GEOMETRIC-BASED REVERBERATOR USING ACOUSTIC RENDERING NETWORKS

Hequn Bai, Gaël Richard*

Institut Mines-Télécom
Télécom ParisTech, CNRS-LTCI, France

Laurent Daudet

Institut Langevin, ESPCI ParisTech
Univ. Paris Diderot, CNRS UMR7587, France

ABSTRACT

Many virtual reality applications incorporate realistic room acoustic simulation to provide increased immersiveness and realism. Traditional geometric methods, although providing modeling accuracy, are usually impractical for use in interactive applications. At the same time, artificial reverberators, with feedback rendering structure, are widely used as a low-cost alternative. This paper presents the design of a geometric-based artificial reverberator inspired by the acoustic rendering equation (ARE) and the feedback delay networks (FDN). The simplified acoustic rendering equation, which models both specular and diffuse reflections, is incorporated with the FDN structure. Our reverberator, despite of modeling the diffuse and late reverberation, is also capable of simulating the early/specular reflections with accuracy. This novel work is among the very few works which are capable to simulate early reflections using feedback delay networks.

***Index Terms*—** room acoustics, reverberation, feedback delay networks, acoustic rendering equation

1. INTRODUCTION

Room acoustics reconstruction considers the problem of modeling the sounds received at the listener's position in a sound scene. Multiple sound sources can be considered as well as their reflections and diffractions from surface materials and objects in order to simulate a realistic sound. The sound propagation effect is commonly divided into two parts: Early Reflections (ER) and Late Reverberation (LR). ER helps in localization, while LR gives an impression of the size of the environment and level of furnishing and absorptivity [1].

Numerous methods have been proposed for modeling the Room Impulse Response (RIR). Amongst geometric methods, the most widely used are the Image Source Method (ISM) [2] and the Ray tracing method [3]. Since the computation time of both methods becomes very large if high reflection orders are needed, they are mostly used for low-order early reflections. Recent advances of beam tracing [4, 5, 6, 7] optimise the computation and achieve interactive modeling performance for both moving sources and receivers.

The late reverberation tail of the RIR usually presents a distinct noise-like character due to the isotropic addition of a multitude of late acoustics reflections. This motivated the deployment of statistical method for modeling the LR [8]. Nevertheless, since the statistical methods do not take into account the detailed room geometry and the precise positions of sound sources and listeners, the simulation accuracy is thus not guaranteed.

In [9], an acoustic rendering equation (ARE) method was presented which can handle both diffuse and nondiffuse reflections. Since the high-order reflections computation is precomputed and decoupled from the run-time computation procedure, this method is well suited for dynamic applications with moving sources/receivers. In [10] the accuracy of the modeled early reflections was studied using the ARE model and the results showed that the ARE model provides good accuracy for both early/specular reflections and late/diffuse reverberation.

Feedback delay networks (FDN), a statistical reverberator utilising feedback loops in the simulation structure, are widely used to model late reverberation in hybrid methods [11] where geometric methods are used to model the more important early/specular reflections. Although it is common to split ER and LR, it would be desirable to have a single generic method that would be accurate for both parts. It can be noticed that, in the ARE model, the sound reflections between surface patches along the discretized direction can be regarded as feedback loops. Thus it is naturel to incorporate the ARE model to the FDN for the synthesis of room reverberation. Our approach inherits the computational efficiency of the FDN structure, but models the specular/diffuse reflections by using the directional discretization of the ARE method. In other words, this novel reverberator, named acoustic rendering networks (ARN), is to model the ARE using the FDN structure and relate the FDN to the physical room layout. This study is an extension of the geometry-based reverberator presented in [12]. [12] estimates the parameters of the FDN to model the diffuse reflections, while the presented work introduces directional discretization and models the specular reflections as well.

The paper is organised as follows. In section 2 we recall the basics of the ARE model and the main computation procedure. In section 3 we propose a new acoustic rendering equation based on FDN. The experimental results are given in section 4 and some conclusions are suggested in section 5.

2. ACOUSTIC RENDERING EQUATION

2.1. Continuous acoustic rendering equation

The ARE [9, 10] models the sound energy received at the listener's location by the contribution of the outgoing sound energy flux on the surface in each direction. Let $L(x, \Omega, t)$ be the time-dependent outgoing sound energy flux at point x along direction Ω. Then the acoustic rendering equation can be described as

$$L(x, \Omega, t) = L_0(x, \Omega, t) + \int_s R(x, x', \Omega, t) L(x', \frac{x - x'}{|x - x'|}, t) \mathrm{d}x',$$

where L_0 is the initial emitted sound energy and L is the total outgoing energy. $R(x, x', \Omega, t)$ is the reflection kernel which describes

*This work was partly supported by the European Commission under contract FP7-ICT-287723 - REVERIE.

978-1-4799-7450-4/15/$31.00 ©2015 IEEE

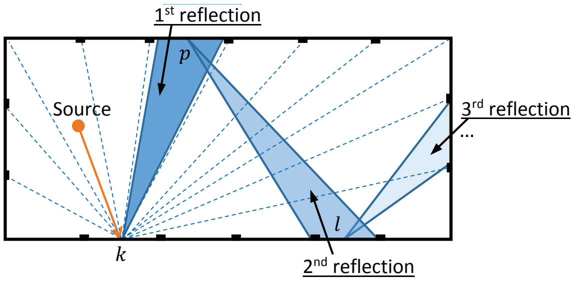

Figure 1: Illustration of discretized specular reflections.

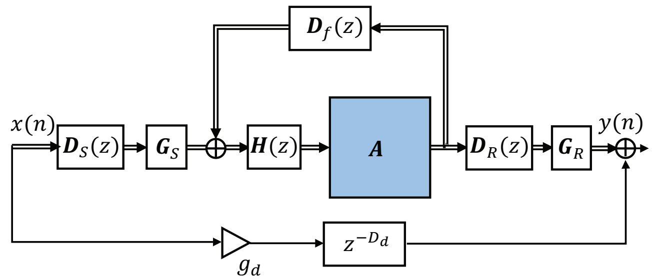

Figure 2: System structure of the acoustic rendering networks.

how the outgoing energy from point x' influences the outgoing energy at point x in direction Ω.

2.2. Discretization

In order to solve the acoustic rendering equation by numerical simulation, the room surface is discretized into patches. The Bidirectional Reflectance Distribution Function (BRDF), which defines how the incoming sound flux is reflected at a surface, is discretized by dividing the hemisphere into solid angles.

Unlike the directional discretization scheme used in [9] where the azimuth and elevation of the hemisphere are uniformly quantized, we divide the hemisphere with respect to all the other visible patches in the enclosure. As an example, the discretized direction for patch k is illustrated by the dotted lines in Figure 1. In our discretion scheme, one quantized direction corresponds to an unique patch-to-patch pair. This directional division scheme reduces the redundancy of the discretization between near patches where multiple quantized solid angles hit onto the same patch, and increases the accuracy of discretization between far patches where one quantized solid angle hits onto multiple patches. This discretization scheme also eases the implementation using feedback loop structure, since the patch-to-patch reflection and the directional discretization are unified.

For simplicity, we further simplify the ARE by assuming that the incoming energy from patch k to patch p will be specularly reflected along only one discretized direction, for example, to patch l, as shwon in Figure 1. This can be done by choosing patch l as the direction which receives the most reflected energy.

2.3. Discretized acoustic rendering equation

Let $\bar{i} = \{k, p\}$ denote the discretized direction element from patch k to patch p, and $\bar{j} = \{l, q\}$ denote the discretized direction element from patch l to patch q. Suppose the initial outgoing energy for the direction $\bar{j}$ is $\hat{L}_0^{\bar{j}}$. The total outgoing energy for the direction $\bar{i}$ due to the reflections from all other patches is

$$\hat{L}^{\bar{i}} = \sum_{\bar{j}=1}^{\infty} F_{\bar{j},\bar{i}} \hat{L}_0^{\bar{j}}. \tag{1}$$

$F_{\bar{j},\bar{i}}$ is determined by the discretized reflection kernel, which models the reflected energy flux along the direction $\bar{i}$ by the contribution of the reflected or emitted energy flux along the direction $\bar{j}$. We put $\hat{L}^{\bar{i}}$ into vector form ϕ and $F_{\bar{j},\bar{i}}$ into matrix F,

$$\phi = \begin{pmatrix} \hat{L}^1 \\ \vdots \\ \hat{L}^N \end{pmatrix}, F = \begin{pmatrix} F_{1,1} & \cdots & F_{1,N} \\ \vdots & \ddots & \vdots \\ F_{N,1} & \cdots & F_{N,N} \end{pmatrix}, F_{\bar{i},\bar{i}} = 0. \tag{2}$$

Let ϕ_0 denote the initial outgoing energy along the directions between all patches due to the arriving sound from the source, and ξ denote the contribution of energy arrived at the listener from the directions of all patches, then the resulting source-to-listener response (excluding the direct sound) can be formed as

$$\chi = (\phi_0 + \sum_{n=1}^{\infty} \phi_0 F^n)\xi^T = \phi_0(I + \sum_{n=1}^{\infty} F^n)\xi^T. \tag{3}$$

We use energy-based ray tracing to compute the initial outgoing energy ϕ_0, final gathering energy ξ and the matrix F as follows. During pre-computation, rays are emitted uniformly along the discretized directions $\bar{i}$ over the patch. When each ray encounters surfaces, it is specularly and/or diffusely reflected, according to the pre-defined BRDF. The reflected energy flux in each outgoing direction $\bar{j}$ is accumulated. The direction which receives the most specularly reflected energy is chosen as the unique specular reflection direction and transports all the reflected energy. This forms the one-order direction-to-direction response $F_{\bar{i},\bar{j}}$.

At the initial shooting stage, rays are uniformly emitted from the sound source. These rays are collected in each discretized direction to form the initial condition ϕ_0. During the final gathering stage, energy is uniformly emitted from all visible patches along the discretized directions and accumulated at the listener's location to form the final gathering vector ξ.

3. ACOUSTIC RENDERING NETWORKS

3.1. Design overview

The acoustic rendering networks based reverberator proposed in this paper uses one delay line to represent a discretized direction between the patches. We name this reverberator acoustic rendering networks because it is a combination of acoustic rendering method and feedback delay network structure. Figure 2 shows a conceptual depiction of the ARN reverberator.

3.2. Feedback delay components

The feedback delay components include the feedback matrix A, unidirectional delay units $D_f(z)$ and absorptive filters $H(z)$. Each discretized reflection direction corresponds to a delay line in the feedback delay networks. Thus the total number of delay lines is $N = \sum_{k=1}^{M} N_k$, where N_k is the number of visible patches from patch k and M is number of patches. Let's place the delay lines in the order of $[1 \to \rho_1, \cdots, k \to \rho_k, \cdots, M \to \rho_M]$, where ρ_k denotes the assemble of visible patches from patch k.

The length of the delay between patch k and p is calculated as $D_{k\to p} = \lfloor F_s \|\mathbf{x}_k - \mathbf{x}_p\|/c\rceil$, where c is the sound celerity, F_s is the sampling rate, $\mathbf{x}_k$ and $\mathbf{x}_p$ are the positions of the central points of patch k and p respectively. Then the feedback delay vector is

$$D_f(z) = [D_{1\to\rho_1}(z), D_{2\to\rho_2}(z), \cdots, D_{M\to\rho_M}(z)]. \quad (4)$$

The feedback attenuation is the caused by the wall material absorption from the emitting patch, which can be a fixed attenuation factor or a frequency-dependent wall absorption filter. The feedback attenuation vector is

$$H(z) = [\underbrace{H_1(z), \cdots, H_1(z)}_{N_1}, \cdots, \underbrace{H_M(z), \cdots, H_M(z)}_{N_M}], \quad (5)$$

where $H_k(z)$ is the wall reflection filter of patch k.

The feedback matrix defines which delay lines are connected via specular or diffuse reflections, and the amount of energy exchanged between them. We first neglect the diffuse reflections. Suppose the incoming energy from patch k to patch p is specularly reflected to the direction from patch p to patch l. Then the $\bar{i}^{th}$ row of the feedback matrix is

$$[\underbrace{0, \cdots, 0}_{N_1}, \cdots, \underbrace{0, \cdots, 1, \cdots, 0}_{N_p}, \cdots, \underbrace{0, \cdots, 0}_{N_M}],$$

where $\bar{i} = \{k, p\}$, and the "1" is at the corresponding index $\bar{j}$ for the direction pair $\{p, j\}$.

Because of the reflection kernel simplification in Section 2.2, each row of the feedback matrix has only one "1" element, and the rest are all zeros. This reduces the feedback matrix to a permutation matrix, namely P. If diffuse reflections are considered, the i^{th} row of the feedback matrix is

$$[\underbrace{0, \cdots, 0}_{N_1}, \cdots, \underbrace{\frac{2}{N_p}, \cdots, \frac{2-N_p}{N_p}, \cdots, \frac{2}{N_p},}_{N_p}, \cdots, \underbrace{0, \cdots, 0}_{N_M}],$$

where the specular reflection contains energy of $1 - \frac{4(N_p-1)}{{N_p}^2}$ and the diffuse reflections contain energy of $\frac{4(N_p-1)}{{N_p}^2}$.

The specular and diffuse energy is distributed in such a way that the corresponding sub-matrix is a permuted householder matrix of order N_p, which is itself unitary and lossless. Thus the feedback matrix A containing diffuse reflection can be constructed as

$$A = \mathrm{diag}([\Lambda_{N_1}, \Lambda_{N_2}, \cdots, \Lambda_{N_M}])P, \quad (6)$$

where Λ_n is the householder matrix of order n and P is the permutation matrix. Since the matrix $\mathrm{diag}([\Lambda_{N_1}, \Lambda_{N_2}, \cdots, \Lambda_{N_M}])$ is diagonalized householder matrices and thus is unitary. Its permutation is also unitary. The unitary feedback matrix guarantees that the FDN is lossless and that the energy decaying trend of the reverberator is only affected by the attenuation vector. This makes it possible to link the frequency-dependent absorption property of the material with the attenuation filter of each delay line.

3.3. Initial and final components

The length of the delay between the source at $\mathbf{x}_S$ and the central point of the patch k at $\mathbf{x}_k$ is determined by the propagation delay $D_{Si} = \lfloor F_s \|\mathbf{x}_S - \mathbf{x}_k\|/c\rceil$. Thus the initial-shooting delay vector is

$$D_S(z) = [\underbrace{D_{S1}(z), \cdots, D_{S1}(z)}_{N_1}, \cdots, \underbrace{D_{SM}(z), \cdots, D_{SM}(z)}_{N_M}]. \quad (7)$$

The initial-shooting amplitude vector is determined by the amount of energy received by each discretized angle from the source. Since the feedback delay networks simulate the reflections in the sense of sound pressure, the received energy is square-rooted to get G_S.

Like the initial shooting delays, the length of the delay line between the central point of the patch k at $\mathbf{x}_k$ and the listener at $\mathbf{x}_R$ is determined by the propagation delay $D_{iR} = \lfloor F_s \|\mathbf{x}_k - \mathbf{x}_R\|/c\rceil$. The final-gathering delay vector is

$$D_R^T(z) = [D_{\rho_1 R}(z), D_{\rho_2 R}(z), \cdots, D_{\rho_M R}(z)]^T. \quad (8)$$

The final-gathering amplitude vector G_R is the square-root of the received energy at the listener from each discretized angle.

3.4. Relation to previous works

After JOT reverberator [13], numerous extensions have been reported. However, among them, only a few have been devoted to model early/specular reflections using feedback networks [14, 15]. In [12], the geometric acoustic model and the parameters of the feedback networks are linked and a geometry-based artificial reverberator is proposed. Compared with these works, the proposed acoustic rendering networks have the following major differences.

The work in [12] relates the geometric acoustic model with the feedback delay network. A geometric-based FDN is constructed by studying the sound energy exchange between each delay line using the acoustic radiance transfer model. However, the radiance transfer model utilised in [12] only includes diffuse reflections, where the ARN model incorporates specular reflections.

In [14], the early/specular reflections are modeled using the FDN whose parameters are estimated using the ISM method. This model relies on two parallel networks, one for the early reflections and another for the late reverberation. In the proposed ARN, early reflections and late reverberations are modeled in one generic feedback network which unifies both specular and diffuse reflections.

[15] uses digital waveguide mesh to model the sound propagation. The method only guarantees the accuracy of the first order specular reflections by estimating the parameters using the ISM. The simulated RIR quickly transits to diffuse-like dense pulses from the second order reflection. In the ARN model, although directional discretization error may exist, the specular and diffuse reflections are modeled to infinite high order using the feedback loops.

4. NUMERICAL EVALUATION

4.1. Early reflections

Two examples are given to illustrate the simulated early reflections.

The Room 1 is a shoebox room with dimension (in m) $L = [l_x, l_y, l_z] = [15, 20, 5]$. The surfaces are discretized into 152 square patches, each of size 2.5×2.5m. The source is placed at $\mathbf{x}_S = [5.78, 14.91, 3.27]$, and the receiver at $\mathbf{x}_R = [8.23, 11.92, 2.02]$. All the walls have a frequency-independent absorption coefficient $\alpha = 0.3$, which corresponds to the reflection

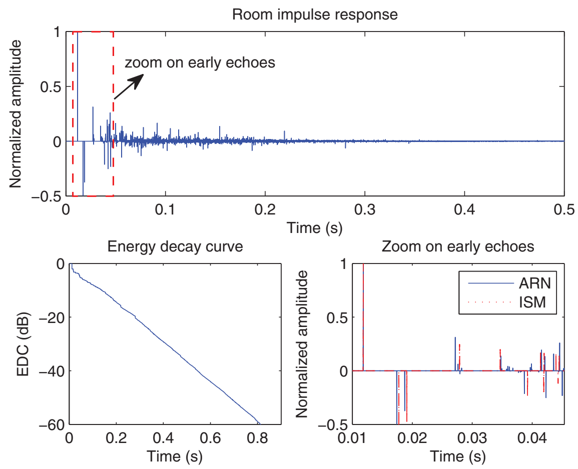

Figure 3: Simulation results of Room 1 with absorption $\alpha = 0.3$.

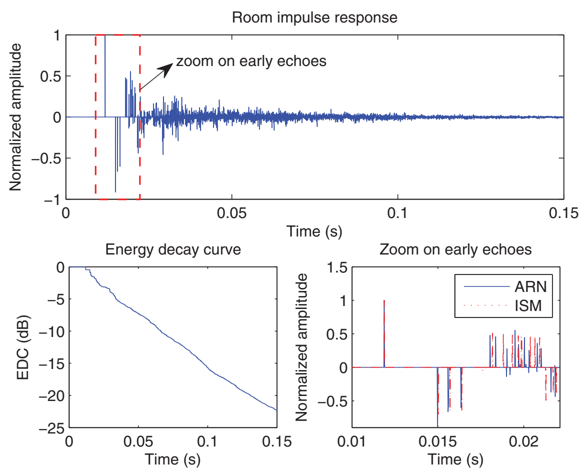

Figure 4: Simulation results of Room 2 with absorption $\alpha = 0.25$.

coefficient $\beta = \sqrt{1-\alpha}$. The simulation results are shown in Figure 3.

As expected, the simulated RIR has distinct early reflections during the first 50ms. The simulated RIR transits to diffuse-like dense pulses gradually, while the energy of distinct specular reflections diminishes. This is in accordance with sound propagation properties. Carefully observing the zoomed-in early echoes, we can see that most of the early reflections are modelled with good accuracy. Slight time shifts, compared to the reference ISM method, is the result of the coarse discretization, where the reflecting point is always chosen as the central point of each patch. However, such time shifts are hard to perceive for human ears.

The Room 2 is a shoebox room with dimension (in m) $L = [l_x, l_y, l_z] = [4, 6, 4]$. The surfaces are discretized into 32 square patches, each of size 2×2m. The source is placed at $\mathbf{x}_S = [1.21, 4.78, 3.02]$, and the microphone at $\mathbf{x}_R = [3.11, 1.92, 0.83]$. All the walls have a frequency-independent absorption coefficient $\alpha = 0.25$. The simulation results are shown in Figure 4. Most of the early echoes are well modeled, except one first order reflection and one second order reflection that are largely under-estimated. The errors of the amplitudes of the modeled early reflections mainly come from the coarse spatial discretization and the simplification of the ARE model. Besides using a denser discretization, improvement of the initial shooting and final gather schemes can also increase the modeling accuracy of the early reflections, as detailed in [10].

The computation times using Matlab on a standard computer (2.8 GHz CPU) to synthesize a one-second RIR are 429s for Room 1 and 16.5s for Room 2. Coarser discretization can further reduce the computation time. For example, the shoebox can be discretized into 6 patches, i.e. each wall being a patch. In this case, the attenuation filter of each delay line need to be adjusted using the ISM model in order to compensate the degradation of accuracy of the modeled early reflections. Besides, the sparsity of the feedback matrix can be explored for more efficient implementations.

4.2. Late reverberation

Reverberation time T_{60} is used as a rough metric to assess the performance of the ARN reverberator in modeling the late reverberation. The reverberation time is predicted by the Sabine formula [1]. The absorption coefficient varies from $\alpha = 0.2$ to $\alpha = 0.6$.

Table 1: Prediction of reverberation time.

Room	α	0.2	0.3	0.4	0.5	0.6
1	Sabine (s)	1.271	0.847	0.635	0.508	0.424
	ARN (s)	1.317	0.823	0.613	0.477	0.392
	Error (%)	3.6	2.8	3.5	6.1	7.6
2	Sabine (s)	0.604	0.403	0.302	0.242	0.201
	ARN (s)	0.703	0.449	0.339	0.263	0.212
	Error (%)	16.3	11.4	12.2	8.6	5.4

Table 1 shows the predicted T_{60} using the ARN model. The prediction results are good for Room 1 where the highest prediction error is 7.6%. However, the prediction for Room 2 has higher error, mainly because the reverberation time is shorter and thus is more sensitive to errors. It may also come from the coarse discretization of the geometry.

Informal listening tests of the late reverberation reveal that, due to the high number of delay lines and the permutation of the feedback matrix, the late reverberation is of satisfying sound quality with smooth reverberant sound.

5. CONCLUSION

In this paper we present a novel geometry-based artificial reverberator which is capable of modeling both early/specular reflection and late/diffuse reverberations. Although coarse discretization is the main source of error, the modeled early echoes have considerable accuracy. The results also shows that the extension to specular reflections does not degrade its performance in late reverberation modeling. The sparsity of the feedback matrix can be explored for more efficient implementations in the future. Besides, some more quantitative measures can be investigated to assess the accuracy of the early reflections.

6. REFERENCES

[1] H. Kuttruff, *Room acoustics*. SPON Press, London, UK, 4th edition, 2009.

[2] J. B. Allen and D. A. Berkley, “Image method for efficiently simulating small-room acoustics,” *The Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.

[3] A. Krokstad, S. Strom, and S. Sørsdal, “Calculating the acoustical room response by the use of a ray tracing technique,” *Journal of Sound and Vibration*, vol. 8, no. 1, pp. 118–125, 1968.

[4] T. Funkhouser, I. Carlbom, G. Elko, G. Pingali, M. Sondhi, and J. West, “A beam tracing approach to acoustic modeling for interactive virtual environments,” in *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, 1998, pp. 21–32.

[5] C. Lauterbach, A. Chandak, and D. Manocha, “Interactive sound rendering in complex and dynamic scenes using frustum tracing,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 6, pp. 1672–1679, 2007.

[6] F. Antonacci, M. Foco, A. Sarti, and S. Tubaro, “Fast tracing of acoustic beams and paths through visibility lookup,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 4, pp. 812–824, 2008.

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

[8] M. R. Schroeder, “Natural sounding artificial reverberation,” *Journal of the Audio Engineering Society*, vol. 10, no. 3, pp. 219–223, 1962.

[9] S. Siltanen, T. Lokki, S. Kiminki, and L. Savioja, “The room acoustic rendering equation,” *The Journal of the Acoustical Society of America*, vol. 122, no. 3, pp. 1624–1635, 2007.

[10] H. Bai, G. Richard, and L. Daudet, “Modeling early reflections of room impulse responses using a radiance transfer method,” in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2013.

[11] V. Valimaki, J. D. Parker, L. Savioja, J. O. Smith, and J. S. Abel, “Fifty years of artificial reverberation,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 5, pp. 1421–1448, 2012.

[12] H. Bai, G. Richard, and L. Daudet, “Late reverberation synthesis: from radiance transfer to feedback delay networks,” *submitted to IEEE Transactions on Audio, Speech, and Language Processing (under revision)*, 2015. Available at https://hal.archives-ouvertes.fr/hal-01142568.

[13] J.-M. Jot and A. Chaigne, “Digital delay networks for designing artificial reverberators,” in *Proceedings of the 90th Audio Engineering Society Convention*, 1991.

[14] F. Menzer, “Binaural reverberation using two parallel feedback delay networks,” in *Proceedings of the 40th Audio Engineering Society International Conference: Spatial Audio: Sense the Sound of Space*, 2010.

[15] E. De Sena, H. Hacihabiboglu, and Z. Cvetkovic, “Scattering delay network: an interactive reverberator for computer games,” in *Proceedings of the 41st Audio Engineering Society International Conference: Audio for Games*, 2011.