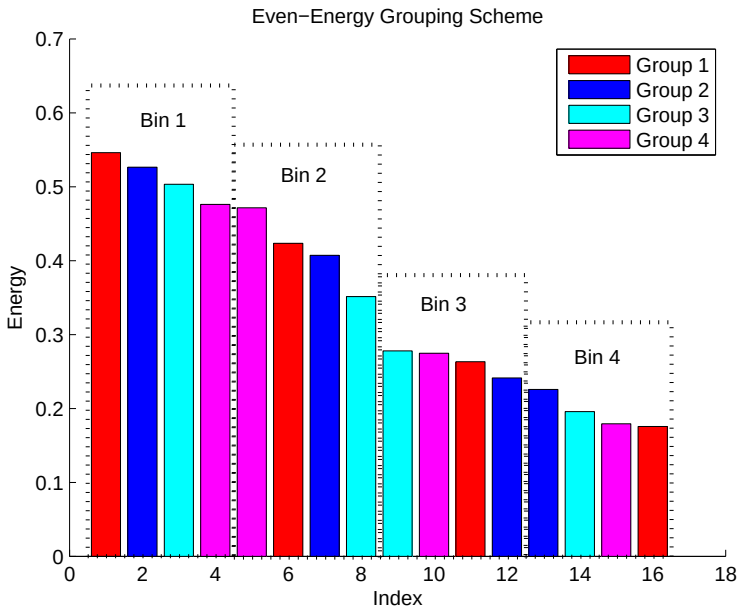


Figure 8: *Even-energy group assignment scheme for FDN-RTM with four delay lines. The samples with the same color denote that they are assigned to the same group.*

1. Decompose the room geometry into patches.

2. Calculate the diffuse form factors for all patch-to-patch reflections using Eqn. (4). Calculate the distribution of the patch-to-patch delays, and choose the delay units around the distribution peaks, or quantized delay values which minimize the average error.

3. Sort the form factors based on the transported energy. Assign them to N groups as shown in Fig. 8.

4. Estimate the FDN-RTM parameters as detailed in Section 2.3: the feedback matrix using Eqns. (5-8), the attenuation factors for each delay line using Eqn. (10), the low-pass filter coefficients using Eqn. (11) or (12), and the average delay length using Eqn. (14).

5. Square-root the feedback matrix and switch the signs of the entries according to the signs of the Hadamard matrix.

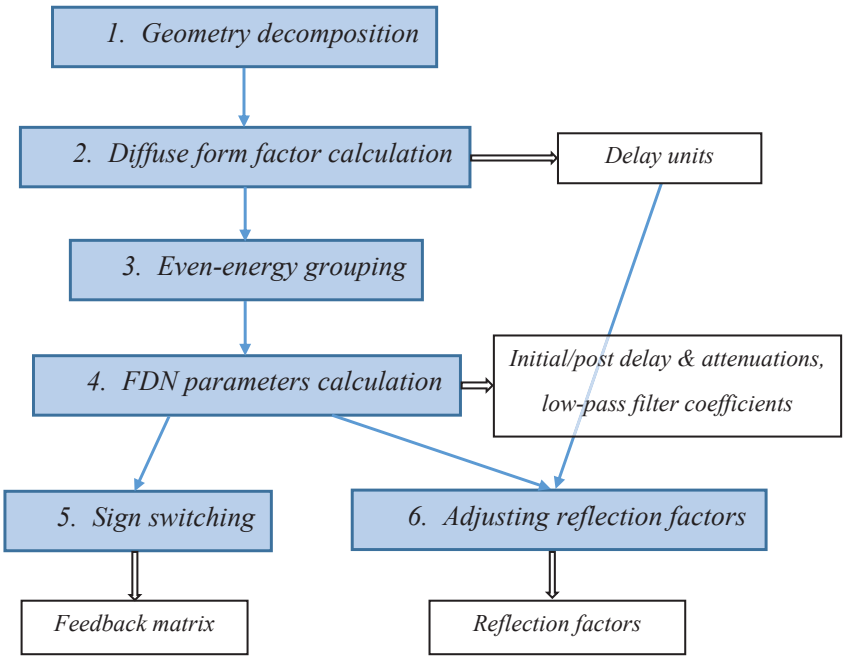


Figure 9: *Process of finding the parameters of the FDN-RTM algorithm.*

6. Using Eqn. (15), adjust the attenuation factors of each delay line based on the average delay length (in step 4) and the chosen delay units (in step 2).

3 Experimental Results

3.1 Experiment setup

We use both the proposed FDN-RTM and the traditional radiance transfer method, taken as reference, to model the diffuse late reverberation of a set of 4 rooms with different shapes and dimensions. The experimental geometries are displayed on Fig. 3.

Room 1 is a non-convex room with dimension $8m \times 6m \times 3m$ and uniform reflection coefficients are 0.9. Room 2 is a rectangular room with dimension $4.5m \times 3m \times 2.5m$. Room 3 is a long corridor with dimension $16m \times 2m \times 2m$. Room 4 is a more complex gallery geometry. The main shape of the gallery is a rectangular room with dimension $19.6m \times 17.7m \times 5.1m$, with four pillars and a few small box-shaped or cones-shaped obstacles on the floor. The triangular mesh of the geometry is first simplified

to 325 triangles in order to remove the fine details of the architecture, and then decomposed into 435 patches with variable size for the computation of the patch-to-patch interactions. The wall reflection factors of the four rooms are shown in Table 1.

Table 1: *Wall reflection coefficients settings of the experimental geometries.*

Room index	Room size	floor	ceiling	wall
1. Non-convex	$8m \times 6m \times 3m$	0.9		
2. Rectangular	$4.5m \times 3m \times 2.5m$	0.4	0.8	0.7
3. Corridor	$16m \times 2m \times 2m$	0.9		
4. Gallery	$19.6m \times 17.7m \times 5.1m$	0.75		

3.2 Simulation results of Room 1

We use Room 1, the non-convex room, as an example to show the parameter estimation process for a FDN-RTM of order 4 and its modeling results. The source and the receiver are both assumed to be located at position $[3, 3, 1.5]$. The surfaces are divided into 164 patches, with uniform patch size of $1m^2$.

The delay units are chosen as the prime number around the peaks of the delay distribution. The feedback matrix is estimated using the even-energy grouping scheme, and is near orthogonal with eigenvalues of amplitudes around 1 ($0.997, 1.003, 1.001, 0.999$).

Fig. 10 shows the synthesized reverberation (normalized) for room 1 using 4 delay lines. Thanks to the feedback structure, the reflection density of simulated echoes increases rapidly as time increases, which gives sufficient echo density for late reverberation. The amplitudes of the echoes decrease smoothly with an exponential rate. Listening tests of synthesized signal for speech and music samples shows satisfying sound quality with smooth reverberant sound.

3.3 Comparison with radiance transfer method

The parameters of FDN-RTM are estimated from the Radiance Transfer Model. Thus, it is of interest to see if the FDN-RTM models the energy decay in a similar way as the Radiance Transfer Method. We

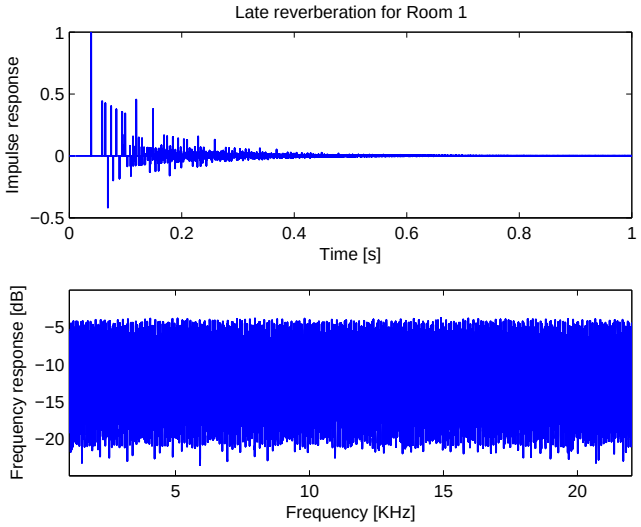


Figure 10: *Room impulse response synthesized for Room 1 with FDN-RTM using four delay lines.*

use the energy decay curve to compare these two methods.

Fig. 11 shows the energy decay curves of the impulse responses synthesized using FDN-RTM and RTM for the four rooms. It is shown that the impulse responses estimated by FDN-RTM is exponentially decreasing, which is shown as linearly decreasing on a logarithmic scale (dB). This is due to the fact that the estimated feedback matrix is near unitary and system poles are of near unit amplitude. The decay rates of FDN-RTM agree well with those simulated using RTM, which shows that FDN-RTM can be used as an alternative method to RTM in application cases where the late reverberation is assumed diffusely reflected and only the general decay rate is of interest to the perception.

Table 2 compares the reverberation time T_{60} estimated by the RTM (reference), FDN-RTM and Sabine formula for the four rooms.

The reverberation times estimated by FDN-RTM agree well with those estimated by RTM and Sabine formula[1]. However, the reverberation times estimated by FDN-RTM are slightly longer. The main reason is that the feedback matrix estimated using even-energy grouping scheme has some eigenvalues

[1]The T_{60} estimated by Sabine formula for Room 4 is an approximated value where the volume of the room is simplified by considering only the volume of the main shoe-box, the four pillars and the windows.

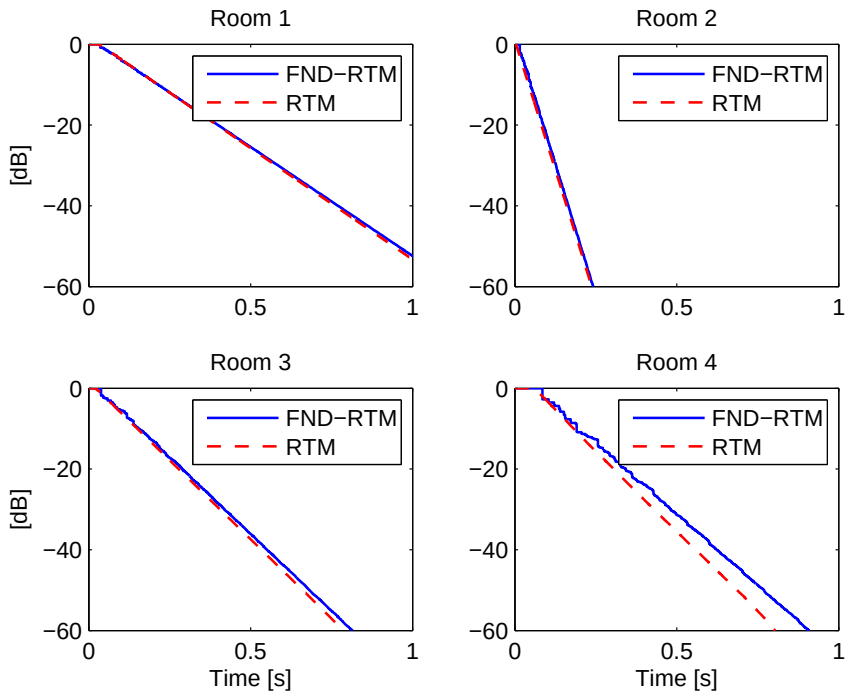


Figure 11: *Energy decay curves of the impulse responses synthesized using FDN-RTM and RTM. The x-axis is kept within $[0, 1]s$ and the y-axis is within the range of $[0, -60]dB$.*

slightly higher than the others, which makes the corresponding delay lines decay slower. For the four rooms, both RTM and FDN-RTM predict similar T_{60} with limited deviation from those predicted by the Sabine formula. While achieving comparable simulation results, FDN-RTM requires a much smaller computation time than the RTM during the pre-computation stage and the real-time rendering. Table 3 compares the computation times for both methods on a standard PC.

Here, the pre-computation time for the RTM includes the decomposition of the geometry, the calculation of the form factors and the computation of high order patch-to-patch responses. The rendering for the RTM includes the generation of the room impulse response and the convolution of the RIR with the dry signal. The pre-computation for FDN-RTM method includes the decomposition of the geometry, the calculation of form factors, the even-energy grouping scheme and the parameters estimation. The rendering for FDN-RTM method includes the feedback loops of the input signal, which requires $N^2 + kN$

Table 2: *Prediction of reverberation time using the RTM, FDN-RTM and Sabine formula. The reverberation times are in seconds, and the errors are in percentage.*

Room	Sabine (s)	RTM (s)	**FDN-RTM (s)**	Error to Sabine (%)	Error to RTM (%)
1	1.178	1.078	**1.110**	5.8	2.9
2	0.246	0.217	**0.228**	7.3	3.2
3	0.758	0.743	**0.785**	1.0	5.7
4	0.861	0.778	**0.849**	1.4	9.1

Table 3: *Computation time consumed by RTM and FDN-RTM (4 delay lines) to compute the late reverberation of a one-second room impulse response. The machine used for this simulation has a 2.8GHz CPU and 8G RAM.*

	RTM			**FDN-RTM**	
Room	Number of patches	Pre-comp.	Render (s)	Pre-comp. (s)	**Render** (s)
1	164	1 h 18 min	30.4	1.93	**0.07**
2	64	13 min	7.1	0.42	**0.09**
3	68	14 min	8.0	0.46	**0.09**
4	435	14 h 51 min	85.8	13.34	**0.09**

multiplications and some additions/delays during the feedback loops, as shown in Fig. 6, where $N = 4$ is the number of delay lines in this simulation.

To render a one-second signal, FDN-RTM uses less than 100ms during run-time, $2 - 3$ order of magnitudes less than needed by the RTM. This makes this geometry-based reverberator suitable for real-time processing.

3.4 Order of delay lines

The FDN-RTM can be regarded as a simplified and efficient implementation of the RTM using only a few delay lines, while the original RTM uses several hundreds to thousands of patch-to-patch interactions. By increasing the number of delay lines, the sound quality should be increased, at the price of a higher computation load. The computation time for FDN-RTM using 4 to 16 delay lines is shown in Table 4. It can be seen that the computation time increases almost linearly with the number of delay lines.

Table 4: *Computation time consumed by FND-RTM using 4, 8 and 16 delay lines to compute the late reverberation of a one-second room impulse response.*

Number of delay lines	4	8	16
Computation time [s]	0.078	0.187	0.476

Informal listening tests[2] using FDN-RTM for speech, singing voice and drum signals reveal that, with the increase of delay lines, the sound quality has noticeable improvement, especially for drum and speech signals. The light metallic sound for $N = 4$ with high reflection coefficient ($\theta > 0.9$) is reduced when $N = 16$, with a smoother reverberant sound.

3.5 Frequency-dependent absorptivity

Instead of using unified absorption coefficients, here the walls of Room 1 are assigned with frequency-dependent absorption materials, as shown in Table 5.

[2]The sound examples can be found at `http://perso.telecom-paristech.fr/~hbai/demo_ASLP2015/`.

Table 5: *Material absorption coefficients from* $125Hz$ *to* $4000Hz$ *for Room* 1.

Surfaces	Material	$125Hz$	$250Hz$	$500Hz$	$1000Hz$	$2000Hz$	$4000Hz$
Floor	Carpet on concrete	0.02	0.06	0.14	0.37	0.60	0.65
Ceiling	Plaster	0.02	0.03	0.04	0.05	0.06	0.08
Wall (type 1)	Curtain 1 and concrete	0.14	0.35	0.55	0.72	0.70	0.65
Wall (type 2)	Curtain 2 and concrete	0.03	0.04	0.11	0.17	0.24	0.35

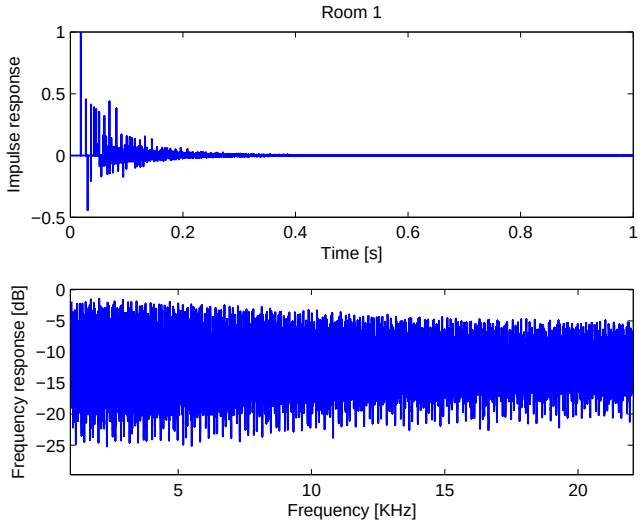


Figure 12: *Impulse response of frequency-dependent absorption.*

The absorption coefficients at $125Hz$ and $4000Hz$ are combined to determine the low-pass filter coefficient b_p. The synthesized room impulse response with frequency low-pass character is shown in Fig. 12.

The frequency-dependent reverberation is well modeled, as seen on the energy decay shown in Fig. 13, where the T_{60} at zero frequency is about 1.15s, while at $22.05kHz$ it is only 0.4s. Still, the first-order low-pass filters cannot model precisely the complex frequency-dependency absorption of real rooms. Higher order IIR filters can be used for more accurate results.

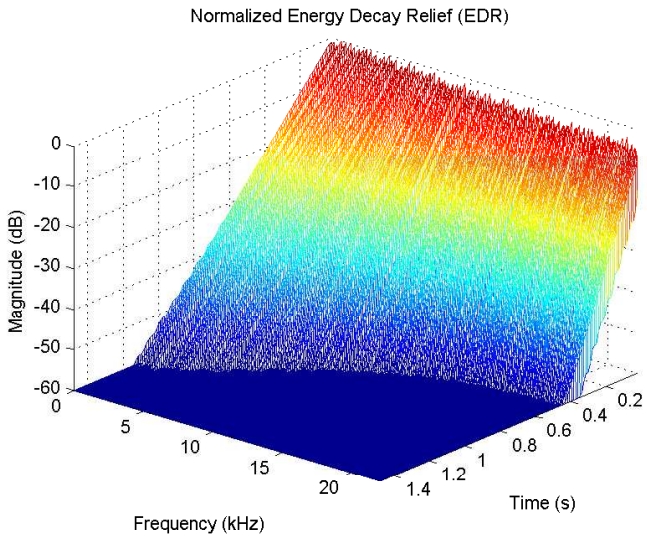


Figure 13: *Frequency-dependent energy decay for Room 1 modeled with the low-pass filter.*

4 Conclusions

In this paper we propose a new method which implements a simplified radiance transfer method using the Feedback Delay Network structure. It inherits the accuracy of the physical model of the acoustic radiance transfer method, at a much lower computational load. The sound quality approaches that of the radiance transfer method, with the increase of delay lines. This method can simulate room acoustic characteristics, and synthesize virtual reverberation in real time, taking into account the physical characteristics of the simulated room. Further extensions can investigate the grouping scheme, in order to obtain a more flexible group design, and possibly to introduce spatial information for each group.

References

1. T. Lokki, L. Savioja, R. Väänänen, J. Huopaniemi, T. Takala, *IEEE Computer Graphics and Applications* **22**, 49 (2002).

2. T. Funkhouser, P. Min, I. Carlbom, *Proceedings of the 26th annual conference on Computer graphics and interactive techniques* (ACM Press/Addison-Wesley Publishing Co., 1999), pp. 365–374.

3. T. Ajdler, L. Sbaiz, M. Vetterli, *IEEE Transactions on Signal Processing* **54**, 3790 (2006).

4. R. Mignot, G. Chardon, L. Daudet, *SPIE Optical Engineering+ Applications* (International Society for Optics and Photonics, 2011), pp. 813808–813808.

5. L. Bianchi, D. Markovic, F. Antonacci, A. Sarti, S. Tubaro, *Workshop on Application of Signal Processing to Audio and Acoustics (WASPAA)* (IEEE, 2013), pp. 1–4.

6. H. Kuttruff, *Room acoustics* (CRC Press, 2009).

7. I. Dokmanić, R. Parhizkar, A. Walther, Y. M. Lu, M. Vetterli, *Proceedings of the National Academy of Sciences* **110**, 12186 (2013).

8. J. B. Allen, D. A. Berkley, *The Journal of the Acoustical Society of America* **65**, 943 (1979).

9. J. Borish, *The Journal of the Acoustical Society of America* **75**, 1827 (1984).

10. A. Krokstad, S. Strom, S. Sørsdal, *Journal of Sound and Vibration* **8**, 118 (1968).

11. P. S. Heckbert, P. Hanrahan, *ACM SIGGRAPH Computer Graphics* **18**, 119 (1984).

12. T. Funkhouser, *et al.*, *Proceedings of the 25th annual conference on Computer graphics and interactive techniques* (ACM, 1998), pp. 21–32.

13. G. I. Koutsouris, J. Brunskog, C.-H. Jeong, F. Jacobsen, *The Journal of the Acoustical Society of America* **133**, 3963 (2013).

14. M. Hodgson, E.-M. Nosal, *The Journal of the Acoustical Society of America* **120**, 808 (2006).

15. E.-M. Nosal, M. Hodgson, I. Ashdown, *The Journal of the Acoustical Society of America* **116**, 970 (2004).

16. S. Siltanen, T. Lokki, L. Savioja, *Acta Acustica united with Acustica* **95**, 106 (2009).

17. L. Antani, A. Chandak, M. Taylor, D. Manocha, *IEEE Transactions on Visualization and Computer Graphics* **18**, 261 (2012).

18. L. Antani, A. Chandak, L. Savioja, D. Manocha, *ACM Transactions on Graphics (TOG)* **31**, 7 (2012).

19. M. R. Schroeder, *Journal of the Audio Engineering Society* **10**, 219 (1962).

20. J. A. Moorer, *Computer music journal* pp. 13–28 (1979).

21. J. Stautner, M. Puckette, *Computer Music Journal* **6**, 52 (1982).

22. J.-M. Jot, A. Chaigne, *Audio Engineering Society Convention 90* (Audio Engineering Society, 1991).

23. L. Dahl, J.-M. Jot, *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00), Verona, Italy* (2000).

24. F. Menzer, *Audio Engineering Society Conference: 40th International Conference: Spatial Audio: Sense the Sound of Space* (Audio Engineering Society, 2010).

25. B. Holm-Rasmussen, H.-M. Lehtonen, V. Välimäki, *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-13)* (2013).

26. E. Vickers, J.-L. L. Wu, P. G. Krishnan, R. N. K. Sadanandam, *Proceedings of the AES 121th Convention* (2006).

27. V. Valimaki, J. D. Parker, L. Savioja, J. O. Smith, J. S. Abel, *IEEE Transactions on Audio, Speech, and Language Processing* **20**, 1421 (2012).

28. D. Rocchesso, J. O. Smith, *IEEE Transactions on Speech and Audio Processing* **5**, 51 (1997).

29. E. De Sena, H. Hacihabiboglu, Z. Cvetkovic, *Audio Engineering Society Conference: 41st International Conference: Audio for Games* (Audio Engineering Society, 2011).

30. M. Chemistruck, K. Marcolini, W. Pirkle, *Audio Engineering Society Convention 133* (Audio Engineering Society, 2012).

31. S. Siltanen, T. Lokki, S. Kiminki, L. Savioja, *The Journal of the Acoustical Society of America* **122**, 1624 (2007).

32. H. Bai, G. Richard, L. Daudet, *Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)* (IEEE, 2013), pp. 1–4.

33. M. Carroll, C. Chien, *The Journal of the Acoustical Society of America* **62**, 1442 (1977).

34. J.-M. Jot, A. Chaigne, *Audio Engineering Society Convention 90* (Audio Engineering Society, 1991).

35. J. Huopaniemi, V. Valimaki, M. Karjalainen, R. Vaananen, *Proceedings of the International Computer Music Conference, Thessaloniki, Hellas* (The International Computer Music Association, 1997), pp. 200–203.

36. Z. Rafii, B. Pardo, *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR)* (2009), pp. 285–290.