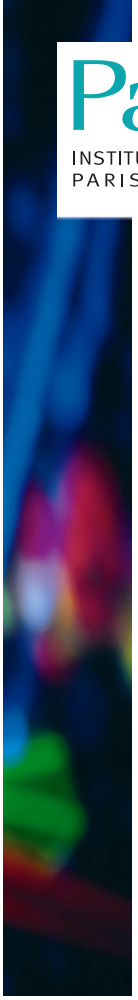
Hybrid models for acoustic reverberation

Hequn Bai



ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

TELECOM ParisTech

2016-ENST-0009

EDITE - ED 130

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

TELECOM ParisTech

Spécialité « Signal et Images »

présentée et soutenue publiquement par

Hequn BAI

le 8 février, 2016

Modèles Hybrides pour la Réverbération Sonore

Directeur de thèse : **Gaël RICHARD**
Co-encadrement de la thèse : **Laurent DAUDET**

Jury

M. Sylvain MARCHAND, Professeur, Université de La Rochelle	Rapporteur
Mme. Rozenn NICOL, Chargé de Recherche, Orange Labs	Rapporteur
M. Julien DE ROSNY, Professeur, Institut Langevin	Président du jury
M. Lauri SAVIOJA, Professeur, Aalto University	Examinateur
M. Olivier WARUSFEL, Chargé de Recherche, IRCAM	Examinateur
M. Nicolas TSINGOS, Directeur de Recherche, Dolby Laboratories	Examinateur
M. Gaël RICHARD, Professeur, Télécom ParisTech	Directeur de thèse
M. Laurent DAUDET, Professeur, Université Paris Diderot	Directeur de thèse

TELECOM ParisTech
école de l'Institut Mines-Télécom - membre de ParisTech
46 rue Barrault 75013 Paris - (+33) 1 45 81 77 77 - www.telecom-paristech.fr

T H È S E

Modèles Hybrides pour la Réverbération Sonore

Hequn BAI

RESUME : Grâce au diverse applications prévues dans le divertissement, l'éducation et les domaines professionnels, la création d'un monde virtuel interactif a intéressé beaucoup de chercheurs dans l'académie et l'industrie. L'espace acoustique trois-dimensionnelle peut être synthétisé à partir de deux perspectives : les approches physiques et les approches perceptives. Méthode de transfert de rayonnement acoustique est une méthode efficace pour modéliser les réflexions diffuses et la réverbération tardive. Une extension de la méthode de transfert de rayonnement est proposée dans cette thèse, qui permet de modéliser les réflexions précoces et spéculaires, en conservant son avantage pour la simulation de réverbération tardive.

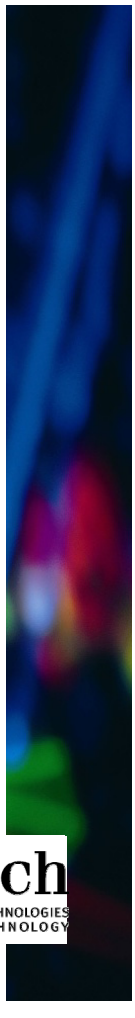
Les approches perceptives utilisent des paramètres acoustiques pour modéliser la perception acoustique principale de l'environnement modélisé. Des réseaux récursifs de lignes à retard sont beaucoup utilisés pour générer de la réverbération tardive. Dans cette thèse, une nouvelle méthode est présentée, qui hérite l'efficacité de la structure des réseaux récursifs de lignes à retard, mais en le même temps, relie les paramètres des réseaux directement à la géométrie de l'environnement modélisé. Puis, l'approche physique et l'approche perceptive ont été combinées. La méthode de transfert de rayonnement acoustique simplifiée, avec extension à modéliser les réflexions spéculaires et diffuses, est incorporée à la structure des réseaux récursifs de lignes à retard. Le nouveau filtre réverbérant, en dépit de la modélisation des réfections diffuses et la réverbération tardive, est également capable de simuler les réflexions précoces et spéculaires avec précision.

MOTS-CLEFS : acoustique des salle, reverberation, audio 3D, realité virtuelle

ABSTRACT : There is an increasing interest in creating interactive virtual worlds due to the wide variety of potential applications in entertainment and education. The 3D acoustic scene can be synthesized from two perspectives : the physical approaches and the perceptual approaches. Acoustic radiance transfer method is an efficient ray-based method to model the diffuse reflections and the late reverberation. An extension of the Radiance Transfer Method (RTM) is proposed in this thesis, which allows modeling the early part of specular reflections while keeping the advantage of the original model for the late reverberation simulation.

Feedback delay networks are widely used structures to generate the late reverberation. A new method is presented in this thesis, which inherits the efficiency of the Feedback Delay Network (FDN) structure, but aims at linking the parameters of the FDN directly to the geometries of the modeled environment. The relation is achieved by assigning a physical meaning to each delay line and studying the sound energy exchange between them. Then the physical approach and the perceptual approach are combined. The simplified acoustic Radiance Transfer Method, with extension for both specular and diffuse reflections, is incorporated with the Feedback Delay Networks. The new reverberator, despite of modeling the diffuse and late reverberation, is also capable of simulating the early and specular reflections with accuracy.

KEY-WORDS : room acoustics, reverberation, 3D audio, virtual reality.



TELECOM ParisTech

EDITE

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

Résumé

Grâce au diverse applications prévues dans le divertissement, l'éducation et les domaines professionnels, la création d'un monde virtuel interactif a intéressé beaucoup de chercheurs dans l'académie et l'industrie. Dans ces scenarios d'applications, il est particulièrement important d'avoir des outils puissants et efficaces pour la synthèse de l'audio trois-dimensionnel pour créer des espaces acoustiques dynamiquement à partir de la géométrie virtuelle de l'espace. Un tel système qui synthèse la propagation du son dépend de la géométrie de l'espace simulée, de la présence d'obstacles dans l'espace et de la position des sources sonores et des récepteurs. Pendent des dernières vingt années, de nombreuses études ont été menées dans le domaine de la simulation de la propagation du son dans les environnements complexes.

L'espace acoustique trois-dimensionnelle peut être synthétisé à partir de deux perspectives : les approches physiques et les approches perceptives. Les approches physiques considèrent le problème de la modélisation de la propagation du son précis dans l'environnement pour toutes les fréquences audibles par l'oreille d'humaine. Méthode de transfert de rayonnement acoustique est une méthode efficace pour modéliser les réflexions diffuses et la réverbération tardive. Une extension de la méthode de transfert de rayonnement est proposée dans cette thèse, qui permet de modéliser les réflexions précoces et spéculaires, en conservant son avantage pour la simulation de réverbération tardive. Les réflexions entre les murs sont pré-calculées et découplés de la course de calcul. Lorsque le caractéristiques et les locations de la source et le récepteur changent, que l'état initial et final doivent être ré-calculés et mis à jour. Cela réduit le calcul pour les scènes dynamiques.

Les approches perceptives, par contre, utilisent des paramètres acoustiques, comme le temps de réverbération, pour modéliser la perception acoustique principale de l'environnement modélisé. A cause de l'efficacité de calcul, des réseaux récursifs de lignes à retard sont beaucoup utilisés pour générer de la réverbération tardive. Dans cette thèse, une nouvelle méthode est présentée, qui hérite l'efficacité de la structure des réseaux récursifs de lignes à retard, mais en le même temps, relie les paramètres des réseaux directement à la géométrie de l'environnement modélisé. La relation est obtenue par étudier la signification physique de chaque ligne à retard et attribuer de l'échange de l'énergie acoustique entre les murs à chaque ligne.

Puis, dans la dernière partie de cette thèse, l'approche physique et l'approche perceptive ont été combinées. La méthode de transfert de rayonnement acoustique simplifiée, avec extension à

modéliser les réflexions spéculaires et diffuses, est incorporée à la structure des réseaux récursifs de lignes à retard. Le nouveau filtre réverbérant, en dépit de la modélisation des réfections diffuses et la réverbération tardive, est également capable de simuler les réflexions précoces et spéculaires avec précision. La matrice récursive définit les réflexions spéculaires entre composants de la géométrie. Car les réflexions entre les composants de la géométrie ont été découplées de la course de calcul, seulement l'état initial et final doivent à être mis à jour lorsque la source / récepteur se déplace. Ainsi, ce filtre réverbérant est approprié pour des applications dans les scènes dynamiques.

Mots-clé:

L'Effet acoustique de salle réverbération, le temps de réverbération, transfert de rayonnement acoustique, réseau récursif de lignes à retard, réalité virtuelle, audio trois-dimensionnel, filtre réverbérant.

Abstract

There is an increasing interest in creating interactive virtual worlds due to the wide variety of potential applications in entertainment and education. In these applications, creating dynamic 3D acoustic scenes from a given 3D graphical set up is of particular importance.

The 3D acoustic scene can be synthesized from two perspectives: the physical approaches and the perceptual approaches. The physical approaches model the accurate sound propagation in the acoustic environment using the wave-based methods or the ray-based methods. Acoustic radiance transfer method is an efficient ray-based method to model the diffuse reflections and the late reverberation. An extension of the Radiance Transfer Method (RTM) is proposed in this thesis, which allows to model the early part of specular reflections while keeping the advantage of the original model for the late reverberation simulation. The high-order wall-to-wall reflections are pre-computed and stored in order to speed up the run-time computation. Two final gathering schemes are studied and compared with respect to the accuracy of the modeled early reflections.

The perceptual approaches, on the other hand, use global acoustic parameters to model the main perceptual aspects of the modeled environment. Feedback delay networks are widely used structures to generate the late reverberation. However, this method relies on a careful tuning of the different synthesis parameters, either estimated from a pre-recorded impulse response, or set manually from experience. A new method is presented in this thesis, which inherits the efficiency of the Feedback Delay Network (FDNs) structure, but aims at linking the parameters of the FDNs directly to the geometries of the modeled environment. The relation is achieved by assigning a physical meaning to each delay line and studying the sound energy exchange between them. The assigning scheme of each delay line is designed in order to keep the feedback matrix unitary. This new method is tested under various room geometries and the results show that the modeled late reverberation is in good agreement with the virtual geometry settings.

Then, in the last part of this thesis, the physical approach and the perceptual approach are combined. The simplified acoustic Radiance Transfer Method, with extension for both specular and diffuse reflections, is incorporated with the Feedback Delay Networks. The sparse feedback matrix defines the specular reflections between wall components. The new reverberator, despite of modeling the diffuse and late reverberation, is also capable of simulating the early and specular reflections with accuracy. Since the high order reflections between wall components have been decoupled from the run-time computation, only the initial and final state of the reverberator need

to be updated when the source or the receiver moves. Thus this reverberator is suitable for 3D acoustic rendering of dynamic scenes.

Keywords:

Room acoustics, reverberation, reverberation time, feedback delay networks, acoustic radiance transfer, virtual reality, 3D audio.

Remerciements

This thesis is based on my research activities carried out at Télécom ParisTech. I would like to express my sincere gratitude to my supervisors, Professor Gaël Richard from Télécom ParisTech and Professor Laurent Daudet from Université Paris Diderot. I appreciate that Gaël gave me the opportunity to pursuit a PhD under his supervision and guidance. The research topic he offered me is exhilarating which does attract me in the first sight. His academic spirit passed me a message, that the great finding comes out of hard work, inspiration, and innovation. It brings me to a level of interest in wanting to carry this message forward.

I am deeply grateful to my second supervisor, Professor Laurent Daudet. His professional knowledge of room acoustics precisely defined the area of my research before I really knew where I stood. Thank him for the great encouragement and inspirations he gave to me which led to various interesting works afterwards.

I would like to thank Professor Sylvain Marchand from Université La Rochelle and Doctor Rozenn Nicol from Orange Labs for being the report readers (rapporteurs) for my thesis. Thanks to the other defense committee members: Prof. Lauri Savioja from Aalto University, Professor Julien de Rosny from Institut Langevin, Professor Olivier Warusfel from IRCAM and Dr. Nicolas Tsingos from Dolby Laboratory. Many thanks to Professor Julien de Rosny who presents as the president of the jury.

I must thank all my colleagues at AAO (Audio, Acoustic and Waves) research group at Télécom ParisTech. Thank them for creating a pleasant working environment and a dynamic research atmosphere. Thank Professor Gaël Richard and Slim Essid for creating and maintaining a strong and experienced AAO research group. I believe strongly that my research gets benefit from the environment they created. I would like to thank Professor Yves Grenier, Slim Essid, Bertrand David, Roland Badeau, who are permanent staff at AAO and carry out research in various topics of audio signal processing. They set good examples for my research. Thanks to Dr. Liying Wei and Dr. Davide Andrea Mauro for working with me on the European project REVERIE and their kind help with my research. Thanks to Dr. Sebastien Fenet, Simon Leglaive, Simon Durand, Mounira Maazaoui, Xabier Jaureguiberry, Aymeric Masurelle, François Rigaud, Nicolás Lopez and Clement Laroche for spending nice time together in the lab, on the balcony of Bareau site and in the bars around after work. I appreciate these experience with you. Thanks also for Dr. Sebastien Fenet, Danny Sullian, Simon Leglaive, Simon Durand for sharing the office

with me. Many thanks to all the current and previous colleagues, including Yousra Bekhti, Karan Nathwani, Jair Montoya Martinez, Romain Serizel, Thomas Fillon, Victor Bisot, Floriane Dardard, Tom Dupré la tour, Paul Magron, Benoit Fuentes, Manuel Moussallam and so on.

Outside Télécom, I would like to express my gratitude to my friends for your warm accompany these years: Jing Ba, Huan Meng, Tingting Guan, Sara Soulignac, Matthew Elgoyhen, Xian Shi, Hao Wang, Peng Jin, Nicolas Kuchmann, Lina Qu, Gioele Puglisi, and to Leana, Rui, Ray and Andreas, some of them are now in Sweden and China. I would give my thanks to my parents, for giving me a tremendous amount of encouragement and tolerance, no matter how far I am away. I thank my brother, for being so supportive for the decisions that I made.

At the end, I thank Professor Arne Leijon at KTH, whose inspirational courses motivated me to do this Ph.D research on audio processing. Besides, I would like to thank my supervisor at University of Electronic Science and Technology of China, Professor Lemin Li, fellow of Chinese Academy of Engineering. His spirit of dedication to the science and education and his attitude of doing research will influence me and other Chinese young researchers in a long time.

Contents

List of Figures

Introduction en Français

0.1 Introduction de la Thèse

Un acousticien évalue d'abord les effets acoustiques d'une salle en frappant ses mains et écouter la réverbération. Cette méthode empirique a été souvent utilisée pour obtenir une première compréhension de la caractéristique acoustique de l'environnement et est très pratique pour percevoir certaines des qualités acoustiques de la salle. La même musique, quand il est joué dans des environnements différents, comme dans un studio, une salle de concert ou en plein air, donne très différentes expériences perceptives. Par conséquent, les effets de la salle dans laquelle nous percevons le son ont une influence importante sur ce que nous entendons, puis, influencent notre évaluation des environnements d'écoute.

Si les effets des salles peuvent être quantitativement reconstruites, nous pouvons reconstruire l'espace acoustique pratiquement, pour que nos oreilles entendent le son dans l'environnement virtuellement modélisé, sans être physiquement présenté là. Cela donne un sujet intéressant appelé « acoustique virtuel », qui est devenu populaire dans la recherche et dans l'industrie en raison de la demande constante de l'évolution des techniques de réalité virtuelle.

Le système acoustique virtuelle de l'environnement comprend la modélisation de la source sonore (sources directionnelles), réverbération de la chambre (son direct, les réflexions, la diffusion, la réverbération) et la modélisation du récepteur (diffusion individuelle / de diffraction causé par la diversité de la structure anatomique de l'homme). Les informations de la source et le récepteur doivent être surveillés et régulièrement mis à jour, ainsi que les effets acoustiques de la chambre. Ainsi, un système acoustique virtuelle sophistiqué doit être interactif, dynamique, individualisé et en mesure d'effectuer le calcul en temps réel.

L'un des défis dans les systèmes acoustiques virtuels est comment simuler les effets acoustiques de la chambre de manière interactive et réaliste. Différentes méthodes existent pour la modélisation des effets acoustiques virtuels, de méthodes géométriques à base de méthodes fondées sur la perception.

Méthodes physiques [Botteldooren, 1995; Savioja et al., 1995] résoudre les équations d'ondes numérique et peuvent fournir ainsi des résultats plus précis. Toutefois, ils sont également des méthodes avec calcul le plus lourd.

Approches géométriques à base de rayons utilisent sur les principes physiques de la propagation du son, mais avec l'hypothèse simplificatrice que les ondes sonores se déplacent sous forme de rayons. L'image Méthode Source (ISM) [Allen and Berkley, 1979] calcule trajets de réflexion spéculaire en considérant un ensemble de sources virtuelles, qui sont positionnés à des emplacements symétriques par rapport à chaque mur. La méthode de lancer de rayons calcule rayons sonores issus de la source à travers la scène acoustique, et les reçoie dans le récepteur. La méthode de transfert rayonnement (en anglais : Rdiance Transfer Method RTM) [Lewers, 1993], provenant de l'algorithme de transport de la lumière, de manière efficace, modèle les réflexions diffuses et la réverbération tardive.

D'autre part, des méthodes statistiques, également appelés comme des méthodes fondées sur la perception, utilisent des paramètres acoustiques, tels que le temps de réverbération, de modéliser les principaux aspects de la perception de la réverbération. Schroeder filtre [Schroeder, 1962] utilise des filtres en peigne récursives et des filtres passe-tout pour simuler les échos denses de la réverbération tardive. Il est généralisée dans une structure appelée réseau récursif de lignes à retard (en anglais : Feedback Delay Networks FDN) [Stautner and Puckette, 1982], qui ont été largement utilisés pour synthétiser la réverbération tardive dans les méthodes hybrides.

Cependant, différentes méthodes ont leurs propres avantages et faiblesses dans la modélisation acoustique. Par exemple, une méthode peut être précis à modéliser les premières réflexions, mais inefficace pour la modélisation de réverbération tardive. Cela motive de nombreuses œuvres qui combinent différentes méthodes afin de simuler une réverbération complète. L'objectif de cette thèse est sur l'enquête et l'innovation des méthodes de modélisation acoustique et leurs modèles hybrides, notamment les méthodes hybrides combinant les méthodes de transfert de rayonnement et des réseaux récursif de lignes à retard.

Contribution de la thèse sur la modélisation acoustique

Bien que différentes méthodes de modélisation acoustiques existent, nous nous concentrons notre étude sur leur capacité d'interactivité et de calcul temps réel. L'accent est mis sur les méthodes de transfert de rayonnement et des réseaux récursif de lignes à retard. Le principal avantage de la méthode de transfert de luminance énergétique est qu'il découple la source et le récepteur à partir de la procédure de calcul, ce qui le rend approprié pour la modélisation acoustique dans

les scénarios dynamique et interactif. Les réseaux récursifs de lignes à retard, en raison de son efficacité de calcul, est adapté pour le rendu en temps réel de la réverbération tardive.

Dans la modélisation de réverbération tardive, le réseau récursif de lignes à retard est une structure efficace pour générer des impulsions qui synthétisent la tendance de décroissance de l'énergie de la réverbération tardive. Cependant, une telle méthode repose sur choisissant soigneusement des paramètres, estimée soit par une mesure de la scène acoustique réel, ou par un réglage manuel oar les acousticiens avec expérience. Dans le même temps, l'échange de son radiosité modélisé par la méthode de transfert de rayonnement peut être considéré comme un énorme réseaux rétroaction de retard où la radiosité de chaque élément est renvoyée à un autre élément dans la chambre. Ainsi, la première contribution de la thèse est **de combiner les deux méthodes et de relier les paramètres du FDN à l'échange d'énergie de la méthode de transfert de rayonnement.** Nous ainsi présentons une réverbération tardive géométrique. La réverbération géométrique hérite de l'efficacité de calcul de la structure de FDN, mais estime les paramètres des FDN directement à partir de la géométrie virtuelle. L'échange de radiosité dans le procédé de transfert de rayonnement est mis en correspondance avec les lignes à retard de la structure de FDN. Cela crée une nouvelle façon d'estimer les paramètres des FDN pour la simulation de réverbération tardive.

La méthode de transfert de rayonnement a été prolongé sous une forme générale dans [Siltanen et al., 2007], nommé l'équation de rendu acoustique (en anglais Acoustic Rendering Equation ARE), qui est capable de modéliser les réflexions diffuses et aussi spéculaires. Cela se fait en discrétisant les angles de réflexion de chaque élément. Cette méthode est très efficace pour la modélisation de la réverbération tardive et la prévision des paramètres acoustiques. Cependant, la précision des premières réflexions modélisés ne sont pas expliquées en détail dans la littérature. Dans la deuxième contribution, **la précision de la modélisation des premières réflexions en utilisant l'équation de rendu acoustique est étudiée.** La procédure finale est spécialement étudié et une nouvelle schéma de collecte uniforme est proposé. Deux schémas de collecte finaux ont été comparés. Utilisation de la schéma de collecte uniforme, une méthode générique unique peut efficacement modéliser à la fois le début et la fin de réverbération.

La dernière contribution est sur **la combinaison de l'équation de rendu acoustique avec les réseaux récursif de lignes à retard pour modéliser les réflexions précoces et spéculaires.** Dans le ARE modèle, les réflexions sonores entre éléments de surface peuvent être considéré comme des boucles de rétroaction dans les réseaux récursif de lignes à retard. Ainsi, il est naturel d'intégrer le ARE modèle avec le FDN pour la synthèse de réverbération. Cependant, quelques modifications et simplifications doivent être fait pour le ARE modèle. Premièrement, le schéma de discrétisation directionnelle est modifié afin de relier les éléments de calcul de ARE

avec les lignes rétroaction de retard. Deuxièmement, la réflexion est simplifiée pour maintenir les réseaux rétroaction de retard stable. L'équation de rendu acoustique simplifiée, qui modélise les réflexions spéculaires et diffuses, est incorporé dans le FDN structure. Ce travail est l'un des rares solutions qui sont capables de simuler les premières réflexions en utilisant les réseaux récursif de lignes à retard.

Les résultats des travaux de cette thèse sont publiés dans les articles de revues et de conférences:

- [Bai et al., 2013] H. Bai, G. Richard, and L. Daudet. Modeling early refections 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)*, pages 1–4, New Paltz, NY, October 2013.

- [Bai et al., 2015b] H. Bai, G. Richard, and L. Daudet. Late reverberation synthesis: From radiance transfer to feedback delay networks. *Audio, Speech, and Language Processing, IEEE/ACM Transactions on*, 23(12):2260–2271, 2015.

- [Bai et al., 2015a] H. Bai, G. Richard, and L. Daudet. Geometric-based reverberator using acoustic rendering networks. In *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 1–4, New Paltz, NY, October 2015.

- H. Bai, G. Richard, and L. Daudet. Reflection-based variable delays in feedback delay networks. will be submitted to *Journal of the Audio Engineering Society - special issue*, 2016.

0.2 Techniques de la Modélisation Acoustique

Un son émis par la source peut subir plusieurs réflexions et réfractions dans la scène acoustique avant qu'il ne soit entendu par l'auditeur. Réaliste modélisation de la propagation du son et de rendu peuvent ajouter une nouvelle couche de réalisme à la visualisation des environnements virtuels.

Le scène acoustique en 3D peut être synthétisé à partir de deux perspectives: les approches physiques et les approches perceptives. Les approches physiques modéliser la propagation du son précis dans l'environnement acoustique en utilisant les méthodes basées sur les vagues ou les méthodes géométriques à base de rayons. Les approches perceptives, souvent cités comme

l'acoustique statistique, d'autre part, utilisent des paramètres acoustiques pour modéliser les principaux aspects de la perception de l'environnement.

Dans chapitre 2, l'état de l'art des techniques de modélisation acoustique est étudié par catégorie. Nous concentrons notre discussion sur les méthodes à base de rayons et les méthodes statistiques. Nous concluons chapitre 2 en discutant du potentiel dans la combinaison de différentes techniques de modélisation et les avantages et les inconvénients des modèles hybrides existants.

Acoustique Numérique à Base d'Onde

Méthodes numériques à base d'onde sont des méthodes les plus précis pour la simulation de la propagation des ondes sonores dans des environnements. La solution numérique directe de l'équation d'onde, atteint par l'utilisation de méthodes d'éléments discrets, fournit un moyen d'émuler le comportement des champs sonores qui intègre intrinsèquement des propriétés ondulatoires. Dans le même temps, ils sont les méthodes les plus compliquées aussi.

Les méthodes de modélisation acoustiques utilisés actuellement peuvent être classés en trois catégories: Finite Element Methods (FEM), Boundary Element Methods (BEM) et méthodes de Finite-Difference Time-Domain (FDTD) [Botteldooren, 1995; Savioja et al., 1995]. Dans FEM, l'espace considéré est discrétisé en éléments volumétriques, tandis que dans BEM, où seulement la frontière de l'espace sont discrétisée. Alors que les FEM et les BEM fonctionnent généralement dans le domaine de fréquence, les FDTD simule l'acoustique de la salle dans le domaine temporel. Le principe des FDTD est que les dérivés dans les équations d'ondes sont remplacés par les différences finies correspondant [Strikwerda, 2004].

Cependant, les méthodes à base d'onde sont également des méthodes les plus compliquées. Ainsi, dans la pratique, ils ne sont généralement utilisés pour la modélisation de l'acoustique de la salle que pour les fréquences basses [Kleiner et al., 1993; Pietrzyk, 1998]. Pour auralization en temps réel, elles ne sont utilisées que pour modéliser le son direct et les premières réflexions. En tenant compte de simplifier le calcul, le FDTD peut être adapté en temps réel auralization [Strikwerda, 2004].

Étant donné que les méthodes à base d'onde ne sont pas l'objet de cette thèse, les auteurs intéressés sont invités à se référer à ces progrès récents [Kowalczyk and Van Walstijn, 2011; Sakamoto et al., 2006; Savioja et al., 2006; Southern et al., 2013; Svensson and Kristiansen, 2002; Van Walstijn and Kowalczyk, 2008].

Acoustique Géométrique sur la Base des Rayons

Dans acoustique géométriques où la taille de la salle est grande par rapport aux longueurs d'onde du son, le concept d'une onde n'est plus d'importance; il est remplacé à la place par le concept d'un rayon [Kuttruff, 2009]. Cette condition est souvent rencontré dans la salle acoustique. Le rayon sonore a direction bien définie de propagation, et est soumise aux réflexions spéculaires et diffuses quand il frappe les surfaces réfléchissantes. La vitesse limitée du son exige que le temps de propagation ne peut pas être négligée. De nombreux effets acoustiques sont liées au temps de propagation, tels que des échos et des réverbérations.

De nombreuses techniques ont été proposées pour simuler la propagation du son à l'aide du concept de rayons [Savioja and Svensson, 2015]. Ces techniques sont appelées des méthodes telles que géométriques. Dans cette section, nous présentons le concept principal des méthodes géométriques classiques. L'accent est mis à la discussion sur la différence entre ces techniques et leurs avantages et inconvénients.

La Méthode de la Source d'Image

La Méthode de la Source d'Image (ISM), d'abord présenté dans [Allen and Berkley, 1979], est la méthode géométrique beaucoup utilisé pour trouver de manière exhaustive tous les chemins de réflexion spéculaire provenant d'une source à un récepteur. Les chemins de réflexion spéculaire sont calculées en tenant compte d'un ensemble de sources virtuelles. Le premier ordre de sources virtuelles est obtenue en considérant les emplacements en miroir de la source par rapport à chaque paroi réfléchissante. Le deuxième ordre de sources d'image est obtenue à partir de la première commande de sources virtuelles: chaque source virtuelle de premier ordre est mis en miroir par rapport aux parois réfléchissantes pour générer des sources virtuelles du deuxième ordre.

L'avantage de la méthode de la source d'image est qu'il donne une liste totalement exhaustive des chemins de réflexion spéculaire à l'ordre de réflexion défini. Le principal inconvénient est que sa complexité de calcul augmente de façon exponentielle avec l'ordre de la réflexion puisque le nombre de sources virtuelles augmente également de façon exponentielle.

Méthodes Déterministes du tracé de rayons

La méthode du tracé de rayons [Krokstad et al., 1968; Kulowski, 1985] est l'une des méthodes géométriques les plus importantes dans la pratique et a plusieurs variations. En partant du concept principal de la méthode de tracé de rayon, de nombreux algorithmes ont été dérivés de celui-ci, parmi lesquels:

- Tracé de rayons stochastique [Dalenbäck, 1996];
- Cone tracing [Martin et al., 1993];
- Sonel mapping [Kapralos et al., 2006];
- Phonon tracing [Bertram et al., 2005];

Le procédé de tracé de rayon émet des rayons sonores provenant de la source, dont chacun porte une certaine quantité d'énergie. Les rayons rebondissent entre les surfaces dans la scène en fonction de la manière de réflexion pré-définie, qui sont finalement collectées par le récepteur. Bien que la réflexion spéculaire est communément utilisé, diffraction et diffusion peut également être considérées [Lehnert and Blauert, 1992; Naylor, 1993].

Acoustics Statistique sur la base de Perception

Les méthodes de modélisation acoustique statistiques sont souvent appelés réverbères artificiels [Välimäki et al., 2012; Välimäki et al., 2016]. L'observation des caractéristiques de bruit comme de la réverbération tardive des réponses impulsionnelles acoustiques de la salle a motivé l'étude de modélisation de la réverbération tardive en utilisant des méthodes statistiques. Méthodes statistiques utilisent des paramètres acoustiques tels que le temps de réverbération pour simuler les effets acoustiques d'une scène. Les méthodes statistiques ont acquis une grande popularité, en particulier dans les applications de réalité virtuelle interactives telles que des jeux vidéo. En général, ils produisent des effets acoustiques perceptivement pertinents, qui répondent à l'expérience d'écoute dans les scénarios d'application interactifs.

Filtre Schroeder

Le filtre Schroeder utilise des filtres en peigne parallèles pour produire des impulsions de décomposition exponentielles pour modéliser les effets de réverbération des salles. Afin d'obtenir une réponse en fréquence plate, plusieurs filtres en peigne, normalement au moins 6 filtres en peigne, sont connectés en parallèle. Les lacunes de la réponse en fréquence en forme de peigne sont remplis par les filtres en peigne ayant différentes longueurs de retard.

Feedback Delay Networks

Dans [Stautner and Puckette, 1982] une structure bien adaptée pour la réverbération artificielle appelée "réseaux récursifs de ligne à retard (FDN)" a été introduit. Cette structure se caractérise par un ensemble de lignes à retard connectés en une boucle de rétroaction à travers une matrice

de rétroaction. Les signaux de toutes les lignes à retard interagissent par le biais de la matrice de rétroaction, ce qui augmente la densité d'échos de la réverbération tardive. Jot [Jot, 1992a,b; Jot and Chaigne, 1991] en outre étudié les FDNs et a développé des techniques associées pour la conception de bonne de qualité de la réverbération.

Méthodes Hybrides

Les méthodes de modélisation acoustique, de la base d'onde, de filtres réverbérants artificiels, à base de rayons, ont tous leurs propres forces et faiblesses.

Dans la pratique, les méthodes hybrides sont souvent utilisées où deux ou trois méthodes sont combinées, l'une est utilisée pour modéliser les premières réflexions et l'autre est utilisée pour modéliser la réverbération tardive. Ou une méthode est utilisée pour modéliser les fréquences basses, et l'autre est pour les fréquence élevée. Il est impossible de donner une liste exhaustive de toutes ces méthodes hybrides. Nous passons en revue quelques méthodes hybrides couramment utilisés [Lewers, 1993; Savioja et al., 1999; Southern et al., 2013].

0.3 Filtres Réverbérants Basé sur la Géométrie

Réseaux récursifs de lignes à retard, comme une structure souvent utilisée pour la synthèse de réverbération artificielle, peuvent efficacement modéliser la réverbération tardive pour la chambre considérée. Dans ce secteur, nous étudions la relation entre la méthode de transfert de rayonnement et les réseaux récursifs de lignes à retard. Une telle relation peut être utilisée pour estimer les paramètres des réseaux récursifs de lignes à retard, et faire des ponts entre les deux méthodes. Cette section est un résumé du chapitre 3.

Chapitre 3 est organisé comme suit. Dans la section 3.1 nous rappelons la connaissance de la méthode de transfert de rayonnement acoustique et sa procédure de calcul pour la modélisation des réflexions diffuses. Dans la section 3.2 nous discutons de notre idée de relier les réseaux rétroaction de retard avec la méthode de transfert acoustique. Ensuite, dans la section 3.3 un paradigme de modélisation statistique géométrique à base est proposé, et la procédure de la conception de l'estimation de paramètre correspondant est décrit. Un accent particulier est mis sur la conception de la matrice de rétroaction qui rend les réseaux rétroaction de retard stable. Dans la section 3.6 nous utilisons certains modèles typiques de chambre pour évaluer cette méthode, à partir des deux points de vue numériques et perspectives perceptives.

Relation entre RTM et FDN

On peut remarquer que, dans la procédure de calcul de la méthode de transfert de rayonnement [Lewers, 1993] dans le domaine temporel, l'itération des échanges d'énergie entre les éléments de surface peut également être considérée comme des boucles de rétroaction dans un réseau récursif de lignes à retard, comme illustré sur la figure 3.5. Ici, chaque ligne de retard représente la réflexion du son sur un élément de surface. La propagation du son entre les éléments de surface est alors représentée sous forme de lignes à retard dans le réseau.

Il est alors un intérêt pratique pour combiner ces deux méthodes, en utilisant un réseau récursif de lignes à retard en tant que structure de synthèse, tandis que les caractéristiques de rétroaction sont estimées par la propriété de transfert acoustique dans la chambre. La méthode hybride utilise ensuite des lignes à retard pour représenter un groupe de réflexions diffuses entre surfaces de la chambre, avec les paramètres acoustiques moyennes par les réflexions diffuses inclus dans le groupe.

Si nous remplaçons chaque interaction élément-à-élément par une ligne à retard, et mis la matrice de rétroaction avec les facteurs de forme correspondants, nous pouvons, en théorie, de construire d'énormes FDN avec des centaines de milliers de lignes à retard, qui serait alors équivalent au modèle de transfert de rayonnement.

Cependant, dans la plupart des cas d'applications, la tendance de la décroissance de l'énergie est suffisante pour percevoir la dimension de l'environnement. Ainsi, nous pouvons ignorer le modèle précis de l'amplitude de chaque impulsion, et se concentrer uniquement sur la modélisation des statistiques d'échange d'énergie. Ceci suggère qu'une simplification de modèle RTM peut être utilisée, en regroupant les interactions énergétiques élément-à-élément, afin de réduire le nombre de lignes à retard. Le paramètre de chaque ligne de retard est déterminé par la valeur moyenne de l'interactions d'élément-à-élément du groupe.

Filtre réverbérant basé sur la Modèle RTM

La filtre réverbérant proposée est illustrée à la figure 3.10. Elle consiste en un FDN-RTM pour la réverbération tardive, comme indiqué dans le cadre supérieur, et une série de filtres pour les premières réflexions, comme indiqué dans le cadre inférieur.

La structure des FDN-RTM est similaire au réseau récursif de lignes à retard proposé dans [Jot and Chaigne, 1991], sauf que pour chaque ligne à retard, une unité de retard initial et une unité de retard finale sont ajoutés avant et après chaque ligne récursif. Chaque ligne à retard représente un groupe d'interactions de réflexion de première. L'unité de retard initial est le retard moyen de la distance entre la source sonore et les éléments de chaque groupe, et l'unité de retard finale

est le retard moyen à partir des éléments groupés au récepteur. Le premier ordre de réflexions sonores doivent être modélisés et ajoutés à l'aide d'un filtre FIR supplémentaire (voir la figure 3.10).

Soit $\wp_n$ désignent l'ensemble des interactions sonores élémént-à-élément dans le groupe n. Ensuite, si le nombre total de groupes est de N, le réseau récursif aura N lignes à retard tel que représenté sur la figure 3.10.

Les paramètres des FDN-RTM proposées sont estimés comme suit à l'aide de l'ordre-1 interactions élément-à-élément dans le modèle de transfert de rayonnement.

1. Matrice de rétroaction

L'élément a_{mn} dans la matrice de rétroaction décrit en fait la proportion d'énergie qui est transportée par les interactions d'élémént-à-élément au sein du groupe m, qui sera réfléchie de manière diffuse et aller à une autre interaction d'élément-à-élément dans le groupe n, au cours d'une ordre-1 réflexion diffuse.

Notons que ℓ_m est l'énergie totale transportée par l'échange élémént-à-élément dans le groupe m et $\ell_{m,n}$ est l'énergie totale reçue par le groupe n grâce aux réflexions diffuses du groupe m. Ils peuvent être estimés à partir de la forme des facteurs $F_{i,j}$ dans le modèle RTM comme:

$$\ell_m = \sum_{i \to j \in \wp_m} F_{i,j}, \tag{1}$$

$$\ell_{m,n} = \sum_{i \to j \in \wp_m} \sum_{j \to k \in \wp_n} F_{i,j} F_{j,k}, \tag{2}$$

où $\wp_m$ et $\wp_n$ sont les ensembles d'interactions sonores dans le groupe m et n respectivement.

Les coefficients a_{mn} est défini comme:

$$a_{mn} = \frac{\ell_{m,n}}{\ell_m}, \tag{3}$$

$$sujet \sum_{n=1}^{N} a_{mn} = 1. \tag{4}$$

Et la matrice de rétroaction est:

$$A = \begin{pmatrix} \sqrt{a_{11}} & \cdots & \sqrt{a_{1N}} \\ \vdots & \ddots & \vdots \\ \sqrt{a_{N1}} & \cdots & \sqrt{a_{NN}} \end{pmatrix}. \tag{5}$$

2. Filtres d'atténuation

Les filtres d'atténuation $H_n(z)$ en FDN traditionnels sont des filtres passe-bas, qui modélisent l'absorption dépendant de la fréquence du matériau de surface et de l'air [Huopaniemi et al., 1997]. Pour FDN-RTM, les atténuations sont les moyennes des coefficients de réflexion de tous les interactions d'élément-à-élément au sein du groupe:

$$\bar{\theta}_n^{\,s} = \sum_{i \to j \in \wp_n} F_{i,j} \theta_i^{\,s}, \tag{6}$$

où $\bar{\theta}_n^{\,s}$ est le coefficient de réflexion moyen pour groupe n de bande de fréquence s du coefficient de réflexion du matériau, et $F_{i,j}$ est le facteur de forme correspondante.

Filtres d'ordre supérieur (FIR ou IIR) peuvent être conçus en utilisant des techniques de filtrage numérique basé sur AR (all-pole), MA (FIR), ou ARMA (pôle-zéro) modélisation. La méthode de Gauss-Newton peut être utilisé pour ajuster les coefficients d'absorption des bandes d'octave et d'optimiser les coefficients de filtre.

Résultats Expérimentaux

Figure 3.16 montre les courbes de décroissance de l'énergie des réponses impulsionnelles synthétisées en utilisant FDN-RTM et RTM pour quatre chambres considerées. Il est démontré que les réponses impulsionnelles estimées par FDN-RTM sont exponentiellement décroissants (diminuant par exemple linéairement sur une échelle logarithmique (dB)). Ceci est dû au fait que la matrice de rétroaction estimée est proche des matrices unitaires. Les tendance de décroissance de FDN-RTM concordent bien avec celles simulées en utilisant RTM, ce qui montre que FDN-RTM peut être utilisé comme une méthode alternative de RTM pour la réverbération tardive.

Les temps de réverbération estimés par FDN-RTM concordent bien avec celles estimées par RTM et la formule Sabine. Cependant, les temps de réverbération estimées par FDN-RTM sont légèrement plus longs. La raison principale est que la matrice de rétroaction estimée a des valeurs propres légèrement plus élevés que les autres, ce qui rend les lignes à retard correspondant désintègrent plus lent. Notez que la réverbération synthétisée des chambres 4 a une erreur plus élevé par rapport à RTM, principalement parce que la matrice de rétroaction estimée est pas

parfaitement unitaire (ses plus grandes valeurs propres étant 1.076).

Pour évaluer plus formellement la qualité sonore de la réverbération tardive, des tests d'écoute subjectifs ont été réalisées en fonction de la stimulation de l'UIT-RBS.1534 avec Invisible Référence et Anchor (MUSHRA) [ITU-1534, 2001]. Les tests ont été effectués en utilisant un casque de qualité professionnelle (Beyerdynamic DT $Pro770$). 10 auditeurs (à l'exclusion des auteurs) ont été impliqués, 9 hommes et 1 femme, âgés entre 20 et 32, toutes avec une certaine expérience des tests d'écoute. Les signaux secs d'entrée sont huit extraits d'enregistrements anéchoïques de la parole, piano, violoncelle et trompette sons choisis parmi [Bang&Olufsen, 1992; Open-AIR], chacune avec une durée d'environ 7-10 s. Les signaux en cours d'évaluation ont été générés par le FDN-RTM en utilisant 4, 8 et 16 lignes à retard, avec des paramètres estimés par les géométries des chambres. Quelques exemples sonores utilisés dans les tests peuvent être trouvés sur le site Web de compagnie. [1]

Figure 3.18 montre les statistiques de la qualité sonore évalué par les auditeurs. Les résultats montrent clairement que la qualité de son augmente avec le nombre de lignes à retard pour la voix et les instruments de musique. Bien que FDN-RTM avec 4 lignes à retard peut produire une excellente qualité sonore dans certains cas, en général, la qualité du son produite est irrégulière et n'est donc pas recommandé pour une utilisation pratique. La qualité sonore a une amélioration notable de 4 lignes à retard à 8 lignes à retard. Avec 16 lignes à retard, le son réverbéré devient plus lisse et est presque aussi bon que le RTM. Certains auditeurs ont confondu le FDN-RTM avec 16 lignes à retard avec la référence. En général, les sons réverbérants sont de meilleure qualité pour les sources musicales que pour les sources vocales, en particulier où un petit nombre de lignes à retard est utilisé. En conclusion, pour obtenir une qualité sonore satisfaisante pour la réverbération tardive, il est suggéré d'utiliser au moins 8 lignes à retard pour le FDN-RTM.

0.4 Filtre Réverbérant avec des Retards Variables

Dans le chapitre 3, les longueurs des lignes à retard de rétroaction sont choisies comme valeurs fixes. Cependant, les trajets de réflexion ont des longueurs de retard qui varie de façon continue, ce qui est difficile à modéliser en utilisant seulement quelques unités de retard dans les réseaux récursifs de lignes à retard. En outre, dans le modèle de transfert de rayonnement, les délais de réflexion entre les patches discrétisées varient aussi. Dans ce chapitre, nous examinons la variation des longueurs de retard dans le modèle de transfert de rayonnement et les réseaux récursifs. Extensions de l'utilisation des longueurs de retard variable sont faites pour les deux méthodes.

[1] http://perso.telecom-paristech.fr/~hbai/demo_ASLP2015/

Afin de parvenir à un système invariant dans le temps réel pour les réseaux récursifs, une nouvelle structure est proposée afin d'intégrer des unités de retard variable dans la réverbération artificielle. Cette section est un résumé du chapitre 4.

Chapitre 4 est organisé comme suit. Section 4.1 aborde les questions des longueurs de retard variant entre patches dans le modèle de transfert de rayonnement. Une extension de retards variables de façon aléatoire est proposé dans le modèle. Section 4.2 analyse le retard des longueurs caractéristiques de variation d'une salle, et propose un système de regroupement différent au chapitre 3 basé sur le délai le plus proche des longueurs principe. Un filtre en peigne à retard variable statistique est proposée et appliquée aux réseaux de rétroaction à retard. Étant donné que les statistiques unités de retard variant dans le temps ne sont pas des systèmes LTI, une nouvelle structure de rétroaction en utilisant des groupes de retard déterministes est proposé. Les résultats de simulation sont donnés dans la Section 4.3 et des conclusions sont tirées dans la Section 4.4.

Filtre Réverbérant avec des Retards Variables

Dans le cas de la modélisation de réverbération tardive diffuse, la modélisation précise des amplitudes de chaque impulsion est souvent difficile et inutile. Nous nous concentrons donc sur la modélisation statistique de l'échange d'énergie globale entre les surfaces. Nous approchons du modèle de transfert de rayonnement, le regroupement des réflexions avec des retards similaires. A titre d'exemple illustré à la figure 4.4, la distribution de probabilité des longueurs de retard de la salle sont divisés en quatre groupes. Chaque groupe est ensuite modélisée par une distribution uniforme.

L'unité des différents réverbères artificiels sont des lignes à retard récursives. La plus simple rétroaction récursive peut être représenté par un filtre de peigne [Schroeder, 1962]. Il est bien connu que le filtre de peigne avec une longueur de retard fixe produit des signaux avec un son métallique, qui est expliqué dans la figure 4.6, par le peigne-forme de sa réponse de fréquence.

Le peigne-forme des réponses de fréquence est causé par la répétition régulière des impulsions avec les mêmes retards. En introduisant variation des longueurs de retard, le peigne-forme des réponses de fréquence peut être éliminé, comme le montre la figure 4.8. Ainsi, nous utilisons une unité de retard variable. La longueur de retard varie au sein d'une distribution uniforme. Le son métallique est beaucoup.

En utilisant les retards variables, chaque groupe de la distribution de réflexion sur la figure 4.4 peut être considérée comme une ligne à retard dans la structure FDNs, dont les longueurs de retard sont variable en fonction de la distribution de délai prédéfini.

Pour la synthèse de son en ligne, les longueurs de retard variable doivent être remplacées par des valeurs déterministes. Nous ainsi proposons une implémentation déterministe où chaque unité de retard variable $z^{-\mathbf{D}_i(m)}$ est remplacé par M sous-unités de retard fixé, connectés en parallèle, comme dans la figure 4.11. Les longueurs de retard des unités de retard $d_i(k)$ sont choisies comme les instances des intervalles de longueur de retard, à la suite de leurs distributions de probabilité. Un exemple montrant la connectivité entre les unités de retard est donnée à la figure 4.11.

Un exemple montrant la connectivité entre les unités de retard est donnée à la figure 4.11. Le groupe de retard 1 représente la variable ligne de retard $z^{-\mathbf{D}_1(m)}$ et contient des retards $[d_1(1), d_1(2), d_1(3), \cdots]$, et le groupe de retard 2 représente $z^{-\mathbf{D}_2(m)}$ et contient des retards $[d_2(1), d_2(2), d_2(3), \cdots]$. Dans cette figure, les atténuations de ligne de retard et la matrice de rétroaction est ignoré pour la concision.

Le signal $x(n)$ entre seulement dans l'entrée de l'une des unités de retard de base dans chaque groupe. Nous avons choisi le 1^{st} unité de chaque groupe de retard de base dans l'exemple. Le signal $x(n)$ passe de 1^{st} de base unité de retard $d_1(1)$ dans le groupe 1. Ensuite, la sortie de la variable ligne de retard $z^{-\mathbf{D}_1(m)}$ étant alimenté à lui-même dans la figure 4.9 peut être représentée de manière déterministe, par la sortie de 1^{st} de base unité de retard $d_1(1)$ dans le groupe 1 étant renvoyé à l'entrée de l'une des unités de retard de base du même groupe. Dans cet exemple, il est renvoyé au 2^{nd} de base unité de retard $d_1(2)$. La sortie de la variable ligne de retard $z^{-\mathbf{D}_1(m)}$ étant renvoyée à la variable ligne de retard $z^{-\mathbf{D}_2(m)}$ dans la figure 4.9 peut être représentée par la sortie du 1^{st} de base unité de retard $d_1(1)$ en groupe 1 étant renvoyés à l'entrée de 2^{nd} de base unité de retard $d_2(2)$ du groupe 2.

Résultats expérimentaux

Il est montré dans la figure 4.15 que la réponse impulsionnelle modélisée par le FDN avec des retards variants a une réponse de fréquence plus régulière par rapport à celui modélisé par le FDN traditionnelle. Les fréquences modales sont plus denses, ce qui améliore la qualité sonore synthétisé et réduit l'audibilité des effets sonores métallique.

Les $4 \times 4(4)$ FDNS de retard variable utilisent un total de 16 retards lignes. La densité des impulsions des RIR synthétisés en utilisant $4 \times 4(4)$ retard variant DNs est plus élevé que les traditionnels 4×4 FDNS mais inférieure à la traditionnelle 16×16 FDNS. Les $4 \times 4(4)$ FDNS de retard variable utilisent 4 lignes à retard d'entrée, qui est le même que les traditionnels 4×4 FDNS, mais recueillent les sorties de tous les 16 lignes à retard, ce qui est le cas pour la traditionnelle 16×16 FDNS.

Une comparaison des temps de calcul utilisé par le FDN de retard variable et les FDNs traditionnels est donnée à la figure 4.17. Sur la figure 4.17, avec le même nombre d'unités de retard, le temps de calcul requis par le FDN de retard variable est inférieur à celui requis par les FDNs traditionnels.

Des tests informels d'écoute révèle que la qualité sonore en utilisant le FDNs variable a de grandes améliorations par rapport aux FDNs traditionnels. La qualité sonore en utilisant les retard variant est beaucoup mieux que les traditionnels où le son métallique est presque pas audible du tout. La qualité sonore en utilisant des retards variantes est mieux, surtout pour les signaux de parole. La raison est que par l'extension de chaque ligne à retard à plusieurs sous-unités prédéfinis par la connectivité, la répétition régulière des impulsions est beaucoup réduite et une réponse de fréquence beaucoup plus lisse peut être obtenu.

0.5 Modélisation des Premières Réflexions

Dans le chapitre 1, méthode de transfert rayonnement (RTM) a été étudiée, ce qui est un outil efficace pour modéliser des réflexions diffuses. Dans [Kiminki, 2005; Siltanen et al., 2007], une équation de rendu acoustique (ARE) a été formulé, comme une extension de la RTM, qui considère les réflexions diffuses et spéculaires. Car les réflexions d' ordre élevé sont pré-calculées et découplées de la procédure de calcul, cette méthode est bien adaptée pour les applications dynamiques avec des sources mobiles.

L'exactitude des premières réflexions modélisées est pas pleinement discuté dans la littérature. Dans ce chapitre, nous étudions la précision de la modélisation des premières réflexions de la méthode ARE et de modifier les schémas de collecte finaux utilisant ray tracing méthodes qui nous permet de modéliser les premières réflexions en conservant les avantages de la méthode originale pour la simulation de réverbération tardive. Cette section est un résumé du chapitre 5.

Chapitre 5 est organisé comme suit. Dans la section 5.1 nous rappelons les bases de la méthode ARE et la procédure de calcul principale. La formulation de la sont présentés dans cette section provient principalement de citep Siltanen 2007. Dans la section 5.1, nous expliquons notre méthode de calcul en utilisant la méthode de traçage de rayons. Dans la section 5.3 la précision de la modélisation de la méthode sont pour les premiers échos est étudié et deux systèmes de collecte finale sont comparés. Certains comparaison expérimentale sont donnés dans la section 5.4.

Les schémas de collecte finaux

Dans [Antani et al., 2012a,b], les rayons sont uniformément envoyés à partir de l'auditeur et reçus par chaque élément pour calculer la réponse énergétique pour la collecte finale, en utilisant un principe de réciprocité. Étant donné que ces travaux utilisent ce méthode pour modéliser les réflexions diffuses pour la réverbération tardive, l'effet de la discrétisation directionnelle dans la phase de collecte finale n'a pas été discutée. Pour la simulation des premiers échos, il est important de modéliser avec précision l'angle de réflexion et l'amplitude de l'énergie pour chaque angle solide. Une telle méthode de collecte finale peut produire échos précoces inexactes. Discrétisation de la géométrie et directions introduit des erreurs en deux façons. Tout d'abord, le flux d'énergie sortante dans une certaine direction est considéré comme constant. En fait, il peut avoir des variations considérables, en particulier pour les éléments proches de la source. Deuxièmement, parce que les sources sonores et les auditeurs sont libres de se déplacer dans l'espace sonore, il existe des positions pour les auditeurs où l'énergie estimé par la phase de collecte finale sont largement sous-estimée ou sur-estimée.

Une façon plus naturelle est d'envoyer les rayons du centre de chaque élément et de les accumuler à la position de l'auditeur, semblable à ce qu'ils ont fait pour la modélisation de la radiosité diffuse. Les rayons sont reçus par un volume de réception à la position de l'auditeur. Cependant, cette méthode produit toujours à de grandes zones où l'auditeur reçoit aucun son, et quelques autres zones où il reçoit de sons de plusieurs éléments, comme on peut le voir sur la figure 5.3.

Pour rendre le son réfléchi distribué plus uniformément et pour mieux couvrir l'espace acoustique, nous proposons d'émettre des rayons uniformément sur chaque élément le long de l'angle solide, au lieu d'émettre tous les rayons du centre du patch. Pour chaque angle solide d'un élément, nous choisissons uniformément K points sur sa surface. Nous envoyons N_R rayons de chaque point le long des directions uniformes dans l'angle solide. Chaque rayon transporte de l'énergie de $e_R = \frac{1}{K \times N_R}$. Ces rayons sont accumulés dans le volume de réception pour générer les réponses de collecte finale. Le schéma final de recueillir uniforme est illustrée à la figure 5.4.

Résultats expérimentaux

Figure 5.6 et la figure 5.7 affiche la reconstruction des échos. Dans la figure 5.6, la méthode proposée montre une meilleure reconstruction. Notez que le premier écho est largement surestimé par la méthode ARE-traditionnelle, mais modélisée avec plus de précision par la méthode proposée. Dans la figure 5.7 le premier écho est totalement manquée par la méthode ARE-traditionnelle, mais est bien préservé en utilisant la méthode proposée. L'amélioration est due à

l'effet de l'émission de moyenne d'une façon uniforme sur la surface de chaque élément. Notez que, comme le montre la figure 5.7, il existe encore quelques impulsions qui ne sont pas bien modélisées par la méthode. Dans ce cas, l'écouteur est très proche de la surface, ce qui conduit à des effets de discrétisation de surface.

Combiné avec des reflets d'ordre élevé, la méthode proposée est capable de modéliser efficacement l'ensemble de RIR. Afin de vérifier que la méthode proposée n'a pas affecté la réverbération tardive, nous avons calculé la durée de réverbération (T_{30}) des deux méthodes, et nous avons trouvé aucune différence significative.

En comparant les deux méthodes dans le même schéma de discrétisation, il est clair que la méthode proposée surpasse la méthode traditionnelle ARE. La méthode proposée peut parfois atteindre une performance similaire à la méthode traditionnelle utilisée avec discrétisation plus fine. En d'autres termes, la méthode proposée est une alternative compétitive à l'aide de petites parcelles, quand une grande précision est nécessaire avec des ressources de calcul limitées.

Il est intéressant d'observer à partir du tableau 5.1 qui divisant le plus fin de l'hémisphère ne met pas nécessairement une amélioration de la précision des premiers échos. Au contraire, les performances sont légèrement dégradée dans certains tests, mais pas de façon systématique. Toutefois, cette observation indique que lorsque plus de puissance de calcul et de mémoire est disponible, d'abord essayer de diviser la surface en plaques plus fines peuvent être une meilleure option.

En comparant les deux méthodes dans le même schéma de discrétisation, il est clair que les ARE technique surpasse la méthode traditionnelle ARE, à un coût très légère en termes d'exigences de calcul. Comme on le voit dans les valeurs moyennes dans les deux derniers modèles de discrétisation, la méthode proposée peut parfois atteindre une performance similaire à la méthode traditionnelle utilisée avec discrétisation plus fine. En d'autres termes, la méthode proposée est une alternative compétitive à l'aide de petites parcelles, quand une grande précision est nécessaire avec des ressources informatiques limitées.

Le grand nombre de lignes à retard des ARN peut entraîner une énorme charge de calcul. Cependant, la matrice de rétroaction est très faible. Ainsi, la majorité des calculs peut être enregistré en ignorant les éléments zéro.

La simulation utilise Matlab sur un ordinateur standard (2.8 GHz CPU) pour synthétiser une seconde d'RIR. Dans notre expérience en utilisant le code Matlab non optimisé, pour synthétiser une seconde d'RIR, Salle 1 simulation consommée 429s, et Salle 2 simulation 16.5s consommés. Ce ne sont pas encore prêts pour la simulation en temps réel, mais l'optimisation potentielle est encore possible.

Discrétisation grossier peut réduire davantage le temps de calcul. Par exemple, la boîte à chaus-

sures peut être discrétisé en 6 élément, à savoir chaque paroi étant un patch. Dans ce cas, le filtre d'atténuation de chaque ligne de retard doivent être ajustés à l'aide du modèle ISM afin de compenser la dégradation de la précision des premières réflexions modélisées.

0.6 Réseaux de Rendu Acoustiques

Les réseaux récursifs de ligne à retard (FDN) ont été beaucoup utilisés pour modéliser la réverbération tardive. Dans le modèle ARE au chapitre 3, les réflexions sonores entre les éléments de surface le long de la direction discrétisée peuvent être considérés comme des boucles de rétroaction. Ainsi, il est naturel d'intégrer le modèle ARE à la FDN pour la synthèse de la réverbération de la salle. Dans ce chapitre, nous présentons les réseaux de rendu acoustiques (ARN), une approche qui hérite de l'efficacité de calcul de la structure FDN, mais modélise les réflexions spéculaires et diffuses à l'aide de la discrétisation directionnelle de la méthode ARE. Cette étude est une extension de la réverbération basée sur la géométrie présentée dans le chapitre 1 [Bai et al., 2015b] et le schéma de modélisation des premières réflexions en utilisant ARE méthode présentée dans le chapitre 3 [Bai et al., 2013]. Cette section est un résumé du chapitre 6.

Chapitre 6 est organisé comme suit. Dans la section 6.1 les avantages et les inconvénients des réseaux rétroaction de retard sont brièvement résumées. L'idée de rejoindre le modèle ARE avec FDNS est alors proposé de surmonter l'un des principaux inconvénients de la suite. Dans la section 6.2 nous rappelons les bases du modèle ARE et la modification que nous avons fait afin de l'intégrer à la structure rétroaction de retard. Dans la section 6.3, nous décrivons les nouveaux réseaux de rendu acoustique et comment les paramètres des réseaux sont liés à la géométrie de la pièce modélisée. Certains résultats de simulation sont donnés dans la section 6.4.

Modification de l'Equation de Rendu Acoustique

Afin de résoudre l'équation de rendu acoustique par simulation numérique, la surface de la salle est discrétisé en plaques. La Fonction Bidirectionnelle de Distribution Réflectance (BRDF), qui définit la façon dont le flux de son entrant se reflète sur une surface, est discrétisé en divisant l'hémisphère dans des angles solides.

Dans [Siltanen et al., 2007], les directions de l'hémisphère d'une surface d'élément, sont divisés en m angles uniforme d'azimut et n angles d'élévation uniformes. Il en résulte qu'un angle solide discrétisé donné à partir d'un élément peut couvrir plusieurs autres correctifs.

Contrairement au schéma de discrétisation directionnel utilisé dans [Siltanen et al., 2007] où

l'azimut et l'élévation de l'hémisphère sont uniformément discrétisée, on divise l'hémisphère par rapport à tous les autres éléments visibles comme le montre la figure 6.3. A titre d'exemple, la direction discrétisée pour l'élément k est illustré par les lignes en pointillés sur la figure 6.3. Dans notre schéma de discrétion, une direction discrétisée correspond à une paire élément-à-élément unique. Ce schéma de division directionnelle réduit la redondance de la discrétisation où multiples angles solides discrétisées ont frappé sur le même élément, et augmente la précision de discrétisation entre les éléments loin.

Ce schéma de discrétisation facilite également la mise en œuvre en utilisant la structure de boucle de rétroaction, puisque la réflexion élément-à- élément et la discrétisation directionnelle sont unifiés.

Pour le raison de simplicité, nous simplifions encore l'ARE en supposant que l'énergie entrante d'élément k à élément p sera réfléchie de façon spéculaire seulement le long d'une direction discrétisée, par exemple, pour élément l, comme le montre la figure 6.3. Cela peut être fait en choisissant Patch l que la direction qui reçoit l'énergie la plus forte. Ceci est une approximation grossière, ce qui peut entraîner une imprécision de l'amplitude des échos précoces modélisés. Cependant, à des fins de auralization pour une application en temps réel, nous supposons que cela est de priorité secondaire.

Réseaux de Rendu Acoustiques

La réverbération des réseaux de rendu acoustique proposée utilise une ligne à retard pour représenter une direction discrétisée entre les éléments. Les paramètres acoustiques des directions élément-à- élément sont modélisées par unités de retard unidirectionnels $D_f(z)$ et filtres absorbants $H(z)$. Les lignes à retard sont interconnectées les uns aux autres par une matrice de rétroaction A. Cette matrice de réaction définit essentiellement des lignes à retard qui sont connectés via des réflexions spéculaires ou diffus et la quantité d'énergie échangée entre eux. En outre, la source initiale $D_S(z)$ et de l'amplitude G_S et microphone collecte finale $D_R(z)$ et de l'amplitude G_R sont reliés à chaque ligne à retard.

Nous nommons ce filtre réverbérant un réseau de rendu acoustique (ARN), car il est une combinaison de la méthode de rendu acoustique et la structure du réseau récursif de ligne à retard. Figure 6.6 montre un schéma conceptuel de ce filtre.

La faible densité de la matrice de rétroaction pour les réseaux de rendu acoustique est illustré sur la figure 6.8. La palette de couleurs de la figure est mise à l'échelle et inversée afin de rendre l'affichage de la couleur des éléments non-zéros plus visibles. Dans la figure ref fig: clairsemés, la couleur blanche représente des éléments zéro, points gris foncé (avec une amplitude d'environ

0.7 avec coefficient de réflexion de 0.7) désignent les composants spéculaires et des points gris clair (avec amplitude inférieure à 0.1) désignent la composants diffus. (Les points gris clair sont si densément placés de telle sorte qu'ils ressemblent à des lignes gris clair.) On peut noter que, pour une matrice de rétroaction de dimension 1106×1106, le nombre d'éléments non nuls est relativement faible. Il n'y a qu'un seul point gris foncé dans chaque rangée. Les zones blanches de forme carrée à la diagonale de la matrice de rétroaction représentent principalement la visibilité dans des directions discrétisées à ces taches qui se trouvent sur la même surface.

Résultats expérimentaux

La salle 1 est une salle rectangulaire avec la dimension $L = [l_x, l_y, l_z] = [15, 20, 5]$. Les surfaces sont discrétisées en 152 patchs carrés, chacun de taille 2.5×2.5m. La source est placée à $\mathbf{x}_S = [5.78, 14.91, 3.27]$, et le récepteur à $\mathbf{x}_R = [8.23, 11.92, 2.02]$. Toutes les surfaces possèdent un coefficient d'absorption $\alpha = 0.3$ indépendant de la fréquence, ce qui correspond au coefficient de réflexion $\beta = \sqrt{1-\alpha}$.

Les résultats de simulation sont présentés dans la figure 6.10.

Comme prévu, le RIR simulé a des premières réflexions distinctes au cours des premiers 50ms. Les transits de RIR simulées à diffuser comme des impulsions deviennent dense progressivement, tandis que l'énergie des réflexions spéculaires distinctes diminue. Ceci est conforme aux propriétés de la propagation du son. Observant soigneusement le zoom pour les premiers échos, nous pouvons voir que la plupart des premières réflexions sont modélisés avec une bonne précision.

Les résultats de simulation pour une chambre rectangulaire de dimension $L = [l_x, l_y, l_z] = [4, 6, 4]$ sont présentés dans la figure 6.12.

D'une manière générale, les premiers échos sont bien modélisés et les impulsions de la réverbération tardive est dense et décroît de façon exponentielle. Cependant, il y a un peu, au moins une réflexion de l'ordre première et une réflexion d'ordre secondaire, qui sont largement sous-estimés. En comparaison avec les résultats de simulation de la salle 1, il est une erreur plus grande. Cela provient principalement de la discrétisation grossière de la géométrie.

Le coefficient d'absorption varie de $\alpha = 0.2$ à $\alpha = 0.6$.

Tableau 1 montre les T_{60} prédits en utilisant le modèle ARN. Les résultats de prédiction en général sont d'accord avec les prédictions Sabine. Il est encore une fois la salle 2 qui a une plus grande erreur de prédiction. Cela provient principalement de la discrétisation grossière de la géométrie.

Les tests informels d'écoute révèlent que la qualité sonore modélisé par ARN a une améliora-

Table 1: Prédiction du temps de réverbération.

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

tion considérable par rapport aux FDNs traditionnels. Les premières réflexions sont perceptibles qui améliore la clarté du son réverbérant. La qualité du son est comparable à celle de la méthod hybride et peut être utilisé comme une alternative à la méthode hybride.

En outre, en raison du nombre élevé de lignes à retard et la permutation aléatoire entre la ligne des commentaires de retard, la réverbération tardive est de satisfaire la qualité de son avec son réverbéré lisse. Si le nombre de lignes à retard est réduite, une telle structure de rétroaction avec permutation aléatoire et la matrice de rétroaction clairsemée peut également être utilisé comme un dispositif de réverbération artificielle de remplacement même que pour modéliser la réverbération tardive.

0.7 Conclusions et Travaux Futurs

Contributions

La contribution principale de cette thèse a été mis sur l'exploration de la réverbération artificielle géométrique basée. Dans la plupart des travaux précédents sur la modélisation acoustique, méthodes géométriques et réverbération artificiels sont souvent étudiés séparément ou sont utilisés de manière hybride où chaque méthode est utilisée pour simuler une région temps-fréquence spécifique de la réponse impulsionnelle. Ainsi très peu l'accent a été mis sur l'interprétation physique des structures et des paramètres des réverbération artificiels, tels que la matrice de rétroaction, les retards et les unités filtres d'atténuation. Dans cette thèse, nous avons mis l'accent sur l'étude des liens entre les modèles physiques et les réverbérations artificiels. La signification physique des différents paramètres dans les filtres réverbérant artificiels sont étudiés. Modèles acoustiques géométriques, y compris la méthode de transfert de rayonnement et l'équation de rendu acous-

tique, sont utilisés pour relier les deux catégories de méthodes, et ont été incorporés dans les structures de rétroaction récursives pour estimer leurs paramètres. Plus précisément, la méthode de transfert de rayonnement a été utilisé pour estimer les paramètres des réverbères artificiels pour modéliser les réflexions diffuses, et le rendu acoustique équation étend en outre les structures de rétroaction récursives et leur permet de modéliser à la fois spéculaire et réflexions diffuses.

La contribution de cette thèse comprend les techniques suivantes:

- Une méthode d'estimation des paramètres pour les réseaux récursifs de ligne à retard et un filtre réverbérant géométrique basée.
- Une mise en œuvre pratique des retards variables pour réverbération artificiels et une nouvelle structure rétroaction de ligne à retard récursive pour améliorer la qualité sonore des réseaux récursifs traditionnels.
- Un schéma de collecte finale uniforme pour l'équation de rendu acoustique.
- Un nouveau réseau de rendu acoustique qui génère à la fois des échos précoces / spéculaires et réflexions tardives / diffuses.
- Un système de rendu acoustique virtuelle pour les applications de social interactif dans la réalité virtuelle.

Travaux futurs

Basé sur les défis des sujets et la réalisation effectuée dans cette thèse, la recherche et le développement peuvent être réalisés à l'avenir dans plusieurs directions.

D'une manière générale, les techniques étudiées dans cette thèse, les techniques de modélisation des effets de diffraction dans les méthodes proposées, en particulier la reverberator géométrique à base et les réseaux de rendu acoustique, ne sont pas bien discuté. Dans les scènes acoustiques dynamiques à géométrie complexe, la modélisation du phénomène de diffraction est important, surtout pour les premiers échos. Etant donné que les lignes à retard des réseaux de rendu acoustique est étroitement liée aux taches superficielles de la géométrie, le noyau de réflexion et la matrice de rétroaction correspondant peut être étendu pour supporter un modèle de diffraction.

Pour la reverberator fin géométrique à base proposée dans le chapitre 3, le schéma de regroupement existant est basé sur le principe de l'énergie uniforme. D'autres extensions peuvent

enquêter sur d'autres schémas de regroupement, afin d'obtenir une conception de groupe plus flexible, et peut-être d'introduire des informations spatiales pour chaque groupe.

Dans certains environnements extrêmes, les fréquences sonores et modaux légèrement métalliques sont les effets sonores souhaités. Ainsi, il est nécessaire d'avoir un paradigme généralisé de choisir les longueurs de retard afin de modéliser les effets sonores de différents scénarios d'application.

Pour augmenter la précision de la modélisation des premiers échos en utilisant l'équation de rendu acoustique au chapitre 6, l'idée de pré-calcul de la subdivision et la construction de l'arbre de faisceau de la méthode de traçage de faisceau peut être utilisé pour optimiser la discrétisation de la géométrie . La limitation sera que cette discrétisation est seulement optimale à un endroit précis de la source. En utilisant la propriété de réciprocité, ce qui peut être appliquée dans des scénarios où au moins une des deux parties de communication est statique et l'autre est dynamique, comme un concert musical et un discours public. L'optimisation de discrétisation est également applicable aux réseaux de rendu acoustiques dans le chapitre 6.

Dans le système de rendu acoustique virtuelle pour l'application de réseautage social dans le chapitre 7, seulement la reverberator géométrique à base est actuellement intégrée pour modéliser les réflexions diffuses et la réverbération tardive. réflexions spéculaires et au début des échos peuvent être inclus en utilisant les réseaux de rendu acoustiques du chapitre 6. optimisation informatique doit être fait pour les réseaux de rendu acoustique, par exemple en utilisant moins de patches et en utilisant la faible densité de la matrice de rétroaction, afin de le rendre applicable pour le rendu en temps réel.

Introduction

An experienced acoustician first investigates the acoustic effects of a room by clapping in his hands and listen. This empirical method was often used to gain a first insight of the acoustics of the place and is very practical to perceive some of the acoustical qualities of a room. The same music, when played in different environments, such as a recording studio, a concert hall or in the open air, gives very different perceptual experiences. At the same time, a well-designed concert hall can be excellent to host an orchestra performance, but may not be suitable to deliver a public speech to an large audience. Therefore the effects of the room in which we perceive the sound have an important influence on what we hear, and then, influence our assessment of the listening environments.

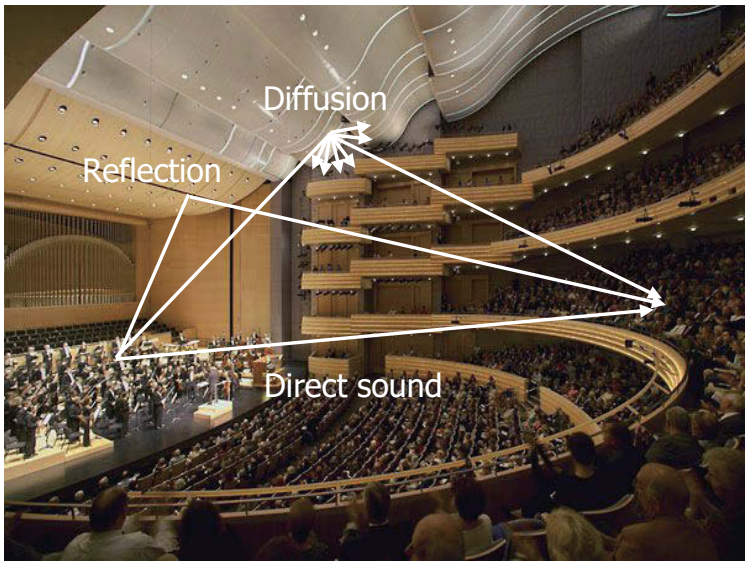


Figure 1: Illustration of acoustic effects in a concert hall of overture center for the arts, in Madison, Wisconsin, USA. Photo by Jeff Goldberg/Esto

What do we hear in a room ?

Sound perception in a room is affected by a number of factors. The originally played sound, ideally recorded in an anechoic studio without the reverberation effects of the environment, is referred to as dry signal. When the same sound is played in different rooms, a room effect, unique for the environment considered, is imposed onto the original signal. If we denote the original dry signal as x and what we finally hear as y, then their relationship can be described as:

$$y = h(x). \tag{7}$$

where h is an operator that gathers all the room effects imposed to x. It is reasonable to assume that the room effects h is independent of the signal played and is considered constant during the period of perception. That is to say, the room effects operator can be regarded as a Linear Time-Invariant (LTI) function.

The room effects operator h is determined by a number of factors related to the room and transportation medium:

$$h = f(\mathbf{x}_s, \mathbf{x}_r, \alpha, \mathcal{G}, \cdots), \tag{8}$$

where $\mathbf{x}_s$ and $\mathbf{x}_r$ are the positions of the sound source and the listener, α denotes the acoustic characteristics of the construction materials of the room surfaces, $\mathcal{G}$ denotes the shape and geometry of the room. There are also a number of other named and unnamed affecting factors, such as the directivity of the source and receiver, the furnishing level, the obstacles between the source and the receiver, the humidity and the temperature in the room, or even a person moving inside the room can affect the received room effect.

Can the sound bring us to a virtual environment ?

If the room effects can be quantitatively rebuilt, we start to wonder if we can rebuild the acoustic space virtually, to make our ears hear the sound in the virtually modeled environment, without physically being there. This arises within an interesting topic called virtual acoustics, which has become popular in research and in industry due to the constant demand of the evolution in virtual reality techniques.

Virtual acoustic techniques have made great progress in the last decade. Existing virtual acoustic systems include IKA-SIM by Ruhr-University Bochum in Germany [Blauert et al., 2000; Silzle et al., 2004], SLAB by NASA in U.S. [Miller and Wenzel, 2002; Wenzel et al., 2000], DIVA by Aalto University in Finland [Savioja et al., 1999], as well as the system by Boston University [Scarpaci and White, 2005].

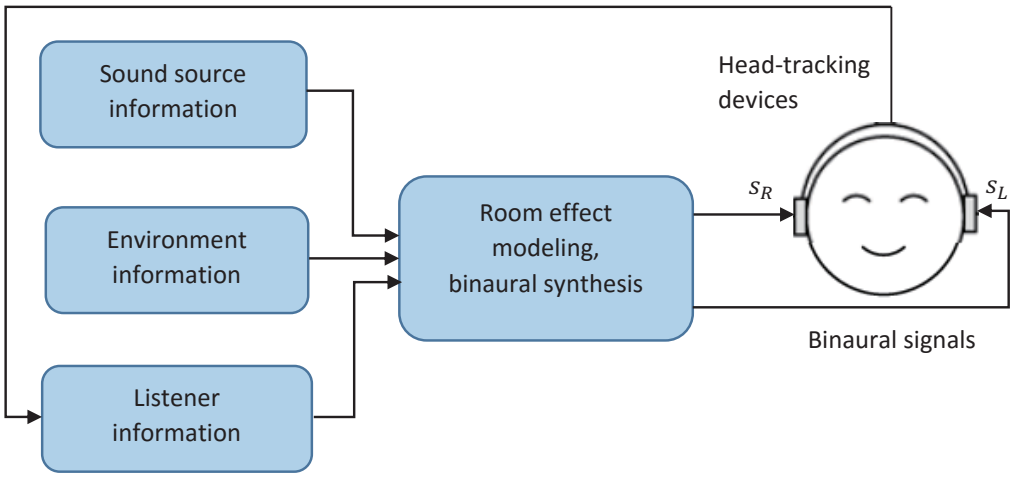


Figure 2: Structure of a dynamic virtual acoustic environment system.

Virtual acoustic environment systems recreate the auditory perceptions as if it would happen in real world by simulating acoustic environment artificially. The virtual acoustic environment system includes sound source modeling (directional sources), room reverberation modeling (direct sound, reflections, diffusion, reverberation) and listener modeling (individual scattering/diffraction caused by diversity of human anatomical structure). The information of the source and the receiver need to be monitored and updated regularly, as well as the room reverberation effects. Thus, a sophisticated virtual acoustic system should be interactive, dynamic, individualized and able to perform real-time rendering.

One of the challenges in virtual acoustic systems is how to simulate realistic room acoustic effects in an interactive manner for real-time rendering. Various methods exist for modeling virtual acoustic effects, from geometric-based methods to perception-based methods.

How to model the room acoustic ?

Many techniques have been proposed to compute the Room Impulse Responses (RIRs) in order to simulate the sound propagation within a room. They can be roughly divided into wave-based methods, ray-based geometric methods and statistical approaches.

Wave-based methods [Botteldooren, 1995; Savioja et al., 1995] solve the wave equations numerically and thus provide the most accurate results. However, they are also the most computationally demanding methods.

Ray-based geometric approaches rely on the physical principles of sound propagation, but with the simplifying assumption that sound waves travel as rays. The Image Source Method

(ISM) [Allen and Berkley, 1979] 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 computes sound rays outgoing from the source through the acoustic scene, and collects them at the receiver. The Radiance Transfer Method (RTM) [Lewers, 1993], originating from the light transport algorithm, efficiently models the diffuse reflections of RIRs and the sound energy decay of the late reverberation.

On the other hand, statistical methods, also referred as perception-based methods, use global acoustic parameters, such as reverberation time, to model the main perceptual aspects of reverberation. Schroeder filter [Schroeder, 1962] uses recursive comb filters and all-pass filters to simulate the dense echoes of the late reverberation. It is generalized into a structure called Feedback Delay Networks (FDNs) [Stautner and Puckette, 1982], which have been widely used to synthesize the late reverberation in hybrid methods.

However, different methods have their own strengthes and weakness in modeling room acoustics. For example, one method can be accurate at modeling the early reflections, but computationally inefficient for late reverberation modeling. This motivates many works which combined different methods in order to simulate a complete RIR. The focus of this dissertation is on the investigation and innovation of the current room acoustic modeling methods and their hybrid models, notably the hybrid methods combining radiance transfer methods and feedback delay networks.

Contribution of the dissertation on acoustic modeling

Although various acoustic modeling methods exist, we focus our study on their ability for interactivity and real-time computing. Special focus is put on the radiance transfer methods and feedback delay networks. The main advantage of the radiance transfer method is that it decouples the source and receiver's information from the bulk of real-time computation procedure, which makes it suitable for acoustic modeling in dynamic and interactive scenarios. The feedback delay networks, due to its computational efficiency, is suitable for real-time rendering of the late reverberation.

In late reverberation modeling, the feedback delay network is an efficient structure to generate exponentially decaying dense impulses which synthesize the energy decay trend of the late reverberation. However such a method relies on carefully choosing a bunch of parameters, estimated by either a pre-recorded impulse response from the real acoustic scene, or by manual setting from the experience of acousticians. At the same time, the exchange of sound radiosity modeled by the radiance transfer method can be regarded as a huge feedback delay networks where radiosity of each patch is fed back to other patches inside the enclosure. Thus, the first

contribution of the dissertation is **to combine the two methods and to relate the parameters of the FDNs with the radiosity exchange of the radiance transfer method.** We present a geometric-related late reverberator. The geometric-related late reverberator inherits the computational efficiency of the FDNs structure, but estimates the parameters of the FDNs directly from the virtual geometry setting. The sound radiosity exchange in the radiance transfer method is mapped to the delay lines of the feedback structure. This creates a new way to estimate the parameters of the FDNs for late reverberation simulation.

The radiance transfer method was extended in a general form in [Siltanen et al., 2007], named Acoustic Rendering Equation (ARE), which is able to model not only diffuse reflections by also specular reflections. This is done by discretizing the reflection angles of each patch. This method is very efficient for modeling the late reverberation and predicting the acoustic parameters. However, the accuracy of the modeled early reflections is not fully discussed in the literature. In the second contribution, **the modeling accuracy of the early reflections using the acoustic rendering equation is studied.** The final-gathering procedure is especially investigated and an uniform final gathering scheme is proposed. Two final gathering schemes were compared with respect to their modeling accuracy of early echoes. Using the uniform final gathering scheme, a single generic method can efficiently model both the early and the late reverberation of the RIRs.

The last contribution is devoted to **the combination of the acoustic rendering equation with the feedback delay networks to model early and specular reflections.** In the ARE model, the sound reflections between surface patches along the discretized direction can be regarded as feedback loops in the feedback delay networks. Thus, it is natural to incorporate the ARE model to the FDNs for the synthesis of room reverberation. However, some modifications and simplifications need to be done to the ARE model. First, the directional discretization scheme is modified in order to relate the computational elements of the ARE with the feedback delay lines. Second, the reflection kernel is simplified to keep the proposed feedback delay networks stable. The simplified acoustic rendering equation, which models both specular and diffuse reflections, is incorporated within the FDNs structure. Thus the hybrid reverberator, first designed to model the diffuse and late reverberation, is also capable of simulating the early/specular reflections with accuracy. This work is among the very few solutions which are capable to simulate early reflections using feedback delay networks.

The outcomes of the works in this thesis are published in the journal and conference articles:

- [Bai et al., 2013] H. Bai, G. Richard, and L. Daudet. Modeling early refections 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)*, pages 1–4, New Paltz,

NY, October 2013.

- [Bai et al., 2015b] H. Bai, G. Richard, and L. Daudet. Late reverberation synthesis: From radiance transfer to feedback delay networks. *Audio, Speech, and Language Processing, IEEE/ACM Transactions on*, 23(12):2260–2271, 2015.

- [Bai et al., 2015a] H. Bai, G. Richard, and L. Daudet. Geometric-based reverberator using acoustic rendering networks. In *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 1–4, New Paltz, NY, October 2015.

- H. Bai, G. Richard, and L. Daudet. Reflection-based variable delays in feedback delay networks. will be submitted to *Journal of the Audio Engineering Society - special issue*, 2016.

Organization of the document

This document is composed of seven chapters. Chapter 1 introduces the background knowledge of the room acoustics which is essential for the explanation in the rest of the dissertation. The different presentations used to characterize the acoustical propagation between the source and the receiver are described. The perceptual aspects related to sound and room acoustics are discussed.

In Chapter 2, the state of the art of the acoustic modeling techniques is overviewed by category. The modeling techniques are divided into wave-based methods, ray-based geometric methods and perception-based methods. We focus our discussion on the ray-based methods and the perception-based methods. The pros and cons of each method are presented and a discussion on the potential in combining different modeling techniques is given.

Chapter 3 presents the proposed hybrid method which combines the diffuse version of the radiance transfer method and the feedback delay networks. The basics of the acoustic radiance transfer method and its computation procedure are given. The idea of the link between the feedback delay networks and the acoustic transfer method is elaborated. The hybrid geometrical-based statistical modeling paradigm is then presented, with the design procedure of the corresponding parameter estimation. A special focus is put on the design of the feedback matrix which is essential for stable feedback delay networks. The method is evaluated in some typical room models, from both numerical and perceptual perspectives.

Chapter 5 investigates the acoustic rendering equation in its performance in early reflection modeling. The basics of the ARE method are reviewed with a discussion on the modeling accuracy of the early echoes. Two final-gathering schemes are compared. The uniform final-gathering

scheme is evaluated in its performance in modeling the early reflections using a rectangular room with different discretization densities.

In Chapter 6, the acoustic rendering networks are presented. The pros and cons of the feedback delay networks are firstly summarized, which permits us to motivate an hybrid approach by integrating the ARE model in the acoustic rendering networks. The modification and simplification of the ARE model are explained and the new geometric-based acoustic rendering networks are presented. The paradigm of how the parameters of the networks are related to the geometry of the room are then detailed. The simulation results are given to evaluate its modeling performance.

Chapter 7 discusses the design of systems for creating a virtual acoustic environment in real time, and links the techniques discussed in the previous chapters with their potential applications.

Finally in Chapter 8, we summarize our contribution in acoustic modeling techniques. The perspectives of extending our work in real-time virtual acoustic rendering are discussed.

Chapter 1

Representation of Sound

In this chapter, we introduce a few physical quantities essential for the understanding of sound propagation and room acoustics. We describe the different concepts used to characterize the acoustical propagation between the source and the receiver. We also describe some essential aspects in room acoustic perception, which often form the guiding principles of spatialisation and auralization techniques.

1.1 The Physics of Sound

1.1.1 Sound propagation

The generation of a sound wave can be modeled by a mass spring system [Cadoz et al., 1993]. When one mass (for example the membrane of a loudspeaker) is displaced by an excitation, it takes kinetic energy and vibrates about its mean position. The vibration causes the neighboring medium to be compressed. The compressed medium stores potential energy and transfers it to its neighboring medium along its propagation direction, and so on, to continue the propagation of the sound wave.

Under the influence of a sound wave, the medium (for example, the air particles) is compressed and decompressed, as shown in Figure 1.1. It causes the medium density and pressure vary rapidly as functions of time and space. The change of the pressure subjects to a temporal and spatial harmonic variation based on a number of physical laws [Kuttruff, 2009]:

$$p(x,t) = \hat{p}\exp[i(\omega t - kx)], \tag{1.1}$$

where $\hat{p}$ is a constant denoting the mean amplitude of the pressure, ω denotes the angular fre-

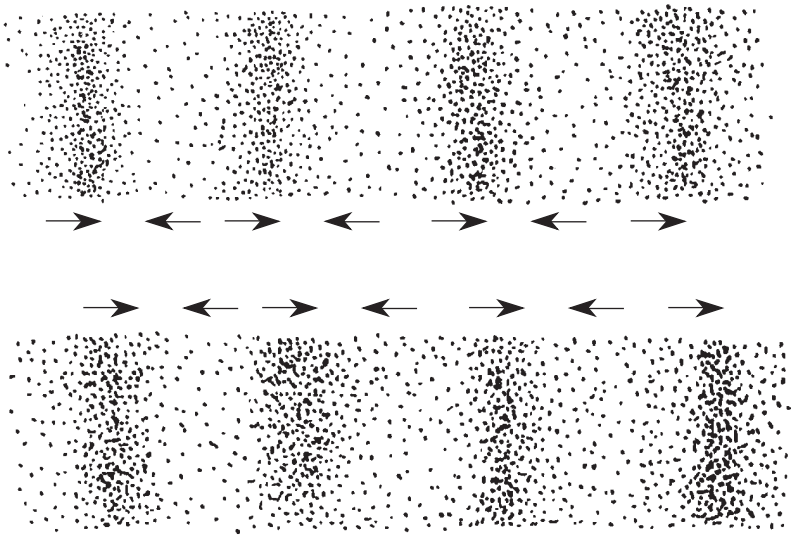


Figure 1.1: Top: the air particles are compressed in some region and spread out in others when sound wave propagates. Bottom: a instant later when the sound propagates slightly to the right, the particle density changes [Everest et al., 2001].

quency of the wave, k is a real constant called propagation constant or the wave number. Equation 1.1 is the wave equation in one dimension, and represents the sound propagation in the positive x-direction.

1.1.2 Sound pressure, intensity and radiance

The propagation of sound wave causes the gas pressure changes rapidly. In room acoustics, the **sound pressure** p at the measurement point is defined as the difference between the instantaneous pressure and the static pressure p_0:

$$p = p_{\text{ins}} - p_0. \tag{1.2}$$

The fluctuation of sound pressure from its static level is much smaller than the static pressure. The static pressure p_0 is $1.013 \cdot 10^5$ Pa in normal laboratory conditions, whereas the sound pressure p rarely exceeds 10 Pa.

The normally rapidly oscillating sound pressure p is usually more useful for theoretical analysis than for practical purpose. In practice the Root Mean Square (RMS) of the sound pressure is often used, which is the square root of the average of the square of the pressure of the sound signal over a given duration. This is also what we can read on our measurement tools (or what our ears perceive).

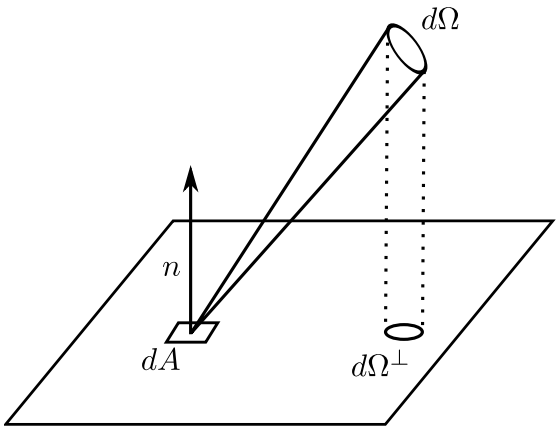


Figure 1.2: Illustration of irradiance and radiance definition.

$$p_{\text{rms}} = \bar{p} = \sqrt{\frac{1}{T}\int_0^T p^2(t)dt}. \tag{1.3}$$

A more frequently used quantity to describe the strength of a sound is the **sound intensity**. It is defined as the rate of the sound energy flowing across a surface with unit area perpendicular to the wave direction. It equals the time average of the product of pressure and the particle velocity:

$$I = \bar{p}\bar{v} \propto \bar{p}^2. \tag{1.4}$$

Sound intensity is also related to the Irradiance and Radiance, which are used to describe the energy flowing when the sound hits a surface.

When the sound hits a surface, there is a certain amount of energy flow to that surface. The amount of the energy flowing to the surface is related to the power of the incident sound, the incident angle and the area of the surface. **Irradiance** $I(\mathbf{x}, \Omega)$ is the incident power on surface $\mathbf{x}$ per surface area:

$$I(\mathbf{x}, \Omega) = \frac{\mathrm{d}\Phi}{\mathrm{d}A}\cos\Omega \tag{1.5}$$

where Ω is the angle between the normal $\mathbf{n}$ to the surface and the measured intensity, $\mathbf{x} = (x, y, z)$ is the coordinates of the radiation point on the surface, $d\Phi$ is the power incident on the surface dA, as shown in Figure 1.2. The irradiance is equal to the sound intensity when the incident direction is perpendicular to the surface.

Radiance $L(\mathbf{x}, \Omega)$ is defined as the radiant power per projected solid angle per unit area.

Radiance is a function of position and orientation:

$$L(\mathbf{x}, \Omega) = \frac{\mathrm{d}^2\Phi}{\mathrm{d}\Omega^{\perp}\mathrm{d}A} = \frac{\mathrm{d}^2\Phi}{\mathrm{d}\Omega\cos\theta\mathrm{d}A} = \frac{\mathrm{d}I}{\mathrm{d}\Omega\cos\theta}. \tag{1.6}$$

The acoustic radiance defined above has the following properties:

- The radiance going from point A to point B equals the radiance incoming at point B from the direction of point A.
- The radiance does not depend on the distance.
- If the radiance is constant, the energy flowing to a surface with unit area diminishes with the square of the distance.

1.1.3 Velocity of sound

When the sound wave is propagated in the medium, the propagation speed is called the velocity of sound. The speed of propagation depends on the characteristics of the particles of the medium. In the air, the velocity of sound can be approximated by [ISO9613-3, 1996]:

$$c = 343.2\sqrt{\frac{273.15 + \mathrm{K}}{293.15}}, \tag{1.7}$$

where c is the velocity of sound in the air in $[\mathrm{m}\cdot\ \mathrm{s}^{-1}]$, and K is the temperature in degrees Celsius.

1.1.4 Sound reflections and diffraction

The sound propagation mechanisms normally include specular reflection, diffusion and diffraction.

Specular reflection An acoustic travelling wave can be specularly reflected by a solid surface. At the boundary of the surface, the sound travel direction is changed. The amplitude of the sound wave is attenuated as part of its energy is absorbed by the surface material. There is a phase change to the reflected sound, which is due to the reflection characteristics of the surface material.

For a plane wave propagating along the positive x-direction, as represented in Equation 1.1, the reflected sound is:

$$p_r(x,t) = R\hat{p}\,\exp[i(\omega t + kx)], \tag{1.8}$$

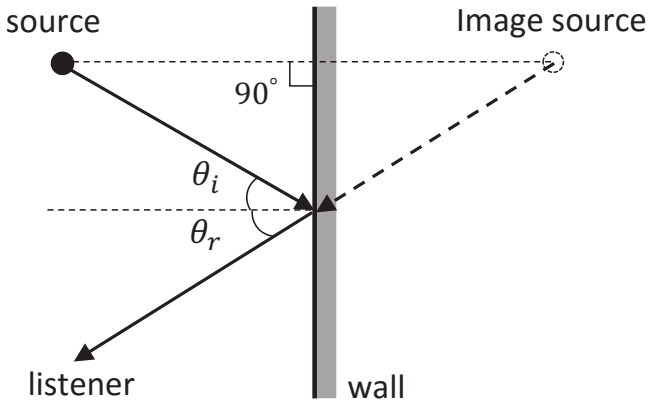


Figure 1.3: Illustration of the specular reflection of sound wave. This can be viewed as a virtual sound source, called "image source", emitting in free field. The angle of incidence θ_i equals the angle of reflection θ_r.

where R is the reflection factor, which describes the amplitude attenuation and phase change between the reflected sound wave and the incident sound wave. The sign of the x term is inversed as the reflected sound travels in the opposite direction of the incident sound.

Specular reflections behave in the same way as light reflections from a mirror, where the angle of the reflected sound with respect to the surface normal equals the angle of the incident sound:

$$\theta_i = \theta_r, \tag{1.9}$$

as shown in Figure 1.3.

In two dimensional space, the original sound wave and the reflected sound wave in Figure 1.3 can be described as:

$$p_i(x,t) = \hat{p} \exp\left[i\left(\omega t - k\left(x\cos\theta_i - y\sin\theta_i\right)\right)\right], \tag{1.10}$$

$$p_r(x,t) = R\hat{p} \exp\left[i\left(\omega t - k\left(-x\cos\theta_r - y\sin\theta_r\right)\right)\right]. \tag{1.11}$$

In near field of the reflection surface, specular reflection can also result in standing waves.

Diffusion A wave is scattered when it encounters a surface which has variations on the scale of the wavelengths of the incident wave. Ideally, when a sound is diffusely reflected by a surface, its reflected energy is distributed uniformly in all directions of the space. Thus a perfectly diffusive sound space is one that has certain key acoustic properties which are the same anywhere in the

space. For a good concert hall, reverberation should be diffuse in order to avoid "standing waves" [Smith, 2010] and coloration.

Diffraction In a sound propagation medium, sound wave-fronts travel in straight lines, except when they encounter an obstacle which has a comparable size to the wavelength of the incoming sound. Obstacles can cause a sound to change direction from its original rectilinear path. Low-frequency sounds diffract more than high-frequency sounds.

Several quantitative descriptions are defined to describe in a macro sense the sound reflections and diffusion. When a sound hits a reflective surface, the amount of reflected sound, both specular and diffuse, is given by the reflection coefficient which is the ratio of the pressure between the reflected sound and the incident sound. The amplitude of the reflection coefficient is:

$$R^2 = \frac{I_r}{I_i}, \tag{1.12}$$

where I_i is the intensity of the incident sound, I_r is the intensity of the reflected sound.

The amount of energy absorbed by the surface material is given by the absorption coefficient denoted as α:

$$\alpha = 1 - R^2. \tag{1.13}$$

The amount of energy which is diffusely reflected, among all the reflected energy, is defined as the scattering coefficient s. Using the conservation of energy, the sound reflection and absorption can be described as:

$$\alpha + s(1-\alpha) + (1-s)(1-\alpha) = 1. \tag{1.14}$$

The absorption coefficient of some commonly used materials can be found in [Vorländer, 2008].

1.2 Representation of Sound Propagation

The propagation of a sound within an enclosed space undergoes several reflections and diffraction at walls and obstacles. This physical property creates the so-called reverberation. The reverberation effects depend on the size and shape of the room, and the construction material of its surfaces. It sometimes also causes coloration effects by the repetition of sound caused by

reflections. The coloration depends on the position of the sound source and listener. Thus it is important to discuss the representation of the sound propagation for the purpose of reverberation effect analysis and sound rendering.

1.2.1 Room Impulse Response

The time-domain representation of the room-dependent sound propagation, between a sound source and a receiver, is classically characterized by the Room Impulse Response (RIR). It is the response of a room when excited by an ideal impulsion. The RIR is important to characterize the acoustic effects of the room. Its convolution with an anechoic sound produces a reverberant sound as perceived in a particular room. The RIR can be parameterized by the following factors:

$$h(t \mid \mathbf{x}_s, \mathbf{x}_r, \mathbb{k}), \tag{1.15}$$

where t denotes the temporal variable, $\mathbf{x}_s$ and $\mathbf{x}_r$ are the positions of the source and the receiver, and $\mathbb{k}$ denotes the room characteristics including its geometry and the surface materials of its walls.

Two main families of methods may be used to obtain a RIR. It is first possible to measure the response, using an impulsive sound, i.e., a gun shot or a balloon pop in an existing room. More accurate methods, such as Maximum Length Sequence (MLS) [Borish and Angell, 1983] or sweep methods [Müller and Massarani, 2001] are often performed for RIR measurements. The other method used to compute a RIR is through numerical simulation. Many techniques have been proposed to compute RIRs in order to simulate the sound propagation within a room. They can be roughly divided into wave-based methods, ray-based geometric methods and statistical methods. We present in Chapter 2 many algorithms that can be used to model a RIR.

Room impulse responses, depending on their time of arrival and their distinct perceptual cues provided, are classically decomposed into three parts: the direct sound, the Early Reflections (ER) and the Late Reverberation (LR), as shown in Figure 1.4.

Direct sound The direct sound is the first contribution that travels from the source to the receiver. Most of the time, the direct sound is processed apart from the rest of the RIR.

Early reflections The ER portion of the RIR is often taken to be the first 100ms or so [Moorer, 1979]. ER helps in localization and conveys spatial information about the environment. It is known that the early reflections have a strong influence on the listener's perception of the shape of listening-space [Dokmanić et al., 2011, 2013]. That is to say, we can "hear" the shape of the

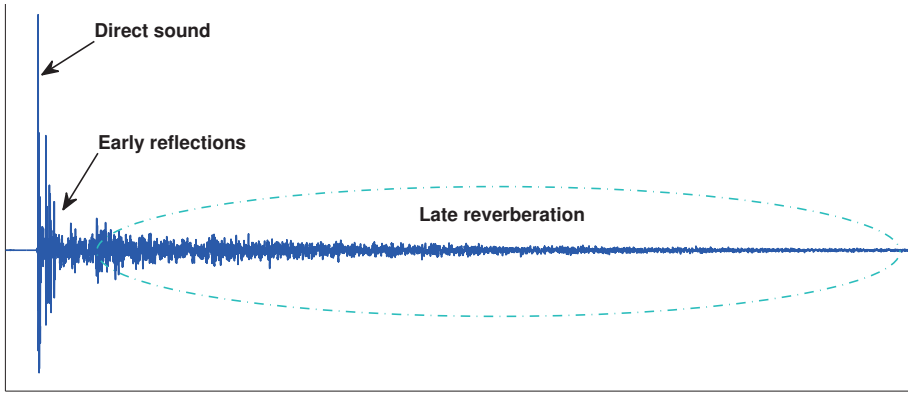


Figure 1.4: Illustration of room impulse response: direct sound, early reflections and late reverberation. The impulse response is from the Aachen Impulse Response Database [Jeub et al., 2009].

room by the aid of the early reflections.

Late reverberation LR enhances immersion, contributing to the perception of the size of the environment, its amount of furnishing and type of absorptivity [Kuttruff, 2009]. It however lacks a clear boundary between the early reflections and the late reverberation. One definition of when the RIR leaving the early reflections and entering into the late reverberation is when it begins to look Gaussian, since a random sum of plane waves uniformly distributed over all angles of arrival yields a Gaussian distributed pressure field [Smith, 2010]. Another desired property of "late reverberation" is that the remaining energy in the RIR, the Energy Decay Curve (EDC) [Schroeder, 1965], approaches to be an exponential decay.

1.2.2 Plenacoustic Functions

One would be interested to know what the room sounds like between any source and receiver's locations in the room. Plenacoustic functions, first mentioned in [Kubovy and Van Valkenburg, 2001], extend the notation of the room impulse response and represent the evolution of RIRs as a function of the three-dimensional spatial positions of source and receiver. Knowing this information, we can easily calculate the reverberant sound produced by the room from the source's position to the receiver's position. The plenacoustic function is thus parameterized by the following factors:

$$P_{\mathbb{k}}(t, \mathbf{x}_s, \mathbf{x}_r \mid \mathbb{k}), \tag{1.16}$$

where t denotes the temporal variable, $\mathbf{x}_s$ and $\mathbf{x}_r$ are the positions of the source and the receiver, and $\mathbb{k}$ denotes the characteristics of room.

Static acoustic scenes, if 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. The plenacoustic fucntions were firstly studied and analysized in [Ajdler and Vetterli, 2003a,b] to answer the question "how many microphones do we need to place in the space in order to completely reconstruct the sound field at any position in the space ?" In [Ajdler et al., 2006], the optimal sampling pattern in time and space was discussed and a quantitative sampling theorem, trading off sampling rate for SNR, has been presented.

Using the modal theory, [Mignot et al., 2011a] uses a Compressed Sensing (CS) approach to sample and to reconstruct the plenacoustic functions at low frequencies, with a number of sensors significantly lower than would be required by Shannon-Nyquist sampling theorem. In [Mignot et al., 2011b], the sparsity of the early part of the RIRs was exploited and the compressed sensing approach was applied to reconstruct the early part of plenacoustic functions. [Bianchi et al., 2013] utilised the deconvolution techniques from aerospace acoustic imaging to improve the resolution capabilities of plenacoustic imaging.

1.2.3 Echogram

The echogram, also referred as reflection diagram, is a temporal representation of the strengths of reflections [Kuttruff, 2009]. It equals the magnitude of the impulse response represented on a logarithmic scale. The echogram is used to represent the response of a room with energetic propagation algorithms such as ray tracing [Krokstad et al., 1968] and radiance transfer methods [Lewers, 1993; Nosal et al., 2004]. As the echogram represents the propagation of the energy, it does not contain phase information.

1.3 Perception of Room Acoustics

1.3.1 Perceived Sound Intensity

Human sensation of loudness corresponds to a logarithmic intensity scale rather than a linear one. For this reason, it is customary to compare the loudness of two sounds by the logarithm of the ratio of their intensity or pressure. The loudness level in free space is thus defined as:

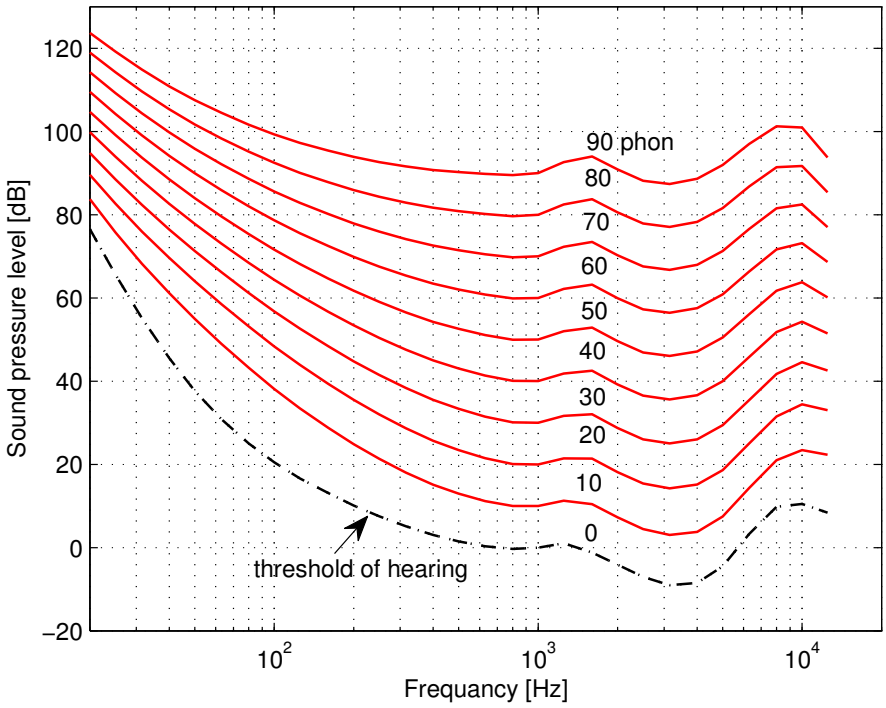


Figure 1.5: Curves of equal loudness level (data from [ISO226, 2003]). The dashed line corresponds to the average threshold of hearing.

$$L = 20 \log |\frac{\bar{p}_1}{\bar{p}_2}| = 10 \log \frac{I_1}{I_2} \text{ dB}, \tag{1.17}$$

where $\bar{p}_2$ is the RMS pressure and I_2 is the intensity of a reference sound. Usually, $\bar{p}_2 = 2 \cdot 10^{-5}$Pa which corresponds to the just audible pure sound at 1 kHz.

1.3.2 Frequency Perception of Sound

It is well known that the perception of frequency content is unequal across the spectrum. Figure 1.5 shows the curves of equal loudness level [Fletcher and Munson, 1933; ISO226, 2003]. They depict the sound pressure levels for different frequencies at equal perceived loudness.

The ear sensitivity reaches its highest value for frequencies in the range 600 to 5000 Hz. This frequency region corresponds to the most salient part of human speech spectra.

1.3.3 Spatial Perception

The perception of sound in 3D involves a number of cognitive mechanisms. The duplex theory [Strutt, 1907] uses interaural time delay (ITD) and interaural level difference (ILD) to model the human sound localization mechanism. The time difference between the onset of non-continuous (transient) sounds or the phase difference of more continuous sounds at both ears is known as ITD [Carlile, 1996]. The difference in level between the sound reaching both ears, mostly caused by the "acoustical shadow" effect of the head, is known as ILD.

However, the duplex theory alone is not able to explain how one can distinguish a number of ambiguous situations such as front-back confusions. For a sound located on the median plane, or directly above or below the listener, both ITD and ILD cues are (nearly) zero. In fact, when the sound reaches the listener, more complex interactions of sound can occur before it enters the eardrum, such as diffraction and attenuation by the head, shoulders, torso and particularly the outer ear. These interactions modify the spectrum and timing of a sound signal, and can be modeled as complex transfer functions called Head Related Transfer Functions (HRTF). The left and right ears' HRTFs H_L and H_R can be defined as:

$$H_L = \frac{\Phi_L(\omega, \delta, \phi, d)}{\Phi_f(\omega)}, \tag{1.18}$$

$$H_R = \frac{\Phi_R(\omega, \delta, \phi, d)}{\Phi_f(\omega)}, \tag{1.19}$$

where $\Phi_L(\omega, \delta, \phi, d)$ and $\Phi_R(\omega, \delta, \phi, d)$ are the sound pressure levels (SPL) at the eardrum of the left and right ears, $\Phi_f(\omega)$ is the reference SPL at the central position of the head when the head is absent. ω is the angular frequency, δ and ϕ are the azimuth and elevation angles and d is the distance from the listener to the source.

HRTFs provide the necessary information to perceive vertical directions and to disambiguate front-back confusions [Moore, 2012]. The problem with HRTF is that they are unique to each person. Much work has been done in order to find a general model to represent the head related characteristics.

1.3.4 Objective Parameters of Room Acoustics

Room impulse responses can be analyzed from many aspects, from time domain to frequency domain. In order to simplify the evaluation process for the acoustical characteristics of a room, several objective indicators have been proposed. Each of them covers a certain aspect of the perception of the acoustic quality of the room.

Reverberation Time

The sound intensity of the room impulse response, especially the late part, has an exponentially decaying trend. Reverberation time (RT), also referred as T_{60}, is a measure of the rate of decay of the intensity. It is defined as the time in seconds required for sound in a room to drop 60 dB from its original level.

The RT can be measured by a recording device. The device simply measures the time required for the sound to decay 60 dB. When the background noise is high, it can measure the time for the first 20-30 dB delay, and then extend it to 60 dB using the exponential decay assumption.

By assuming a perfectly diffused sound field, the Sabine equation relates the RT with the volume of the room and the absorption condition of the room [Kuttruff, 2009]:

$$T_{60} = \frac{0.161V}{A}, \tag{1.20}$$

where T_{60} is the reverberation time in second, V is the volume of the room in m^3, and A is the total equivalent area of absorption, which is calculated as

$$A = A_1\alpha_1 + A_2\alpha_2 + A_3\alpha_3 + \cdots, \tag{1.21}$$

where A_i is the area and α_i is the absorption coefficient for all the absorptive surfaces of the room.

The reverberation times for typical room types are 0.6-0.7 s for classrooms, 0.9-1.2 s for movie theaters and 1.7-2.2 s for concert halls.

Definition D_{50}

The definition is the percentage of energy reaching the listener during the 50 first milliseconds over total energy:

$$D_{50}(\%) = 100\left(\frac{\int_0^{0.05} p^2(t)dt}{\int_0^{+\infty} p^2(t)dt}\right). \tag{1.22}$$

It helps to characterize the intelligibility of speech. This information is essential for classroom or amphitheater modeling. In concert hall acoustics it refers to the degree to which individual strands in a musical presentation can be differentiated from each other.

Clarity C_{80}

The clarity expressed in dB is based on a ratio of energy during the first 80 milliseconds over the energy after 80 ms:

$$C_{80}(dB) = 10 \log \left(\frac{\int_0^{0.08} p^2(t)dt}{\int_{0.08}^{+\infty} p^2(t)dt} \right) \tag{1.23}$$

C_{80} is often used to assess the reverberation property for music signals. In an anechoic room where little/no reverberation is present, the perceived music will be very clear and C_{80} will have a large positive value. If the reverberation in the room is large and the listener is far away from the speaker, the sound will be unclear and C_{80} will have a relatively high negative value. C_{80} becomes 0 dB, if the early reflections and the late reverberation contain equal energy.

For orchestral music a C_{80} of 0 dB to -4 dB is often preferred, but for rehearsals conductors often express satisfaction about a C80 of 1 dB to 5 dB, because every detail can be heard. For singers, all values of clarity between +1 and +5 seem acceptable. C80 should be generally in the range of -4dB and +4dB.

For speech, in comparison to music, the clarity will be measured as the ratio of the first 50 ms (C_{50}) instead of 80 ms (C_{80}) for music.

Chapter 2

Acoustic Modeling Techniques

A sound emitted from the source may undergo several reflections and diffractions within the acoustic scene before it is heard by the listener. Realistic sound propagation modeling and rendering can add a whole new layer of realism to the visualization of the virtual environments.

The 3D acoustic scene can be synthesized from two perspectives: the physical approaches and the perceptual approaches. The physical approaches model the accurate sound propagation in the acoustic environment using the wave-based methods or the ray-based geometric methods. The perceptual approaches, often mentioned as statistical acoustics, on the other hand, use global acoustic parameters to model the main perceptual aspects of the environment.

In this chapter, the state of the art of the acoustic modeling techniques is overviewed by category. We focus our discussion on the ray-based methods and the statistical methods. We conclude this chapter by discussing the potential in combining different modeling techniques and the pros and cons of the existing hybrid models.

2.1 Wave-based Numerical Acoustics

Wave-based numerical methods are the most accurate modeling methods for the simulation of sound wave propagation in bounded environments. The direct numerical solution of the wave equation, achieved through the use of discrete element methods, provides a means of emulating the behavior of sound fields that inherently incorporates wave properties. At the same time, they are the most computationally-demanding methods too.

The currently used wave-based acoustic modeling methods can be classified into three categories: Finite Element Methods (FEM), Boundary Element Methods (BEM) and Finite-Difference Time-Domain methods (FDTD) [Botteldooren, 1995; Savioja et al., 1995]. In FEM, the complete

space considered need to be discretized into volumetric elements, while in BEM, only the boundary of the space are discretized. Typically, the requirements of computation resource and memory capacity for the FEM are higher than for the BEM. While the FEM and the BEM typically operate in the frequency domain, the FDTD simulates the room acoustic in the time domain. The main principle of the FDTD is that the derivatives in the wave equations are replaced by the corresponding finite differences [Strikwerda, 2004].

However, the wave-based methods are also the most computationally-demanding methods. Thus, in practice, they are usually only used for modeling the room acoustics in the low frequency range [Kleiner et al., 1993; Pietrzyk, 1998]. For real-time auralization, they are only used to model the direct sound and the early reflections. With consideration of simplifying the computation, the FDTD can be suitable in real-time auralization [Savioja, 2010].

Since the wave-based methods are not the focus of this thesis, interested authors are encouraged to refer to these recent advances [Kowalczyk and Van Walstijn, 2011; Sakamoto et al., 2006; Savioja et al., 2006; Southern et al., 2013; Svensson and Kristiansen, 2002; Van Walstijn and Kowalczyk, 2008].

2.2 Ray-based Geometric Acoustics

In geometrical room acoustics where the room size is large compared to the wavelengths of sound, the concept of a wave is of minor importance; it is replaced instead by the concept of a sound ray [Kuttruff, 2009]. This condition is often met in room acoustics. For example in a room with dimension of $4 \times 4 \times 3$, if we consider a sinusoid sound of frequency 1 kHz, its corresponding wavelength is 34 cm, which is much smaller than the dimension of the walls of the room. The sound ray, like the light ray, is a concept which simplifies the sound wave to the limiting case of vanishingly small wavelengths. The sound ray has well-defined direction of propagation, and is subject to specular and diffuse reflections when it hits the reflective surfaces. The limited velocity of sound requires that the propagation time cannot be neglected when simulating the propagation process using sound rays. Many acoustical effects are related to the propagation time, such as echoes, reverberations and so on.

Many techniques have been proposed to simulate the sound propagation using the concept of sound ray [Savioja and Svensson, 2015]. These techniques are referred to as geometric methods. In this section, we present the main concept of the classical geometric methods. Focus is given to the discussion on the difference between these techniques, and their pros and cons.

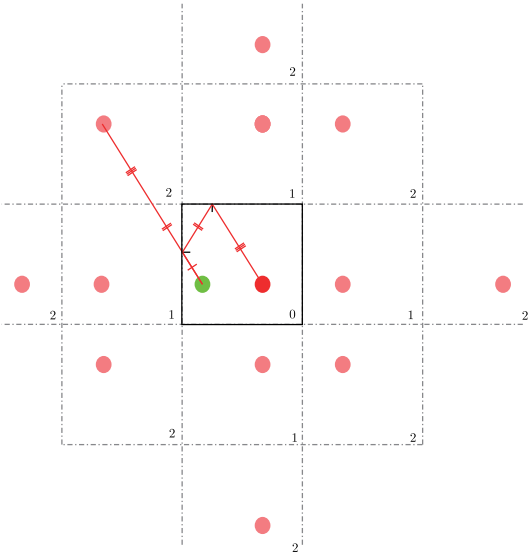


Figure 2.1: Illustration of the first and second order of virtual sources for image source method [Loyet, 2012]. The red dots are the sound source and its virtual sources, and the green dot is the receiver.

2.2.1 Image Source Method

The Image Source Method (ISM), first presented in [Allen and Berkley, 1979], is the mostly used geometric method to exhaustively find all the specular reflection paths from a source to a receiver. The specular reflection paths are computed by considering a set of virtual sources. The first order of virtual sources is obtained by considering the mirrored locations of the source with respect to each reflecting wall, as illustrated in Figure 2.1. The second order of image sources is obtained from the first order of virtual sources: each first-order virtual source is mirrored with respect to the reflecting walls to generate second-order virtual sources. The process continues until enough reflection paths are found, or the reflected sound energy is small enough to be neglected.

The advantage of the image source method is that it gives a totally exhaustive list of the specular reflection paths up to the defined reflection order. The main drawback is that its computational complexity increases exponentially with the order of reflection since the number of virtual sources also increases exponentially [Rindel, 2000]. The exact number of virtual sources N_S after N_r^{th} order of reflections is given by:

$$N_S = 1 + N_W + \sum_{n=2}^{N_r} (N_W - 1)^n, \tag{2.1}$$

where N_W is the number of rigid walls in the scene.

Another problem with the image source method is that, for more complex geometries than shoeboxes, not all the virtual sources generated using the mirroring positions are visible at the receiver. Thus, the visibility needs to be checked for each newly created virtual source. [Borish, 1984] proposed the use of the ray tracing method to verify visibility. When a virtual source is created, a virtual ray is created at the receiver's location and sent towards the direction of the real sound source. This virtual ray is propagated in the scene. If the virtual ray reaches the real source with the same order of reflections, this virtual source is determined as valid for the computation of the RIR.

Many extensions were introduced after the proposition of the original ISM. [Lee and Lee, 1988] improved the computational efficiency of the image source method by early discrimination of invisible virtual sources during computation. [Lehmann and Johansson, 2008] uses the ISM model to predict the energy decay trend of the late reverberation and a diffuse reverberator is proposed in [Lehmann et al., 2010] based on this study. [McGovern, 2009] proposes a fast image source method to calculate the RIRs of shoebox rooms. [Heinz, 1993] used ISM for Binaural auralization.

In general, the image source method, like all the other ray-based geometric methods, only holds for high frequencies. Besides, it does not allow the preservation of wave characteristics such as diffraction and diffusion. Based on its advantages and its shortcomings, in real-time auralization, the ISM is usually used to model the early specular refelctions, leaving the diffusion and the late reverberation to other methods such as radiance transfer methods (Section 2.2.2 and 3.1) and artificial reverberators (Section 2.3).

2.2.2 Deterministic Ray Tracing Methods

The ray tracing method [Krokstad et al., 1968; Kulowski, 1985] is one of the most important geometric methods in practice and has several variations. Starting from the main concept of the ray tracing method, many algorithms have been derived from it, among which:

- Stochastic ray tracing [Dalenbäck, 1996];
- Cone tracing [Martin et al., 1993];
- Sonel mapping [Kapralos et al., 2006];
- Phonon tracing algorithm [Bertram et al., 2005];

The basic ray tracing method emits sound rays from the source, each of which carries a certain amount of energy. The rays bounce between the surfaces in the scene according to the

pre-defined reflection manner, which are finally collected by the receiver. Although specular reflection is the commonly used reflection kernel, diffraction and diffusion can also be accounted for [Lehnert and Blauert, 1992; Naylor, 1993].

The sound rays are emitted from the source either in a pre-defined uniform pattern, or a randomized uniform pattern. A directional sound source can also be modeled by modifying the emission pattern of sound rays from the source. The receiver is modeled as a receiving volumetric object. It is usually a sphere, in order to have an omnidirectional collecting element. According to [Vorländer, 1989], the minimum number of rays used in a simulation for a sphere of radius r_L can be approximated by:

$$N_r \geq \frac{4(T_h c)^2}{r_L^2}, \tag{2.2}$$

where T_h is the length of the RIR we intend to model using ray tracing and c is the sound velocity. To generate a RIR of duration 1 s using a sphere of radius 1 m as receiving volume, at least $4.5 \cdot 10^6$ rays are needed. [Lesoinne and Embrechts, 2008] use a time-varying receiving volume which progressively modifies the receptor size as the ray's travel grows.

Compared with the exponential increase of the computation load for the ISM, the computational complexity increases linearly with the order of reflections. However, large number of rays need to be used in order to reduce the systematic error due to the finite number of rays [Kulowski, 1982]. Note that the ray tracing method is source and receiver dependent. When the source or the receiver moves, the rays need to be emitted and traced again. These computational constrains make such methods inappropriate for real-time interactive applications.

2.2.3 Beam Tracing

Beam tracing method [Funkhouser et al., 1998; Heckbert and Hanrahan, 1984] is an optimal solution to the problem of finding the early specular reflections [Laine et al., 2009]. It produces exactly the same results as the image source method, while optimizing the visibility calculation so that only the image sources which are part of the valid reflection path are considered [Funkhouser et al., 1998].

The beam tracing method pre-computes the spatial subdivision, where the 3D space is partitioned into convex polyhedral regions, as illustrated in Figure 2.2. A beam tracing algorithm recursively traces the pyramidal beams through the spatial subdivision to construct a beam tree data structure representing the region of space reachable by each potential transmission sequence and specular reflection path.

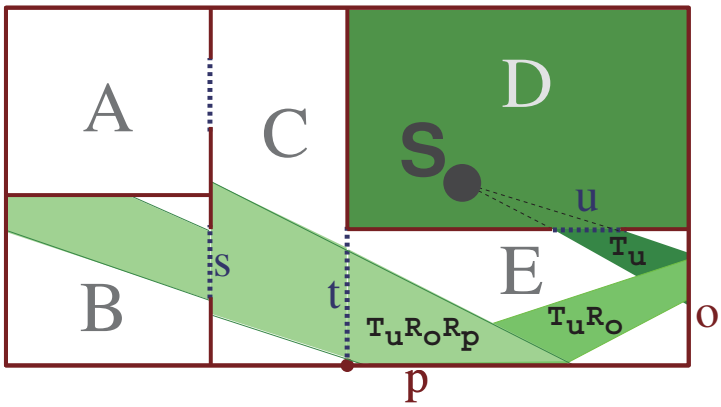


Figure 2.2: Illustration of the spatial subdivision of the beam tracing method [Funkhouser et al., 1998]. The sound first reaches cell E after leaving cell D. It is reflected in cell E and enters cell C and finally cell B.

The original beam tracing handles low-order specular reflections for fixed sources. [Tsingos et al., 2001] extended it to include diffraction via the geometric theory of diffraction, for edges which are much larger than the wavelength. An accelerated beam tracing algorithms is optimised by using the spatial coherence in path validation [Laine et al., 2009]. [Chandak et al., 2008] present an interactive algorithm which supports moving sources and real-time update rate for complex scenes. [Antonacci et al., 2008; Markovic et al., 2016] use an iterative lookup of a special data structure that describes the global visibility between reflectors to speed up the generation of the beam tree and is designed for real-time acoustic rendering where both the receiver and the source are moving.

2.2.4 Radiance Transfer Method

Radiance Transfer Methods (RTM) have been firstly used in computer graphics to model the diffuse reflection of the radiosity between patches [Goral et al., 1984]. Basically, the radiance transfer methods consider every patch a potential emitter and reflector of radiosity. Radiance transfer methods were firstly applied in room acoustic modeling in [Lewers, 1993], where the author used the radiance method to model the diffuse reverberant tail of the impulse response in a hybrid model. This method is further explored by [Nosal et al., 2004] to support arbitrary polyhedral geometries. After that, the radiance transfer methods have been extensively used to model the diffuse reflections and late reverberation in room acoustic modeling [Antani et al., 2012a,b; Hodgson and Nosal, 2006; Koutsouris et al., 2013a].

The basic idea of the radiance transfer method is to dicretize the surfaces into patches and

utilise the energy exchange between patches to simulate the propagation of the diffuse radiance. For each pair of patches A and B, a patch-to-patch energy response is computed which measures the fraction of the radiant energy leaving patch A that arrives at patch B (detailed in Section 3.1). A set of propagation equations can be yielded which are used to solve the radiance of each patch and the energy response from the source to the receiver, via the diffuse reflection patches.

The main advantage is that the patch-to-patch interactions formulated by the radiance transfer method are independent of the source and receiver positions. By pre-computing a linear operator that defines how a sound emitted from the surface patch affects the radiance of other patches, the high-order patch-to-patch reflections can be decoupled from the run-time computation. During the run-time computation, the high-order patch-to-patch interactions are summed up, with additional attenuations and delays related to the locations of the source and receiver, to simulate the whole the sound propagation.

[Siltanen et al., 2007] reformulated the radiance transfer method in a more general form, named Acoustic Rendering Equation (ARE), which supports the modeling of specular reflections. It is done by discretizing the reflection directions of each patch into solid angles, as illustrated in Figure 2.3. The incoming radiosity of a patch can be specularly and/or diffusely reflected along the pre-defined solid angles (detailed in Section 5.1).

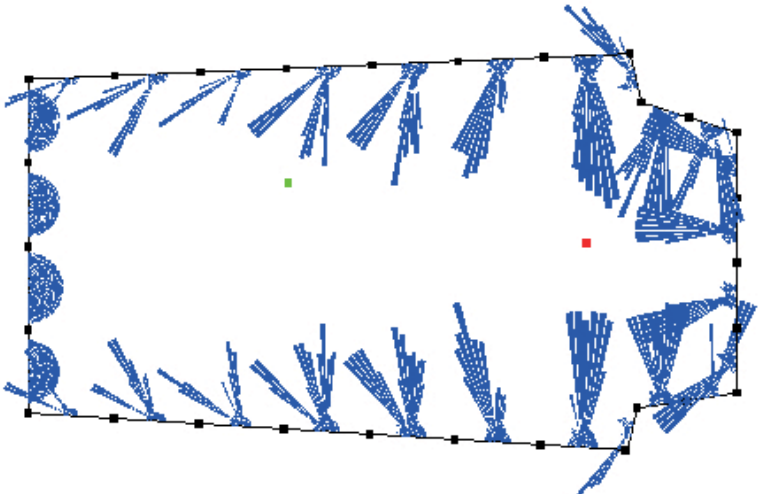


Figure 2.3: Illustration of the acoustic rendering equation [Siltanen et al., 2010a]. The reflection directions are discretized into solid angles and the sound emitted from the source (the red dot) propagates along the solid angles between patches before it arrives at the receiver (the green dot).

Although the acoustic rendering equation is able to model general sound propagation, like all ray-based methods, it is unable to recover the phase information of RIRs. The main limitation of

the method is its high requirement in terms of computation and memory capacity. Depending on the complexity of the geometry, the radiance transfer method requires hundreds to thousands of convolutions at run-time computation to synthesize the room impulse response. [Siltanen et al., 2009] use the radiance transfer method in the frequency domain to synthesize the impulse response. [Antani et al., 2012a] exploit the sparsity of the frequency-domain patch-to-patch transfer operator matrix and use a Karhunen-Loeve transform to compactly represent the sound propagation. [Siltanen and Lokki, 2008; Siltanen et al., 2010b] extend the method to support diffraction modeling. Other improvements have been made to the radiance transfer method, including the extension to model the time-dependent reflections [Siltanen et al., 2012; Southern et al., 2013], the improvement in computational efficiency using GPU [Savioja et al., 2010], and various hybrid methods [Southern et al., 2011, 2013].

The radiance transfer method will be further discussed in Section 3.1, and the complete formulation of the acoustic rendering equation will be given in detail in Section 5.1.

2.3 Perception-based Statistical Acoustics

The statistical acoustic modeling methods are often referred to as artificial reverberators [Välimäki et al., 2012; Välimäki et al., 2016]. The observation of the noise-like characteristics of the late reverberation of the acoustic room impulse responses motivated the study of modeling the late reverberation using statistical methods. Statistical methods use acoustic parameters such as reverberation time to mimic the acoustic effects of a scene. Statistical methods have gained great popularity, especially in interactive virtual reality applications such as video games. In general, they produce perceptually relevant acoustic effects, which satisfy the listening experience in interactive application scenarios.

2.3.1 Schroeder Filter

The landmark work in artificial reverberation was done by Schroeder [Schroeder, 1962] more than fifty years ago, where he introduced recursive comb filters and allpass filters to simulate the multiple echoes. Moorer further improved this algorithm [Moorer, 1979].

The Schroeder filter utilises parallel comb filters to produce exponential decaying pulses to model the reverberant effects of rooms. The building block is a comb filter, as shown in Figure 2.4, and has the following transfer function and frequency response:

$$H(z) = \frac{z^{-d}}{1 - \rho z^{-d}}, \tag{2.3}$$

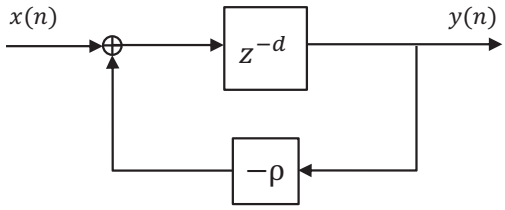


Figure 2.4: The comb filter: building block of Schroeder filter.

$$|H(\omega)|^2 = \frac{1}{1+\rho^2-2\rho\cos\omega d}, \tag{2.4}$$

where $\omega = 2n\pi/d$ ($n = 0, 1, 2, \cdots$).

A single comb filter has a comb-shaped frequency response with its maxima of $1/(1-\rho)$ and minima of $1/(1+\rho)$. In order to obtain a flat frequency response, several comb filters, normally at least 6, are connected in parallel. The gaps of the comb-shaped frequency response are filled by the comb filters with different delay length.

Another issue with Schroeder filters is that the produced echo density is too low compared to the real room. This leads to a "fluttering" of the reverberated sound. In order to solve this issue, a series of all-pass filters are usually connected after the comb filters to increase the density of echoes.

2.3.2 Feedback Delay Networks

In [Stautner and Puckette, 1982] a structure well suited for artificial reverberation called "feedback delay networks (FDNs)" was introduced. This structures is characterized by a set of delay lines connected in a feedback loop through a feedback matrix, as illustrated in Figure 2.5. The signals of all delay lines interact through the feedback matrix, which increases the density of echoes for the late reverberation. Jot [Jot, 1992a,b; Jot and Chaigne, 1991] further studied the FDNs and developed associated techniques for designing good quality reverberators.

For artificial reverberation, a few considerations are taken into account as the designing criteria [Rocchesso and Smith, 1997b]:

- Precise modeling of early reflections;
- Minimal coloration due to uneven mode distribution in the frequency domain;
- Smooth change of decaying speed at all frequencies;

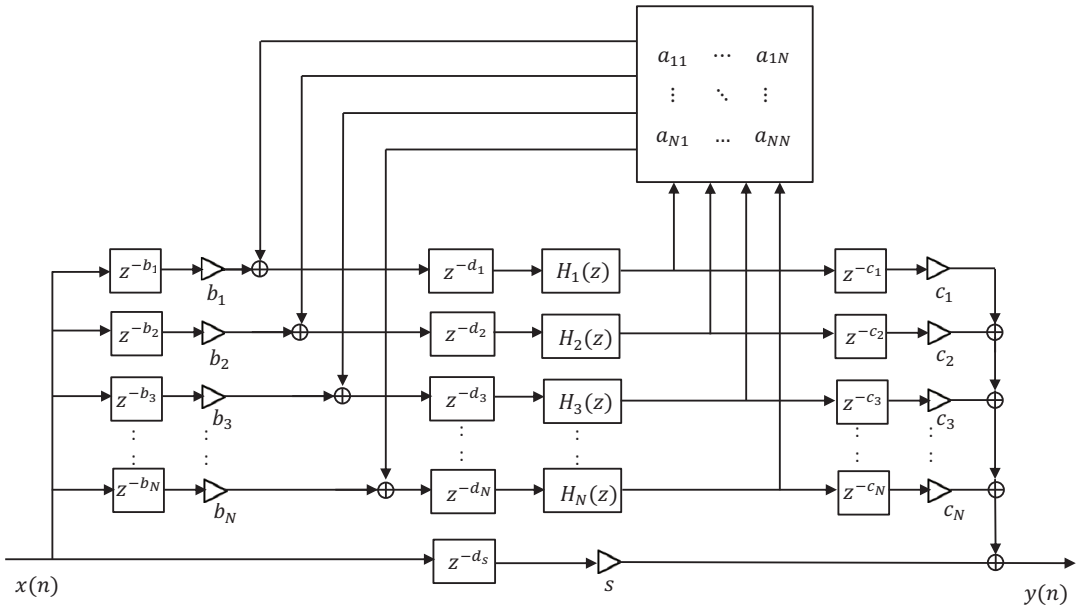


Figure 2.5: Order N feedback delay network.

- Smooth, rich echo density in the late part of the modeled reverberation.

A FIR filter is usually added to the FDNs to synthesize the early reflections, whose delays and amplitudes are estimated by other geometric methods, such as ISM. To obtain minimal coloration and rich echo density, it is suggested to use at least 8 delay lines [Rocchesso and Smith, 1997b]. Low-pass filters can be used in each delay line to model the smooth change of the absorption in different frequency bands of the wall materials.

Many interesting works have been done after the introduction of FDNs. The Digital Waveguide Mesh (DWM) and the FDNs are linked in the work of [Smith, 1985] and a circulant feedback matrix was proposed in [Rocchesso and Smith, 1997b]. [De Sena et al., 2011; Hacihabiboğlu et al., 2011] utilised this hybrid scheme to model the room reverberation using a scattering delay networks, whose parameters are estimated from the model of the DWM. [Menzer, 2010; Menzer and Faller, 2009] created a binaural reverberator using two parallel feedback delay networks.

There are also works which proposed novel feedback delay structures [Dahl and Jot, 2000] or feedback matrices [Menzer and Faller, 2010; Rocchesso, 1997]. [Frenette, 2000] proposed to use time-varying delay units to construct the FDNs which can reduce the coloration of the modeled reverberant sound and reduce the number of delay lines. Other approaches exist, including the recently proposed sparse velvet noise reverberator [Holm-Rasmussena et al., 2013] and the artificial reverberator in the frequency domain to model the spectral magnitude decay [Vickers et al.,

2006]. A comprehensive review of other artificial reverberators is given in [Välimäki et al., 2012].

In a recent study [Chemistruck et al., 2012], a genetic algorithm was used to automatically search for an optimal setting of the parameters for order-4 FDNs. [Primavera et al., 2014] automatically finds the optimal FDNs parameters where the parameter estimation performance was evaluated by both objective and subjective measures.

2.4 Hybrid Methods

The acoustic modeling methods, from wave-based, ray-based to artificial reverberators, all have their own strengths and weaknesses. The frequency and temporal ranges in which each of these methods are good at are plotted in the Figure 2.6. There is no strict boundaries between these categories. The idea is that in a certain frequency range or temporal range, one method is better than the other in some aspects.

In practice, hybrid methods are often used where two or three methods are used, one is used to model early reflections and the other is used to model the late reverberation. Or one method is used to model the low frequency range, and the other is for the high frequency range. It is not possible to give an exhaustive list of all these hybrid methods. We review a few commonly used hybrid methods.

Beam tracing/ISM and radiance transfer methods are often combined to model the complete impulse response. [Lewers, 1993] used a beam tracing method to model the specular reflections and radiance transfer method to model the diffuse reflections. However, there is no clear indication of the optimal point to switch from the specular to the diffuse model. More computationally efficient approaches use beam tracing model for the early specular reflections, and the radiance transfer methods for the late reverberation.

When using the FDNs to model the reverberation, apart from the feedback loops, it is common to use an additional FIR filter to model the early echoes, whose delays and amplitudes are often estimated by ISM [De Sena et al., 2011; Jot, 1997; Menzer, 2010; Murphy and Stewart, 2007; Savioja et al., 1999].

[Heinz, 1993] uses the image source method to model the specular early parts of the reflections. The missing diffusely scattered energies are handled by a method based on the result of a sound particle tracing procedure with a low time resolution.

[Southern et al., 2013] validated the use of a 3D FDTD scheme in conjunction with optimized ISM and acoustic radiance transfer method (RTM) models for RIR synthesis. In this implementation, the FDTD simulation is limited to low frequencies, while ISM and RTM models are applied

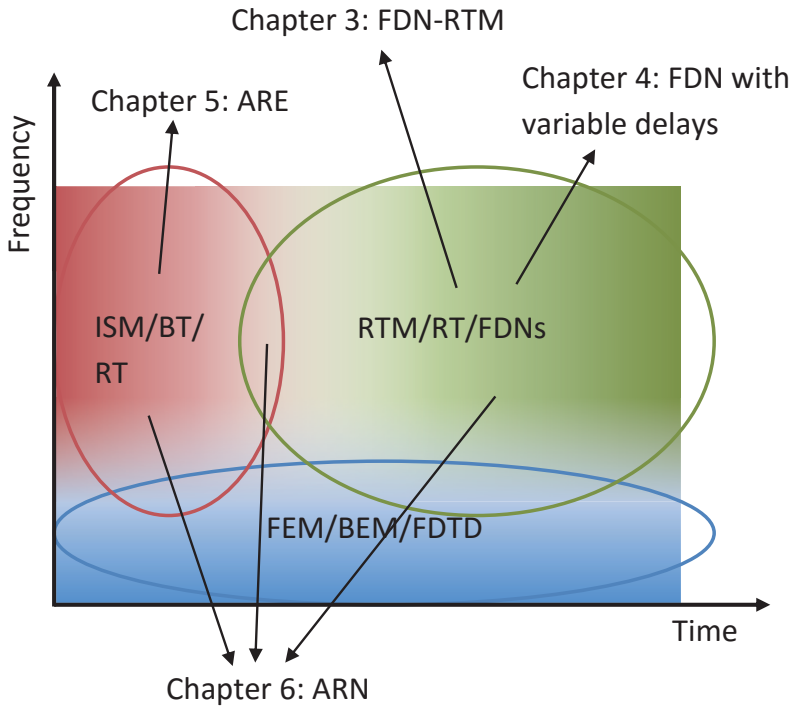


Figure 2.6: Different modeling methods at different frequency range and temporal ranges. The circles symbolise the open time-frequency domain of operation. The regions in which the contribution of each technical chapter lies are also marked in the figure. ISM = image source method, BT = beam tracing method, RT = ray tracing method, RTM = radiance transfer method, FDNs = feedback delay networks, FEM = finite element method, BEM = boundary element method, FDTD = finite-difference time-domain methods, ARN = Acoustic Rendering Networks.

to compute, respectively, the high frequency response for the early and late stages of sound propagation.

Chapter 3

Geometric-Based Late Reverberator: From Radiance Transfer to Feedback Delay Networks

Feedback delay networks, as a widely used structure for artificial reverberation synthesis, can efficiently model the late reverberation of the room impulse response. In this chapter, we study the relation between the radiance transfer method and the feedback delay networks. Such a relation can be used to estimate the parameters of the feedback delay networks, that will synthesize the late reverberation for the room considered.

This chapter is organised as follows. In section 3.1 we recall the knowledge of the acoustic radiance transfer method and its computation procedure for the modeling of the diffuse reflections. In section 3.2 we discuss our idea of linking the feedback delay networks with the acoustic transfer method. Then in Section 3.3 a geometrical-based statistical modeling paradigm is proposed, and the designing procedure of the corresponding parameter estimation is described. A special focus is put on the design of the feedback matrix which makes the feedback delay networks stable. In section 3.6 we use some typical room models to evaluate this method, from both numerical perspectives and perceptual perspectives.

3.1 Radiance Transfer Method

As seen in chapter 2, the Radiance Transfer Method (RTM), a ray-based geometric method, can efficiently model the diffuse reflections of RIRs and the sound energy decay of the late reverberation. The RTM decouples the sound source and receiver from the sound propagation by pre-computing a linear operator that defines how a sound emitted from surface patches affects the radiance of other surfaces. As an energy-based method to model the sound interaction between room surfaces, the phase information of the waves during propagation is assumed irrelevant. This section reviews the basics of the radiance transfer method. The sections 3.1.1-3.1.3 are briefly summarized from the work in [Hodgson and Nosal, 2006; Koutsouris et al., 2013b].

3.1.1 Analytical Formulation

The radiance transfer method is based on the assumption that all boundaries are diffusely reflecting and governed by Lambert's law:

$$I(\theta, R) = I(0, R)\cos\theta, \tag{3.1}$$

where $I(\theta, R)$ is the intensity of sound scattered by a surface element along the direction θ $(0 \leq \theta \leq \pi/2)$ from the surface normal $\mathbf{n}$ at distance R from the element.

The rate of energy that leaves a unit area of surface is defined as **radiation density** [Kuttruff, 2009]. It is a scalar quantity that is homogeneous to a sound intensity. In this chapter, the radiation density is seen as a position and time variant quantity, which is noted as $B(\mathbf{x}, t)$, where $\mathbf{x}$ is the position vector.

Analytically, the radiation density of an infinitesimal wall element, $\mathbf{x}$, is defined as its initial radiation density received directly from the source, plus the contribution of the radiation density from all other wall elements of the enclosure at some earlier time.

The radiation density of element at $\mathbf{x}$ contributed from the radiation density of another element at $\mathbf{x'}$, as shown in Figure 3.1, is:

$$B_{\mathbf{x'}}(\mathbf{x}, t) = (1 - \alpha(\mathbf{x}))B(\mathbf{x'}, t - \frac{|\mathbf{x} - \mathbf{x'}|}{c})e^{-mR}\frac{\cos\theta\cos\theta'}{R^2}, \tag{3.2}$$

where the term e^{-mR} accounts for the air absorption exponent over a distance R (assumed frequency-independent), $\alpha(\mathbf{x})$ is the absorption coefficients for the surface element at $\mathbf{x}$, R is the distance between $\mathbf{x}$ and $\mathbf{x'}$, θ and θ' are the angle between the line jointing the two elements and their surface normals, and c is the sound velocity in air. Integrating the contribution from

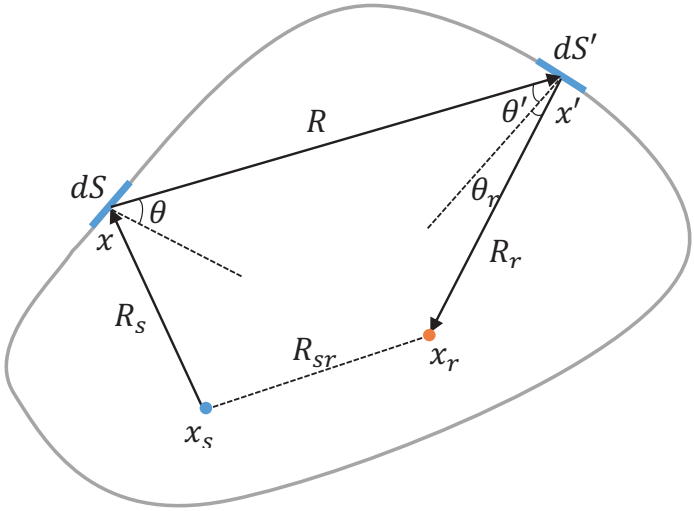


Figure 3.1: Integration of radiance density over the room surface.

the source and all the other elements in the enclosure S, the radiation density of element at $\mathbf{x}$ is:

$$B(\mathbf{x},t) = B_0(\mathbf{x},t) + \frac{1-\alpha(\mathbf{x})}{\pi}\int_S B(\mathbf{x'},t-\frac{|\mathbf{x}-\mathbf{x'}|}{c})e^{-mR}\frac{\cos\theta\cos\theta'}{R^2}\mathrm{d}x', \tag{3.3}$$

where the contribution from the source at $\mathbf{x}_s$ is calculated as:

$$B_0(\mathbf{x},t) = \frac{W(t-\frac{R_s}{c}\cos\theta_s)}{4\pi R_s^2}(1-\alpha(\mathbf{x}))e^{-mR_s}, \tag{3.4}$$

where $W(t)$ is the radiation power of the sound source, and R_s is the distance from the source to the surface element at $\mathbf{x}$. Once $B(\mathbf{x},t)$ is known for all elements on the surface S, the intensity at the receiver is:

$$I(\mathbf{x}_r,t) = I_0(\mathbf{x}_r,t) + \frac{1}{\pi}\int_S B(\mathbf{x},t-\frac{R_r}{c})e^{-mR_r}\frac{\cos\theta_r}{R_r^2}\mathrm{d}x, \tag{3.5}$$

where R_r is the distance from the element at $\mathbf{x}$ to the receiver, and the direct contribution from the source $I_0(\mathbf{x}_r,t)$ is:

$$I_0(\mathbf{x}_r,t) = \frac{W(t-\frac{R_{sr}}{c})}{4\pi R_{sr}^2}e^{-mR_{sr}}, \tag{3.6}$$

3.1.2 Numerical Formulation

To numerically simulate the radiation exchange, the room surface S is discretized into M finite-size small planar patches. The average radiation density of patch i is given as the discretized form of Equation 3.3:

$$B_i(t) = B_i^0(t) + (1-\alpha_i)\sum_{j=1,j\neq i}^{M} F_{i,j}B_j(t-\frac{R_{i,j}}{c}), \tag{3.7}$$

where $B_i^0(t)$ is the radiation density of patch i contributed directly from the sound source, i.e., the initial state, and $F_{i,j}$ is called the form factor, which defines the fraction of energy leaving patch j and incident to patch i and will be detailed in section 3.1.3.

The numerical simulation of the acoustic radiation can be regarded as an iterative procedure. In each iteration, after receiving the radiation density from other patches, patch i radiates its energy to other patches in the same diffuse manner. The radiation density of patch i after n iteration can be written in discretized form as:

$$B_i^{(n)}(t) = B_i^{(n-1)}(t) + (1-\alpha_i)\sum_{j=1,j\neq i}^{M} F_{i,j}B_j^{(n-1)}(t-\frac{R_{i,j}}{c}), \tag{3.8}$$

where $R_{i,j}$ is the average distance between patch i and j, and $F_{i,j}$ is a form factor which defines the fraction of energy leaving patch j and incident to patch i.

The total radiation density from patch i can be written as:

$$I_i(t) = \sum_{n=0}^{\infty} I_i^{(n)}(t), \tag{3.9}$$

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

3.1.3 Form Factor

The form factor, $F_{i,j}$, which defines the fraction of energy leaving patch j and incident to patch i, is given by:

$$F_{i,j} = \frac{1}{A_i}\int_{S_i}\int_{S_j}\frac{\cos\theta\cos\theta'}{\pi R^2}\mathrm{d}S'\mathrm{d}S, \tag{3.10}$$

where the integrals are taken over the patch areas S_i and S_j, and where A_i is the area of patch i, R is the distance between the points of integration on S_i and S_j, θ and θ' are the angles between

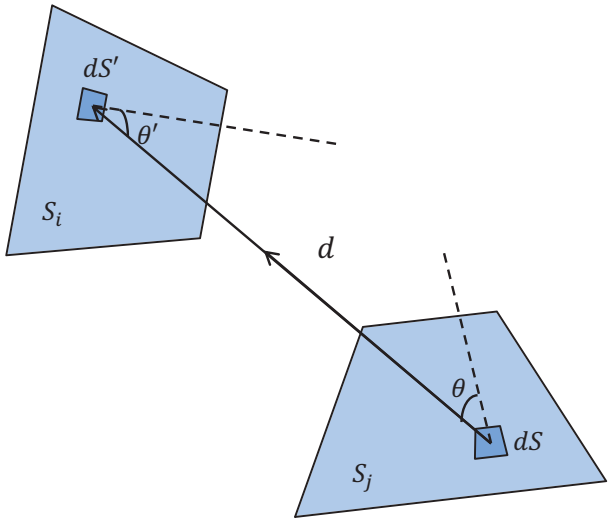


Figure 3.2: Geometric parameters for the form factor for diffuse reflections. $\mathrm{d}S$ and $\mathrm{d}S'$ are the areas of the patches in consideration, θ and θ' are the angles between the normal to the surface and the line joining the centres of the two patches, and d is the distance between the two patches.

the line joining the points of integration and the surface normals of the two patches, as shown in Figure 3.2.

It is easily proven that the form factors have the following properties:

$$A_i F_{i,j} = A_j F_{j,i}, \tag{3.11}$$

$$F_{i,i} = 0, \tag{3.12}$$

$$\sum_{j=1}^{M} F_{i,j} = 1. \tag{3.13}$$

Under the assumption that the dimensions of the patches are much smaller than the distances between them, Equation 3.10 can be approximated as:

$$F_{i,j} \approx \int_{S_j} \frac{\cos\theta \cos\theta'}{\pi R^2} \mathrm{d}S \approx A_j \frac{\cos\theta \cos\theta'}{\pi R^2}, \tag{3.14}$$

where θ and θ' are the angles between the line joining the central points of patch i and patch j,

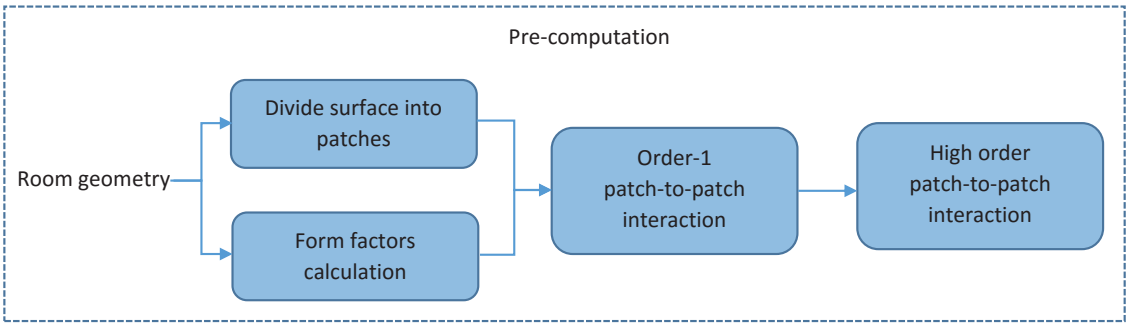


Figure 3.3: The pre-computation of the radiance transfer method.

and R is the distance between the central points of the two patches.

3.1.4 Simulation Procedure

The numerical simulation can be divided into an off-line pre-computation and a run-time computation. During the off-line computation, the geometry of the room is processed and the form factors $F_{i,j}$ are computed to form the order-1 patch-to-patch interaction. The room surfaces are divided into planar patches. The curved surfaces can be approximated using a number of small planar patches.

The high order patch-to-patch interactions can be pre-computed and stored by convolving the order-1 patch-to-patch responses. Frequency-domain operations can be used to calculate Equations 3.7-3.9 [Antani et al., 2012b; Siltanen et al., 2009]. The time domain response $B_i(t)$ can be represented in the frequency domain using its Discrete Fourier transform (DFT).

The run-time computation consists of the initial stage, the convolution stage and the final stage. At the initial stage, the sound energy from the source is distributed to each surface patch, forming the so-called initial shooting matrix. The initial shooting matrix is computed by emitting an impulse of unit energy from the source to all its visible patches. The impulse originating from the source can be modeled by a Dirac delta function, and the elements in the initial shooting matrix are also attenuated and time-shifted pulses. The initial stage can be performed in the frequency domain. The linearity of the Fourier transform assures that scalar multiplication and addition can be performed similarly to the time domain operation.

Then, the sound energy propagates within the enclosure and is exchanged between surface patches using the pre-computed high-order patch-to-patch reflections. It is done by convolving the elements of the initial shooting matrix with the pre-computed high-order patch-to-patch interaction responses, or by multiplication of their DFTs. After high-order reflections, the resulting

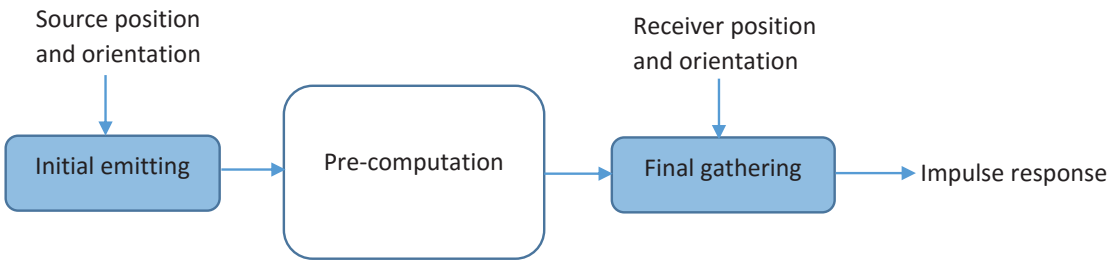


Figure 3.4: The computational procedure of the radiance transfer method.

responses are gathered from each surface patch to the listener, using the final gathering matrix. It is worth noting that the high-order reflection responses are 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.

In order to speed up the computation, [Siltanen et al., 2009] downsampled the pre-computed high-order patch-to-patch responses by a factor K ($K = 2, 4, 8, \cdots$). Such a downsampling scheme removes the high frequency range of the responses by assuming that the high frequencies of the responses are of little interest for the reverberation and that their frequency responses follow the same matter as in the low frequency ranges. In order to use less memory to store the high-order patch-to-patch responses in time domain, variable sampling rate can be used [Siltanen et al., 2010b]. Since the energy of the acoustic reverberation decays exponentially, it is reasonable to use a denser sampling rate for the early part of the reverberation and to use a coarser sampling rate for the late part, where the sampling rate decays exponentially as well. [Siltanen et al., 2010b] also discussed the possible extensions to model diffraction using the radiance transfer model.

Several studies have shown that the radiance transfer method is effective in predicting the reverberation characteristics of a given room. In [Nosal et al., 2004], predictions were made in three spheric enclosures of 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 [Carroll and Chien, 1977]. In [Hodgson and Nosal, 2006], 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.

3.2 From RTM to FDN

From this section, we begin to present our idea of how to link the radiance transfer method with the feedback delay networks. A new interpretation of the delay lines is first given from the physical point of view. Then the delay lengths are further discussed with the analysis of the geometries of the modeled rooms.

3.2.1 Relation between RTM and FDN

The RTM decouples the sound source and receiver from the sound propagation which makes it suitable for dynamic acoustic scenes synthesis where the sound source and the receiver may both move. 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.

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. The statistical artificial reverberator can be generalized into a structure called Feedback Delay Networks (FDNs) [Jot and Chaigne, 1991; Stautner and Puckette, 1982]. This structure is characterized by a set of delay lines connected in a feedback loop through a feedback matrix. 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 [Rocchesso and Smith, 1997a], 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.

It can be noticed that, in the computational procedure of the radiance transfer method in the time domain, the iteration of energy exchanges between surface patches can also be regarded as feedback loops in a feedback delay network, as illustrated on Figure 3.5. Here, each delay line represents the 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 is then 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 combined room geometry-based statistical method then uses delay lines to represent a group of diffuse reflections between room surfaces, with the acoustic parameters averaged by the diffuse reflections included in the group.

For the FDNs, the convolution of the room impulse response with the dry signal is replaced by

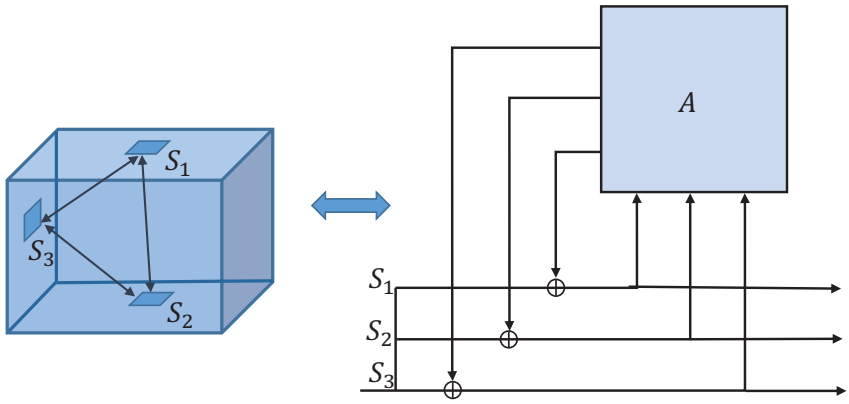


Figure 3.5: Relation between the Radiance Transfer Method and feedback delay networks. Left: In the Radiance Transfer Method, the sound bounces back and forth between surface patches; right: In a feedback delay network, each patch-to-patch interaction is abstracted as a feedback loop.

$N^2 + kN$ multiplications and a few additions per input sample, where N is the number of delay lines and k is a constant which is related to the additional attenuation operations in each delay line. Thus, in another way, the combined structure can also be regarded as an approximated but computationally efficient structure of the radiance transfer method.

3.2.2 Grouping the patch-to-patch interactions

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 huge FDNs 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.

3.2.3 Delay Lengths of the FDN

The range and distribution of the delay lengths can also guide the choice of delay units in the FDNs structure. To illustrate this, we have computed the distribution of occurrence and energy

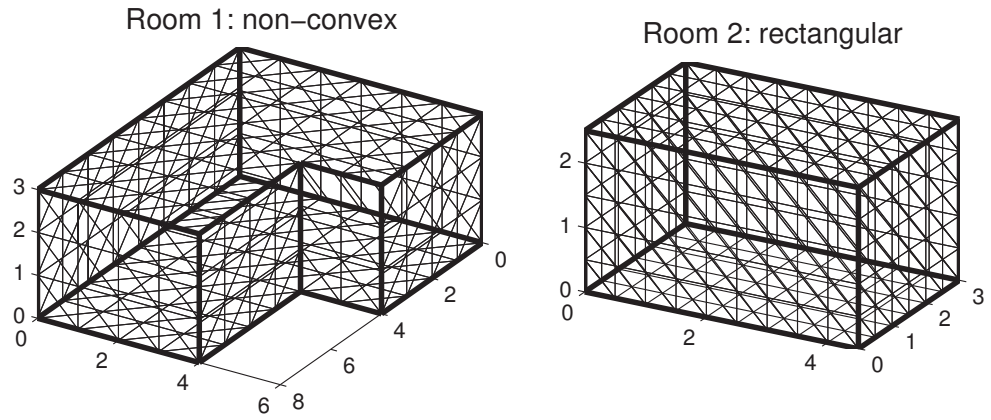


Figure 3.6: The experimental geometries. Room 1: non-convex room of outside dimension 8m×6m×3m. Room 2: rectangular room of dimension 4.5m×3m×2.5m.

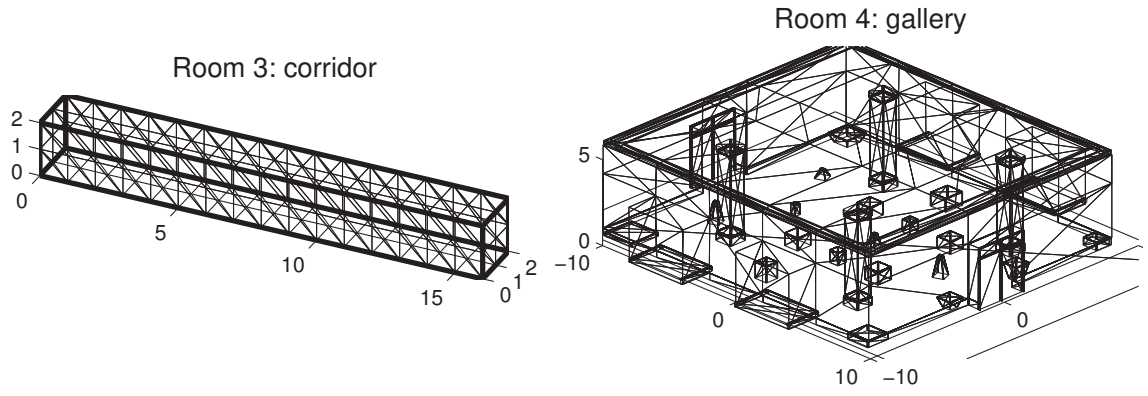


Figure 3.7: The experimental geometries. Room 3: corridor of dimension 16m×2m×2m. Room 4: gallery of outside dimension 19.6m×17.7m×5.1m.

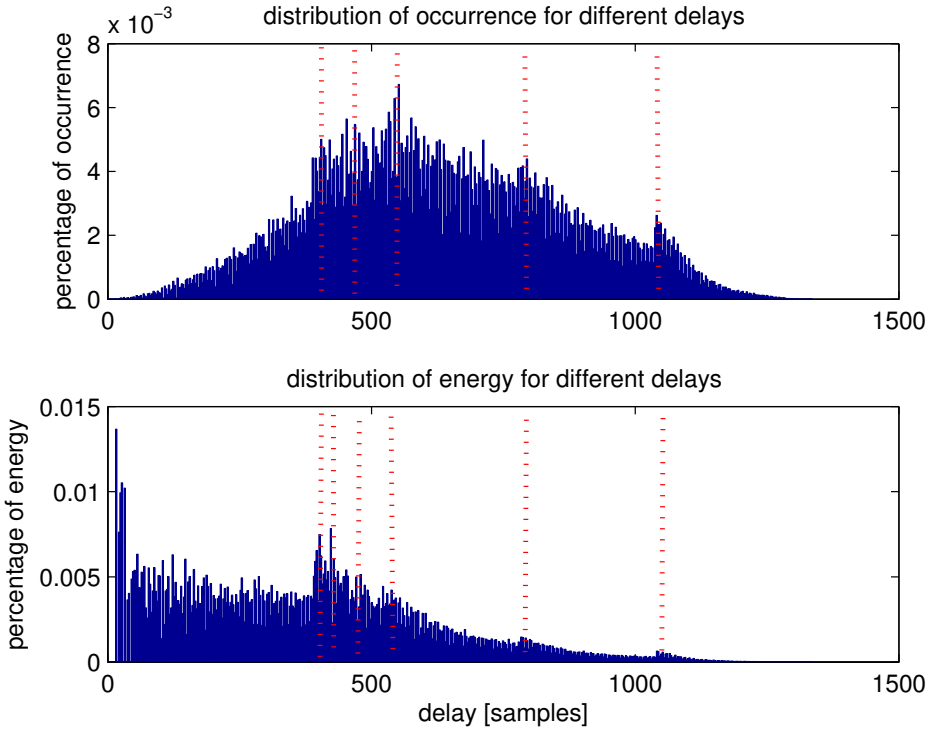


Figure 3.8: Distribution of occurrence and energy for different patch-to-patch delays due to diffuse reflections for Room 1. Top: histogram of occurrence. Bottom: histogram of energy. The dotted lines illustrate some peaks of the distributions. The sampling rate is 44.1 kHz.

for different delays of the non-convex room displayed in Figure 3.6 - Room 1. This room is modeled using 5248 triangular patches, with the edge of 25 cm. (It should be noted that the triangular patches in Figures 3.6 and 3.7 do not illustrate the real patching density during the computation of the RTM algorithm.) The distribution of occurrence and energy is shown in Figure 3.8. 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 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 Equation 3.10. A closer look at Figure 3.8 reveals that both the occurrence distribution and the energy distribution have small peaks at around 390, 540, 800 and 1050 samples, which correspond to distances of 3, 4, 6 and 8 m, 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 in Figure 3.7 - Room 3. The delay distribution, as shown in Figure 3.9, has a strong peak at around 250 samples, which corresponds to 2 m, the width of the corridor. In such cases, the delays can be chosen around the strong peak values to model the acoustic modes.

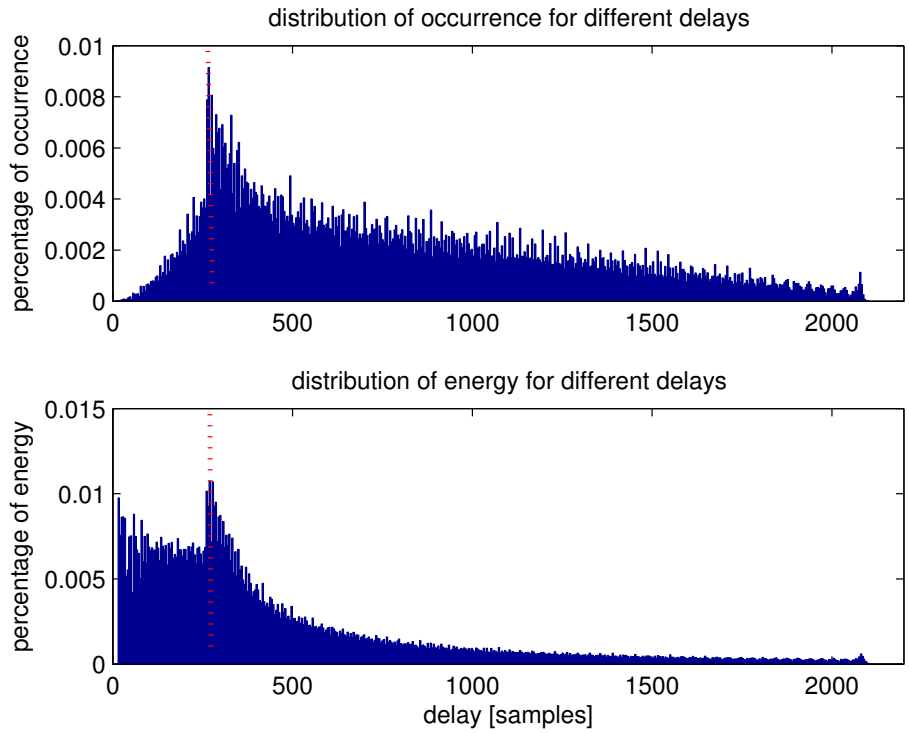


Figure 3.9: Distribution of occurrence and energy for different patch-to-patch delays due to diffuse reflections for Room 3. Top: histogram of occurrence. Bottom: histogram of energy. The dotted lines highlight the main peak of the distributions. The sampling rate is 44.1 kHz.

3.3 RTM Model-Based Late Reverberator

In this section, we discuss the link between the RTM model and the FDNs structure and present a new geometrical-based statistical model, named FDN-RTM. The name of the model comes from two parts. On the one hand, it borrows the FDNs structure for its efficient implementation and flexible parameter settings. On the other hand, the order-1 patch-to-patch responses from the radiance transfer method are used to estimate the parameters of the FDNs structure, in order to

synthesize a realistic impulse response as expected from the specified room geometry.

3.3.1 System structure

The proposed reverberator is illustrated in Figure 3.10. 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 [Jot and Chaigne, 1991], except that 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 an additional FIR filter (see Figure 3.10).

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 in Figure 3.10.

3.3.2 Parameter estimation

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

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:

$$\ell_m = \sum_{i \to j \in \wp_m} F_{i,j}, \tag{3.15}$$

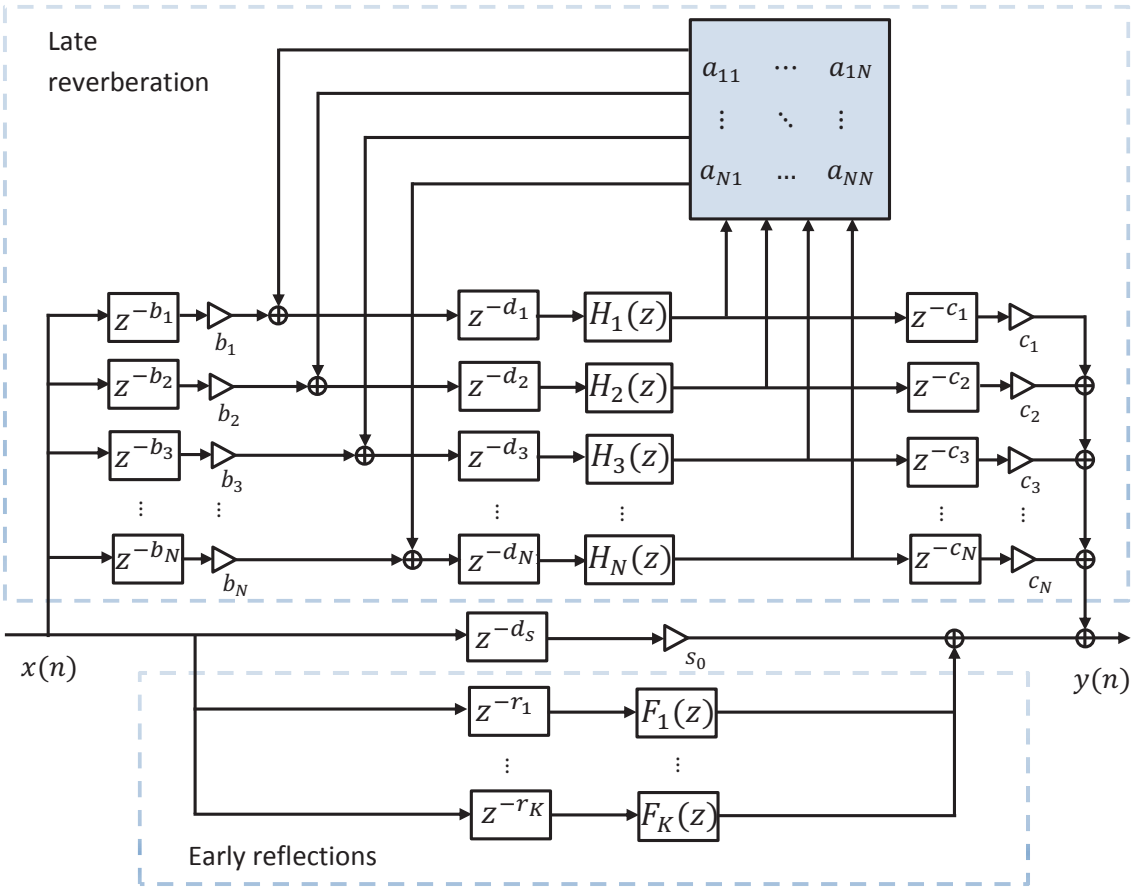


Figure 3.10: Structure of the FDN-RTM method. The input dry signal $x(n)$ in the left undergoes feedback delay loops to generate the late reverberation (upper part) and a series of attenuated and delayed taps to synthesize the early reflections (lower part). The estimation of the parameters in the structure is detailed in Section 3.3.2.

$$\ell_{m,n} = \sum_{i \to j \in \wp_m} \sum_{j \to k \in \wp_n} F_{i,j} F_{j,k}, \tag{3.16}$$

where $\wp_m$ and $\wp_n$ are the sets of patch-to-patch sound interactions in group m and n respectively.

The feedback coefficient a_{mn} is defined as:

$$a_{mn} = \frac{\ell_{m,n}}{\ell_m}, \tag{3.17}$$

$$subject\ to \sum_{n=1}^{N} a_{mn} = 1. \tag{3.18}$$

According to the energy conservation principle, (3.18) will be satisfied if the form factors are computed with ideal precision. Using the form-factor calculation in [Nosal et al., 2004], 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, FDNs 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 contain 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.

$$A = \begin{pmatrix} \sqrt{a_{11}} & \cdots & \sqrt{a_{1N}} \\ \vdots & \ddots & \vdots \\ \sqrt{a_{N1}} & \cdots & \sqrt{a_{NN}} \end{pmatrix}. \tag{3.19}$$

The signs of these entries can then be chosen according to the procedure described in section 3.4.3.

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 [Huopaniemi et al., 1997]. In the case of FDN-RTM, the attenuations are energy-weighted mean reflection coefficients of all the emitting patches within the group:

$$\bar{\theta}_n^{\,s} = \sum_{i \to j \in \wp_n} F_{i,j} \theta_i^{\,s}, \tag{3.20}$$

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 Equation 3.20, 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.

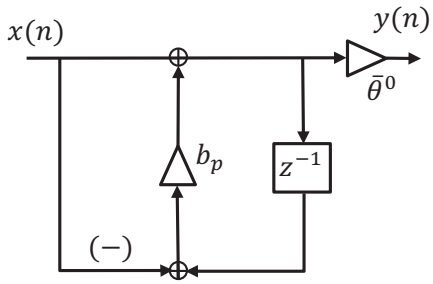


Figure 3.11: Implementation of first order low-pass filter.

The frequency-dependent absorption of the surface material can be modeled by a first order low-pass filter, as shown in Figure 3.11. Its frequency response is:

$$|H(e^{j\omega})|^2 = \frac{(1-b_p)^2}{1+b_p^2-2b_p\cos(\omega)}, \tag{3.21}$$

which has a ratio $(1-b_p)/(1+b_p)$ between zero and Nyquist frequency. Thus, the factor b_p can be estimated using the reflection coefficients of materials at zero and Nyquist frequency as:

$$\frac{\bar{\theta}^{fs}}{\bar{\theta}^0} = \frac{1-b_p}{1+b_p}. \tag{3.22}$$

However, acoustic absorption properties of materials are often only available from 125 Hz until 4 kHz or 8 kHz. When desired, the target response at Nyquist frequency can be approximated from the 4 kHz or 8 kHz octave band absorption coefficients.

Higher order filters (FIR or IIR) can be designed using digital filtering techniques based on AR (all-pole), MA (FIR), or ARMA (pole-zero) modeling. The selection of a filter depends on the target criteria of the design and the available data. The Gauss-Newton method can be used to fit the absorption coefficients of the octave bands and to optimise the filter coefficients.

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:

$$\bar{d} = 4V/S, \tag{3.23}$$

where V is the total volume of the room, and S is total surface area enclosing the room. In [Rafii and Pardo, 2009] 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:

$$\bar{d}_n = \frac{\sum_{i \to j \in \wp_n} F_{i,j} d_{i,j}}{\sum_{i \to j \in \wp_n} F_{i,j}}. \tag{3.24}$$

The delay lengths calculated using Equation 3.24 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 Figure 3.8, 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 needs to be modified accordingly when its length is adjusted:

$$\theta' = \bar{\theta}^{d'/\bar{d}}, \tag{3.25}$$

where $\bar{d}$ is the mean delay length for a certain delay line calculated by Equation 3.24, $\bar{\theta}$ is the mean attenuation by order-1 reflection, d' is the adjusted delay length and θ' is the attenuation factor that needs to be modified accordingly.

4. Early reflections

Early reflections need to be accurately modelled since they play a key role in 3D sound perception in a room. To that aim, realistic geometric-based model should be used. For simple geometries, the delays $z^{-r_1}, ..., z^{-r_K}$ and attenuations filters of early reflections $F_1(z), ..., F_K(z)$ can be estimated, up to the third order for interactive update rate, using traditional ISM [Allen and Berkley, 1979]. These parameters can also be estimated by beam tracing [Funkhouser et al., 1998] or other recent methods [Antonacci et al., 2008; Laine et al., 2009] since they achieve interactive performance for both moving sources and receivers. Since the modeling of early reflections is not the focus of this chapter, interested authors are encouraged to refer to [Lauterbach et al., 2007; Lehmann and Johansson, 2008]. Moreover, an extension of the acoustic rendering equation to model early echoes is proposed in Chapter 5.

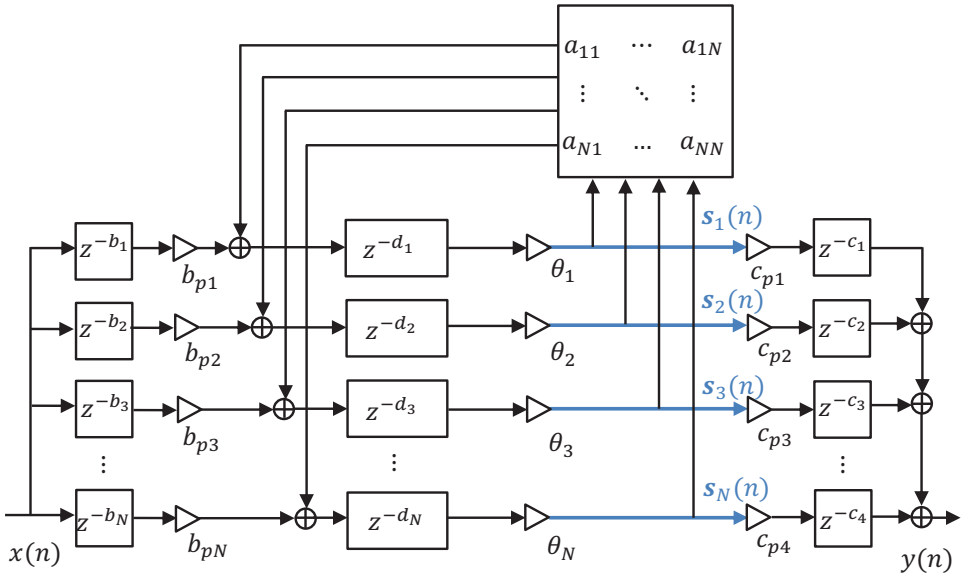


Figure 3.12: Feedback delay networks.

5. Pre-mixing and post-mixing coefficients

We name $z^{-b_1}, ..., z^{-b_N}$ the pre-mixing delays, and $b_1, ..., b_N$ the gains in Figure 3.10, 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 $z^{-b_1}, ..., z^{-b_N}$ are chosen as the average integer number of such initial delays. The coefficients $b_1, ..., b_N$ are the amounts of energy emitted from the source which go to each group after the initial shooting.

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

3.4 Stability of the Reverberator

Although the parameters estimated above satisfy the energy-conservation principle, the feedback loops in the feedback delay networks, as depicted on Figure 3.12, impose further constraints to keep the system stable.

3.4.1 System transfer functions

The relations between the ports of the FDNs in Figure 3.12 can be formulated by:

$$y(n+c_i) = \sum_{i=1}^{N} c_{pi} s_i(n), \tag{3.26}$$

$$s_i(n+d_i) = \sum_{j=1}^{N} a_{i,j}\theta_i s_j(n) + b_{pi}\theta_i x(n-b_i), \tag{3.27}$$

where $s_i(n)$ are the output of the feedback loops of delay line i at time sample n, and θ_i is the attenuation coefficient for delay line i.

Using the z-transform in matrix form, we rewrite the system as:

$$Y(z) = C(z)C_p^T S(z), \tag{3.28}$$

$$S(z) = D(z)\Theta\left[AS(z) + B(z)B_p^T X(z)\right], \tag{3.29}$$

The diagonal matrix $D(z)$ is the delay matrix:

$$D(z) = \begin{pmatrix} z^{-d_1} & & & 0 \\ & z^{-d_2} & & \\ & & \ddots & \\ 0 & & & z^{-d_N} \end{pmatrix}, \tag{3.30}$$

and Θ is the attenuation coefficient matrix at a certain frequency band:

$$\Theta = \begin{pmatrix} \theta_1 & & & 0 \\ & \theta_2 & & \\ & & \ddots & \\ 0 & & & \theta_N \end{pmatrix}. \tag{3.31}$$

$S(z) = [s_i(z), \cdots, s_N(z)]^T$ is the vector denoting the output of the feedback loops of each delay line, $B_p = [b_{p1}, \cdots, b_{pN}]^T$ is the vector of the initial-shooting gains of each delay line, $C_p = [c_{p1}, \cdots, c_{pN}]^T$ is the vector of the final-gathering gain of each delay line, and $\theta = [\theta_1, \cdots, \theta_N]^T$ is the vector of the attenuation coefficients of the patches corresponding to the patch-to-patch interactions. $B(z) = [b_1(z), \cdots, b_N(z)]^T$ is the initial delay vector, and $C(z) = [c_1(z), \cdots, c_N(z)]^T$ is the final delay vector. The matrix $A = [a_{i,j}]_{N\times N}$ is the feedback matrix.

Combining Equation 3.26 and 3.27, we obtain the system transfer function as:

$$\begin{aligned} H(z) &= \frac{Y(z)}{X(z)} \\ &= C(z)C_p^T \left[I - D(z)\Theta A\right]^{-1} D(z)B(z)\Theta B_p^T. \end{aligned} \tag{3.32}$$

Let $G = \Theta A$ where A is the feedback matrix and Θ is the attenuation coefficient matrix, the matrix series expasion of $\left[I - D(z)\Theta A\right]^{-1}$ can be written as:

$$\left[I - D(z)\Theta A\right]^{-1} = \left[I - D(z)G\right]^{-1} \tag{3.33}$$

$$= I + D(z)G + D^2(z)G^2 + \cdots + D^n(z)G^n + \cdots \tag{3.34}$$

Since $\|D(z)\| \leq I$ and $\|G\| < I$, by using triangle inequality and the Cauchy-Schwarz inequality, we have:

$$\begin{aligned} \| \left[I - D(z)\Theta A\right]^{-1} \| &= \|I + D(z)G + D^2(z)G^2 + \cdots + D^n(z)G^n + \cdots \| \\ &\leq \|I\| + \cdots + \|D^n(z)G^n\| + \cdots \\ &\leq 1 + \cdots + \|D(z)\|^n \|G\|^n + \cdots \\ &\leq 1 + \cdots + \|G\|^n + \cdots \\ &= \frac{1}{1 - \|G\|}, \end{aligned} \tag{3.35}$$

where $\|G\|^n$ can be regarded as the attenuation of the reflection of order n. Therefore, as long as $\|G\|^n$ decays with n, the stability of the system transfer function is guaranteed.

If we decompose G using an eigenvalue decomposition:

$$G = P\Lambda P', \tag{3.36}$$

where the diagonal elements of Λ are the eigenvalues, and P is the eigenmatrix, the high order reflections in Equation 3.34 are expressed as:

$$\begin{aligned} \|G^n\| &= \|(P\Lambda P')^n\| \\ &= \|\underline{P}\ \Lambda\ \underline{P'P}\ \Lambda\ \underline{P'P}\ \Lambda\ \underline{P'}\ \ldots\ \| \\ &\leq \|\Lambda\|^n \end{aligned} \tag{3.37}$$

.

Since the product of the eigenmatrix P and its inverse (underlined in Equation 3.37) does 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 or equal to 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.

3.4.2 Choice of the feedback matrix

Normalized 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 [Lee et al., 2009]. If H_N is the $N \times N$ Sylvester Hadamard matrix, then it is well known that Sylvester Hadamard matrices can be recursively constructed using Kronecker product as:

$$H_{2N} = H_2 \otimes H_N = \begin{pmatrix} +1 & +1 \\ +1 & -1 \end{pmatrix} \otimes H_N = \begin{pmatrix} H_N & H_N \\ H_N & -H_N \end{pmatrix}. \tag{3.38}$$

The Hadamard matrices of first and second order can be:

$$H_1 = \begin{pmatrix} 1 \end{pmatrix}, \tag{3.39}$$

$$H_2 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \tag{3.40}$$

$U_N = \frac{1}{\sqrt{N}} H_N$ is the normalized Sylvester Hadamard matrix so that the eigenvalues of U_N are

±1's.

[Jot and Chaigne, 1991] uses the Householder feedback matrix as the feedback matrix. The Householder matrix is given as:

$$A_N = I - \frac{2}{N}{\mu_N}^T \mu_N, \tag{3.41}$$

where N is the order of the matrix, I is identity matrix of dimension N and $\mu_N = [1\ 1\ ...\ 1]$.

When $N = 4$, the Householder matrix is

$$A_4 = \frac{1}{2}\begin{pmatrix} -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{pmatrix}, \tag{3.42}$$

which is also a Hadamard matrix. The higher order Hadamard matrix can be construct from A_4 using Equation 3.40. Hadamard matrix constructed from A_4 has uniform energy for each entry, which leads to the possibility to design a feedback matrix with uniform entries.

3.4.3 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.

Since the form factors contain 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

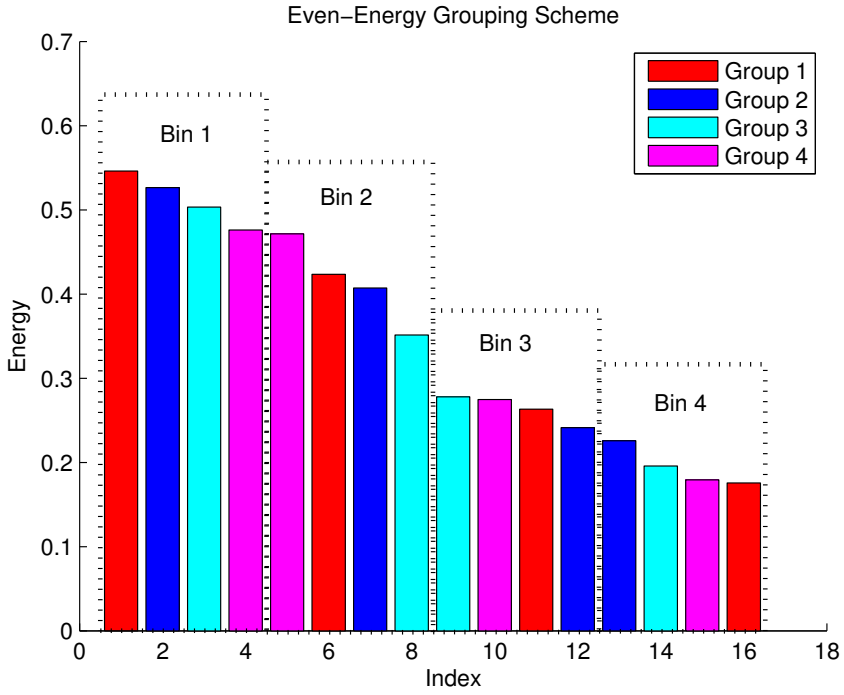


Figure 3.13: Even-energy group assignment scheme for FDN-RTM with four delay lines. The form factors with the same color denote that they are assigned to the same group.

that the sorted energy is nearly linearly descending with the same decreasing step. We take the first N form factors as Bin 1 and assign them to different groups, where N is the order of the desired feedback matrix. Then for the $N+1$ to $2N$ form factors, Bin 2, we do the same, assigning each of them to a different group, but in a circular order.

Take $N = 4$ for example. The 4×4 patch-to-patch energy elements sorted in descending order are divided into 4 bins, as shown in Figure 3.13. Since all adjacent samples decrease with the same step, to guarantee that each group be assigned with equal amount of energy, we only need to design a grouping scheme where each group take one largest energy, one second largest, one third largest, etc., and one smallest energy from each bin. Group 1 takes the first, second, third and fourth form factors from the four bins respectively. Group 2, in a circular order, takes the second, third, fourth and first form factors from the four bins, and so on. The same procedure continues for the next 16 form factors. The groups are denoted by different colors in Figure 3.13. The order assignment is summarized in Table 3.1. So eventually Group 1 takes the patch-to-patch energy elements with index $1, 6, 11, 16$, Group 2 takes the patch-to-patch energy elements with index $2, 7, 12, 13$, Group 3 takes the patch-to-patch energy elements with index $3, 8, 9, 14$ and Group 4 takes the patch-to-patch energy elements with index $4, 5, 10, 15$.

Table 3.1: *Group assignment for four delay lines.*

Group index	Element assignment index				Orders in each bin			
Group 1	1	6	11	16	1st	2nd	3rd	4th
Group 2	2	7	12	13	2nd	3rd	4th	1st
Group 3	3	8	9	14	3rd	4th	1st	2nd
Group 4	4	5	10	15	4th	1st	2nd	3rd

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.

3.5 FDN-RTM Algorithm

The different steps of FDN-RTM algorithm are summarized in Figure 3.14:

1. Decompose the room geometry into patches.

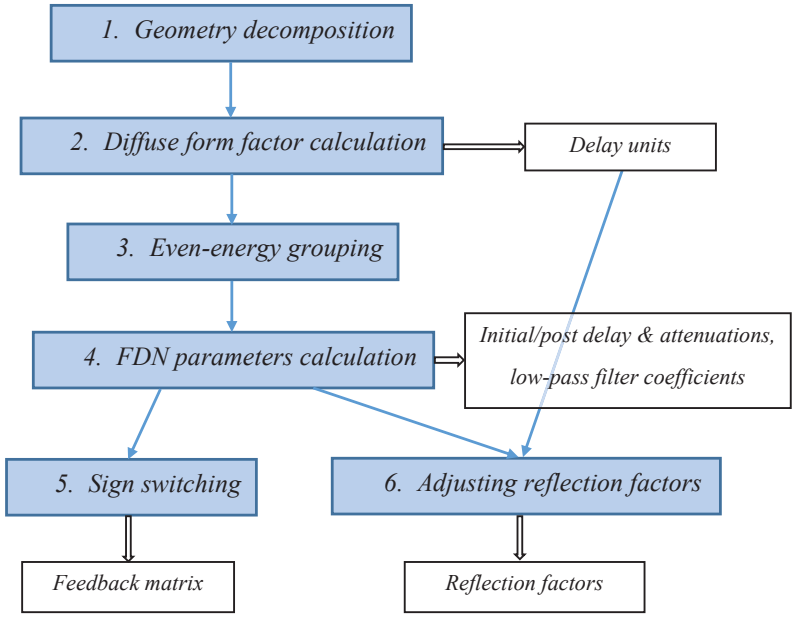


Figure 3.14: Process of finding the parameters of the FDN-RTM algorithm.

2. Calculate the diffuse form factors for all patch-to-patch reflections using Equation 5.15. 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 Figure 3.13.

4. Estimate the FDN-RTM parameters as detailed in Section 3.3.2: the feedback matrix using Equations 3.15-3.18, the attenuation factors for each delay line using Equation 3.20, the low-pass filter coefficients using Equation 3.22, and the average delay length using Equation 3.24.

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

6. Using Equation 3.25, 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).

Table 3.2: Wall reflection coefficients settings of the experimental geometries.

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

3.6 Experimental Results

3.6.1 Experimental geometries

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 in Figure 3.6 and Figure 3.7.

Room 1 is a non-convex room with dimension 8m×6m×3m and uniform reflection coefficients are 0.9. Room 2 is a rectangular room with dimension 4.5m×3m×2.5m. Room 3 is a long corridor with dimension 16m×2m×2m. Room 4 is a more complex gallery geometry. The main shape of the gallery is a rectangular room with dimension 19.6m×17.7m×5.1m, with four pillars and a few small box-shaped or cone-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 3.2.

3.6.2 Simulation examples

We use Room 1, the non-convex room, as an example to show the parameter estimation process for a FDN-RTM of order 8 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 1 m^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, 0.997, 1.001, 0.999, 0.999, 1.000, 1.004, 1.003).

Figure 3.15 shows the synthesized reverberation (normalized) for room 1 using 8 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.

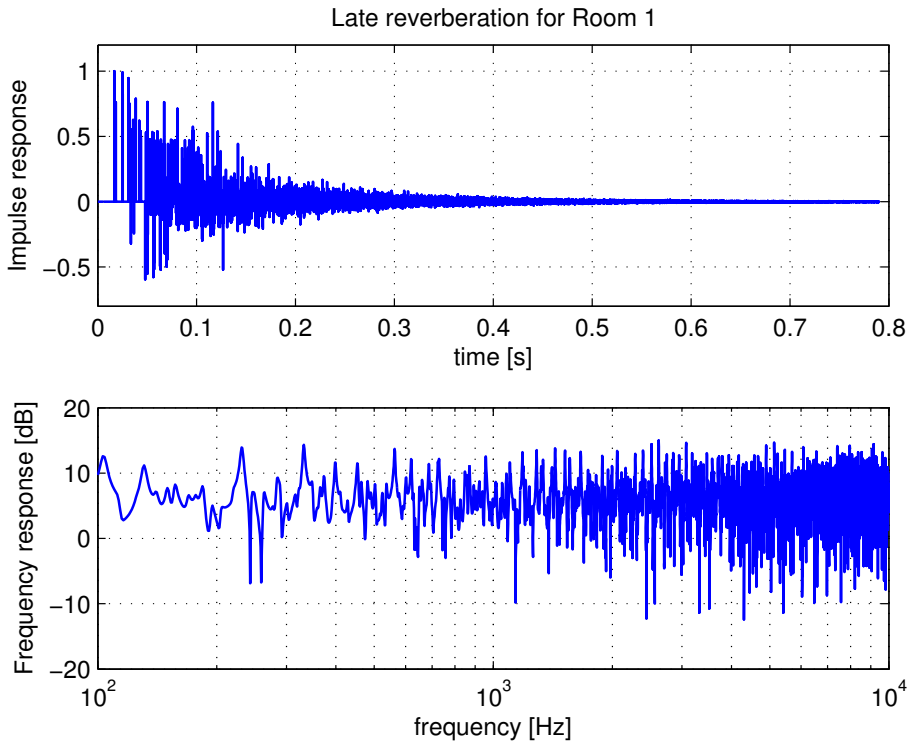


Figure 3.15: Room impulse response synthesized for Room 1 with FDN-RTM using 8 delay lines.

3.6.3 Evaluation of the decaying trend

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 use the energy decay curve to compare these two methods.

Figure 3.16 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 are exponentially decreasing (e.g. 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 3.3 compares the reverberation time T_{60} estimated by the RTM (reference), FDN-RTM and Sabine formula for the four rooms. 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

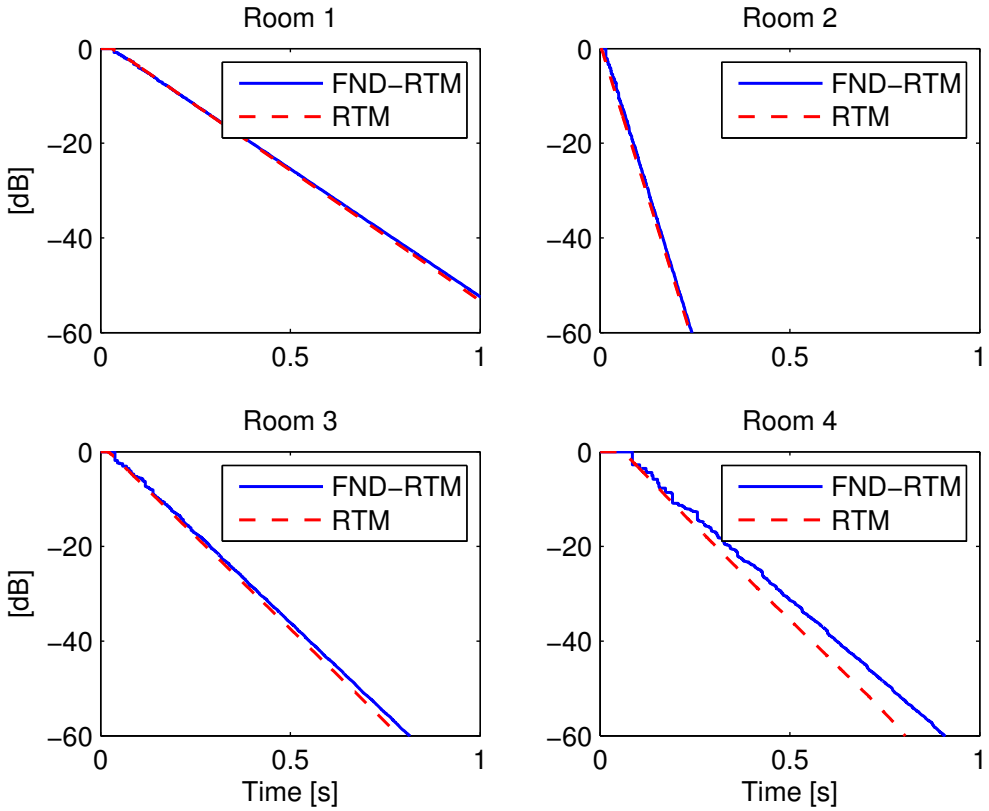


Figure 3.16: Energy decay curves of the wideband impulse responses synthesized using FDN-RTM and RTM. The x-axis is kept within [0,1] s and the y-axis is within [0, -60] dB.

Table 3.3: 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

of the main shoe-box, four pillars and the windows. Thus the Sabine prediction is not guaranteed to be accurate for Room 4.

The reverberation times estimated by FDN-RTM agree well with those estimated by RTM and Sabine formula. However, the reverberation times estimated by FDN-RTM are slightly longer.

Table 3.4: Computation time consumed for RTM to compute a one-second room impulse response.

Room	N. patches	RTM	
		Pre-comp.	Render
1	164	1 h 18 min	30.4 s
2	64	13 min	7.1 s
3	68	14 min	8.0 s
4	435	14 h 51 min	85.8 s

Table 3.5: Computation time consumed for FDN-RTM with 4, 8 and 16 delay lines to compute a one-second room impulse response.

Room	N. patches	FDN-RTM N=4		**FDN-RTM N=8**		FDN-RTM N=16	
		Pre-comp.	Render	**Pre-comp.**	**Render**	Pre-comp.	Render
1	164	1.93 s	0.078 s	**2.72 s**	**0.187 s**	4.64 s	0.476 s
2	64	0.42 s	0.091 s	**0.49 s**	**0.178 s**	0.67 s	0.451 s
3	68	0.46 s	0.097 s	**0.58 s**	**0.192 s**	0.66 s	0.504 s
4	435	13.34 s	0.090 s	**24.08 s**	**0.190 s**	30.48 s	0.493 s

The main reason is that the feedback matrix estimated using even-energy grouping scheme has some eigenvalues slightly higher than the others, which makes the corresponding delay lines decay slower. Note that the synthesized reverberation of Room 4 has a higher error compared to RTM, mainly because the estimated feedback matrix is not perfectly unitary (its largest eigenvalues being 1.076).

Although the experimental geometries have high reflection coefficients, the FDN-RTM networks also work for more absorptive materials in the same way as for the low absorptive materials. The feedback matrix estimated using the even-energy grouping scheme in Section 3.4.3 is nearly unitary, and thus is lossless in the feedback delay networks structure. The energy decaying trend of the reverberation is affected by the estimated attenuation filter of each delay line (estimated by the absorption characteristics of the walls).

3.6.4 Computation performance

While achieving comparable simulation results for late reverberation, FDN-RTM requires a much smaller computation time than the RTM during the pre-computation stage and the real-time rendering. Table 3.5 compares the computation times for both methods on a standard PC. The computer used for this simulation has a 2.8 GHz CPU and 8 GB RAM.

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 $N^2 + kN$ multiplications and some additions/delays during the feedback loops, as shown in Figure 3.10, where N is number of delay lines and k is a constant.

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 increases, at the price of a higher computation load. The computation time for FDN-RTM using 4 to 16 delay lines is shown in Table 3.5. The pre-computation time for the FDN-RTM depends on the complexity of the geometry, i.e. the number of patches. After the parameters of the FDN-RTM are estimated during the pre-computation stage, the rendering time is independent of the complexity of the geometry, and increases almost linearly with the number of delay lines.

To render a one-second signal, FDN-RTM takes about 100 to 500 ms during run-time on a standard computer, depending on the number of delay lines. This makes this geometry-based reverberator suitable for real-time processing. It should be emphasized that this is typically 2 to 3 orders of magnitude less than needed by the RTM.

3.6.5 Subjective evaluation using listening tests

In this section, we evaluate the perceptual quality of reverberated sounds as synthesized by the FDN-RTM model, for both speech and music signals. Informal listening sessions indicate that, using only 4 delay lines, the reverberant sounds feature a slightly metallic quality. When the number of delay lines increases, this metallic attribute is reduced, with more natural-sounding reverberant signals.

To evaluate more formally the sound quality of the late reverberation, subjective listening tests (Figure 3.17) were conducted according to the ITU-RBS.1534 Multiple Stimulus with Hidden Reference and Anchor (MUSHRA) recommendation [ITU-1534, 2001]. The tests were performed using professional-grade headphones (Beyerdynamic DT770 Pro). 10 listeners (excluding the authors) were involved, 9 males and 1 female, aged between 20 and 32, all of them with some previous experience of listening tests. The input dry signals are eight excerpts of anechoic recordings of speech, piano, cello and trumpet sounds selected from [Bang&Olufsen, 1992; Open-AIR], each with a duration of about 7-10 s. The instrumental signals are mostly legato. Room 1 and Room 4 were used as simulation environments. The reflection coefficients of Room 4 were mod-

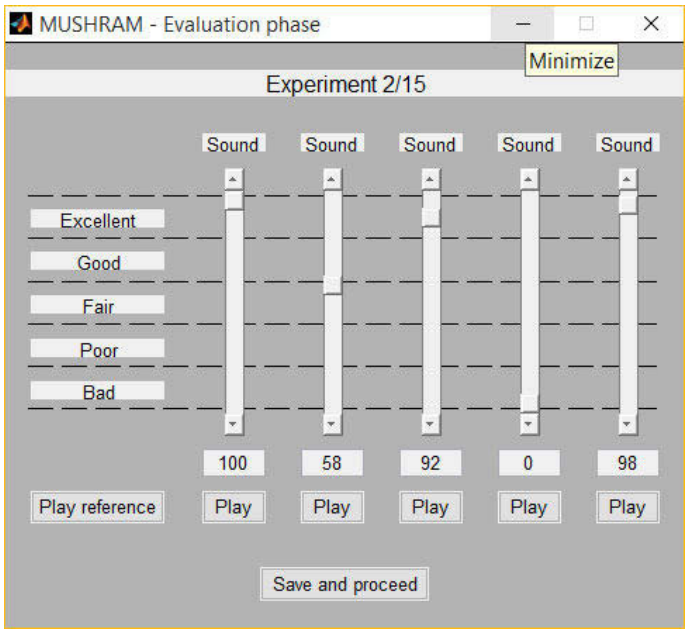


Figure 3.17: Subjective evaluation interface.

ified to have T_{60}=2.34 s and the T_{60} of Room 1 is kept at 1.10 s. Sixteen test sessions were carried out by synthesizing the reverberant sounds of the eight input signals in the two testing rooms. In each session, the reverberation generated by the RTM was used as reference. As suggested in [ITU-1534, 2001], the anchor signal was obtained by low-pass filtering the reference signal with a cut-off frequency of 3.5 kHz. The signals under evaluation were generated by the FDN-RTM using 4, 8 and 16 delay lines, with parameters estimated by the room geometries. Some sound examples used in the tests can be found on the companion website.[1]

Before the main test, listeners had to perform a training phase: the listeners were exposed to the nature of the reverberant signals used in the test in order to gain experience in listening. During the main test, the listeners were asked to rate the sound quality of the various reverberant samples. The experiment takes 20-25 mins for each listener.

Figure 3.18 shows the statistics of the sound quality rated by the listeners. The boxplots shown in the diagrams contain the median, upper, lower quartiles and outliers. The results of piano, cello and trumpet are averaged on a single diagram for the sake of conciseness, and also for the reason that they show similar results. The results clearly show that the sound quality increases with the number of delay lines for both voice and musical instruments. Although FDN-RTM with 4 delay lines can produce excellent sound quality in some cases, in general, its produced sound quality is irregular and thus is not recommended for practical use. The sound quality has noticeable

[1] http://perso.telecom-paristech.fr/~hbai/demo_ASLP2015/

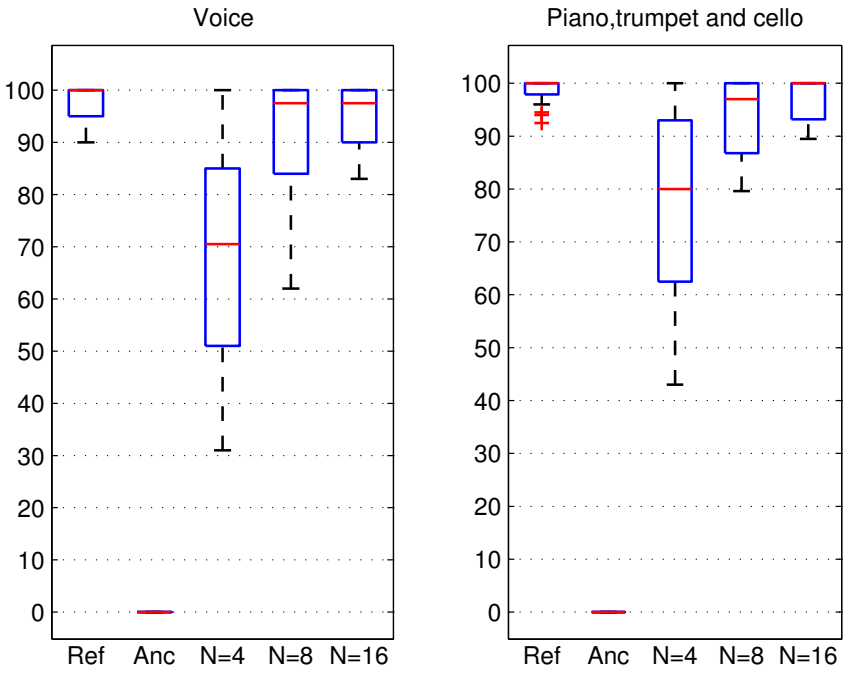


Figure 3.18: Listening test results for reverberation using FDN-RTM with 4, 8 and 16 delay lines. Ref denotes the reference signals which are computed with RTM, and Anc denotes the anchor signals which are the low-pass filtered reference signals, with a cut-off frequency at 3.5 kHz.

Table 3.6: Material absorption coefficients from 125 Hz to 8000 Hz for Room 1.

Surfaces	Area	Material	125 Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	8 kHz
Floor	40 m^2	carpet	0.07	0.31	0.49	0.81	0.66	0.54	0.48
Ceiling	40 m^2	Concrete	0.02	0.02	0.03	0.03	0.04	0.05	0.05
Wall	84 m^2	Curtain	0.06	0.12	0.35	0.45	0.38	0.36	0.36

improvement from 4 delay lines to 8 delay lines. With 16 delay lines, the reverberant sound becomes smoother and is nearly as good as the RTM. Some listeners have confused the FDN-RTM with 16 delay lines with the reference. In general, the reverberant sounds are of higher quality for musical sources than for vocal sources, especially where a small number of delay lines is used. In conclusion, to obtain satisfactory sound quality for the late reverberation, it is suggested to use at least 8 delay lines for the FDN-RTM.

3.6.6 Frequency-dependent reverberation for realistic materials

For realistic reverberation simulation, the reverberation time is modeled in different octave bands following the magnitude response characteristics of the wall absorption filter.

As an example, the walls of Room 1 are selected with frequency-dependent absorption materials [Vorländer, 2008], as shown in Table 3.6. The source and the receiver are both assumed to be located at position [3,3,1.5] m.

A first-order low-pass filter is designed using Equation 3.22 to model the frequency-dependent characteristics of the absorption materials. The absorption coefficients at 125 Hz and 8000 Hz are combined to determine the low-pass filter coefficient b_p. The Energy Decay Relief (EDR) of the synthesized room impulse response is shown in Figure 3.19. The T_{60} at zero frequency is about 2.10 s, while at 22.05 kHz it is about 0.45 s.

Still, this first-order low-pass filter provides only a very crude approximation of the complex frequency-dependent absorption of real rooms. Using the Gauss-Newton method to fit the absorption coefficients and design the frequency-dependent filter can increase the modeling accuracy of the frequency characteristics. Figure 3.20 shows the T_{60} of octave bands of the modeled impulse response using third-order filters, comparing to the Sabine calculated octave band T_{60} and the RTM modeled T_{60}. The Sabine prediction of the reverberation time in different octave bands is calculated as [Kuttruff, 2009]:

$$T_{60,Sab} = \frac{0.161V}{\sum_i A_i \alpha_i(\omega)} = \frac{0.161V}{\sum_i A_i (1 - |H(e^{j\omega}|^2)}. \tag{3.43}$$

It can be seen that the low-pass property is well modeled using a third-order filter, although

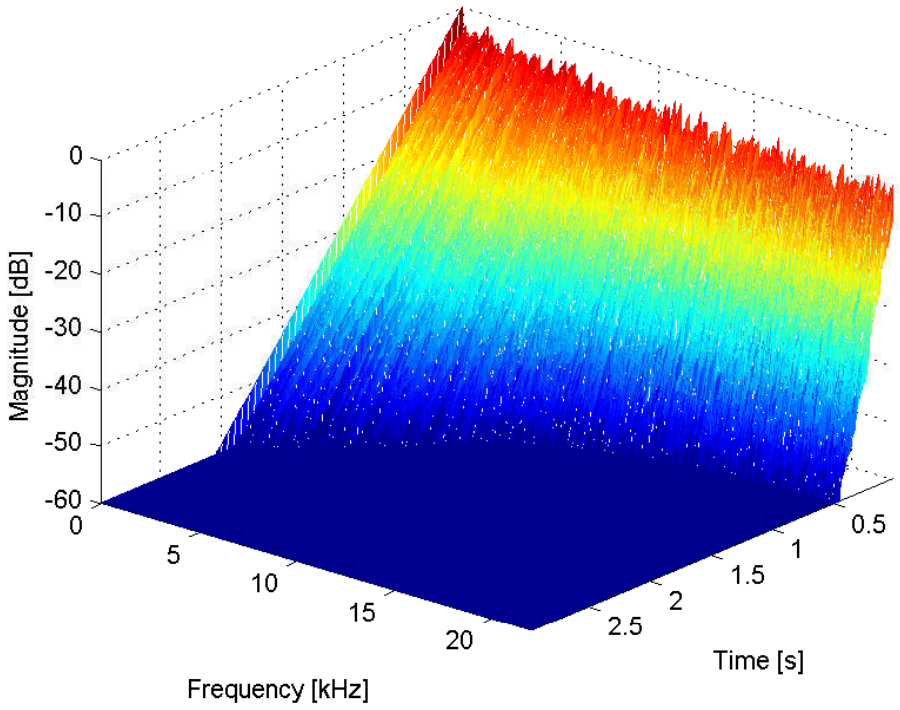


Figure 3.19: Frequency-dependent energy decay for Room 1 modeled with the low-pass filter.

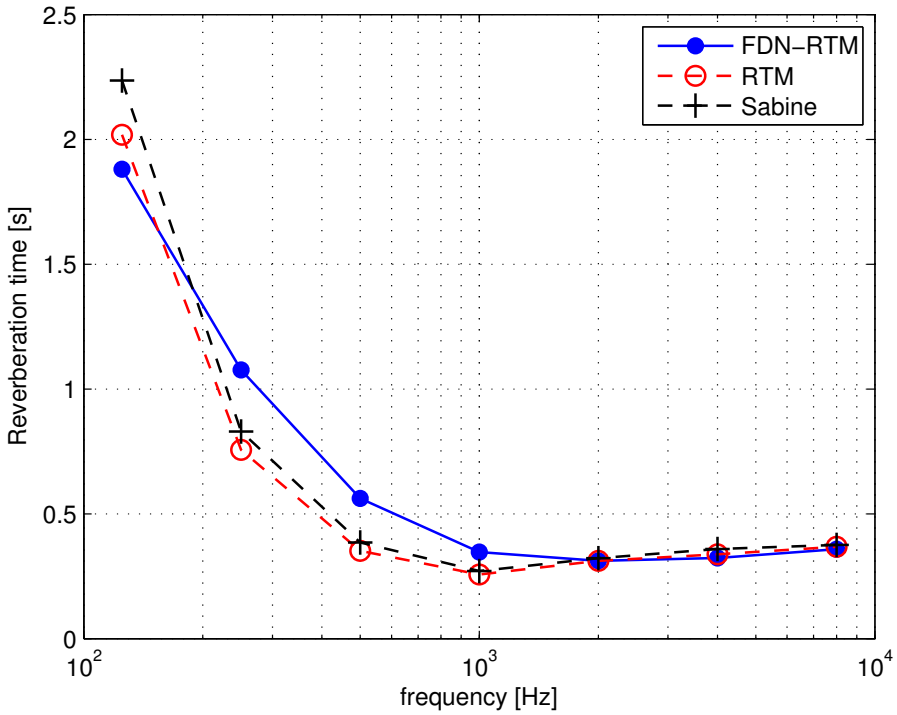


Figure 3.20: Reverberation time in different octave bands. The FDN-RTM uses third-order filters to model the frequency-dependent characteristics of the wall materials.

Table 3.7: Computation time consumed for FND-RTM using 8 delay lines with various orders of frequency-dependent filters to compute a one-second RIR.

Filter order	no filter	1^{st} order	2^{nd} order	3^{rd} order
Computation time (s)	0.187	0.214	0.234	0.257

the transition from the high-amplitude frequency band to the low-amplitude frequency band is slower compared to the ideal RTM and Sabine estimations. In our experiments, we choose the third-order filter for realistic reverberation synthesis, as it is sufficient to model the frequency dependent characteristics for most materials. Further increasing the order of the filter brings in little improvement to the modeled frequency characteristics while also increasing the computation time. The computation time using filters of increasing order is summarized in Table 3.7.

3.7 Conclusions

In this chapter we presented 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.

Chapter 4

Reverberator Using Variable Delays

In Chapter 3, the sound reflection paths between room surfaces are represented by feedback delay lines and the delay lengths of the delay lines are chosen as fixed values. However, the reflection paths have delay lengths that vary continuously within a whole interval, which is difficult to model by using only a few delay units in the feedback delay networks. Moreover, in the radiance transfer model, the reflection delays between discretized patches also vary over a certain range. The normal practice of modeling them using a fixed deterministic delay length can be problematic if the variation range can not be neglected. In this chapter, we investigate the variation of delay lengths in the radiance transfer model and the feedback delay networks. Extensions of using variable delay lengths are made to both methods. In order to achieve a linear time-invariant system for the feedback delay networks, a new structure is proposed in order to incorporate variable delay units in the artificial reverberator.

This chapter is organized as follows. Section 4.1 discusses the issues of the varying delay lengths between patches in the radiance transfer model. An extension of randomly variable delays is proposed to the model. Section 4.2 analyzes the delay lengths variation characteristics of a room, and proposes a grouping scheme different to Chapter 3 based on the nearest delay lengths principle. A statistical variable-delay comb filter is proposed and applied to the feedback delay networks. Since the statistical time-varying delay units are not LTI systems, a new feedback structure using deterministic delay groups is proposed. Simulation results are given in Section 4.3 and conclusions are drawn in Section 4.4.

4.1 Radiance Transfer Model Using Variable Delays

The radiance transfer method, described in Section 3.1, is an efficient model to simulate the diffuse reflections and the late reverberation. For numerical simulation, the geometry is discretized into patches, which results in discretization errors when calculating the radiosity exchange between patches, especially for the patches close to each other. In this section, we discuss such discretization errors, and propose an extension of the radiance transfer model using randomly variable delay lengths to reduce the modeling errors of the reflection delays.

4.1.1 Variable delays of the radiance transfer model

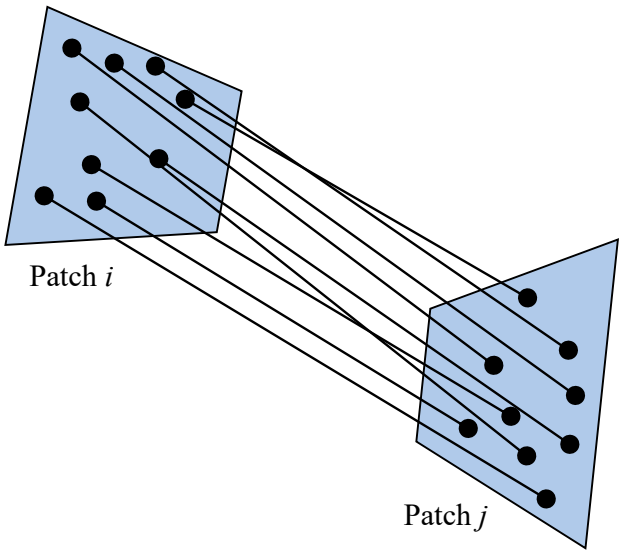


Figure 4.1: The surface of each patch is uniformly sampled and the distance between the sampling points are calculated. The distances between points of two patches, patch i and j, can vary in a wide range from its mean distance or central distance.

The numerical simulation of the radiance transfer model requires that the geometry of the room be discretized into small planar patches. Then the radiation density of patch i is given as its initial radiosity $B_i^0(t)$, plus the contribution of the reflected radiosity from other patches (as given in Section 3.1.2):

$$B_i(t) = B_i^0(t) + (1-\alpha_i) \sum_{j=1, j\neq i}^{M} F_{i,j} B_j(t - \frac{R_{i,j}}{c}). \tag{4.1}$$

In Equation 4.1, the distance $R_{i,j}$ between patch i and patch j is often taken as the mean dis-

tance between the two patches or the distance between the centroid of the two patches. However, this results in quantization errors since the distance between the points of two patches varies in a wide range, especially for the patches close to each other. Take two patches as illustrated in Figure 4.1 for example. Each patch surface is uniformly sampled and the distances between the sampling points are used to calculate the distribution of distances between patches.

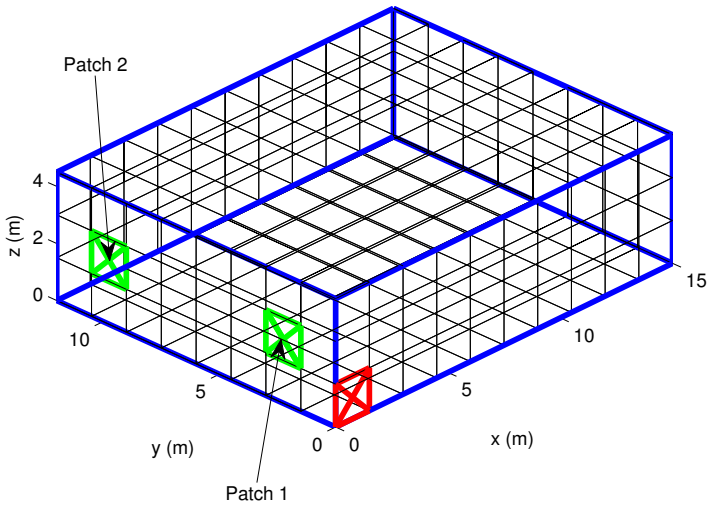


Figure 4.2: Geometry of a rectangular room with dimension $15 \times 12 \times 4.5$m. The distances between patch marked in red and the patches marked in green (patch 1 and patch 2) are calculated to show the distribution of the distances between them.

The distribution of distances between patches can be used to better allocate delay lengths in the radiance transfer model or the feedback delay networks method. Take a geometry of rectangular room with dimension $15 \times 12 \times 4.5$m, as in Figure 4.2, for example. The distribution of the distances between the patch marked in red in Figure 4.2 and the two patches marked in green are plotted in Figure 4.3. The distances between the red patch and Patch 1 (the one closer to the red patch) is nearly Gaussian distributed, while the distribution for patch 2 (the further one) has a rather flat distribution between the range of $[1230, 1400]$ samples. In both cases, it is clear that distances between patches are distributed over wide ranges of delays. Thus it is natural to extend the distance term $R_{i,j}$ in Equation 4.1 and use a time-varying term $R_{i,j}(t)$. The term $R_{i,j}(t)$ varies randomly according to the delay distributions of the two patches considered. With the extension of variable delays, Equation 4.1 becomes:

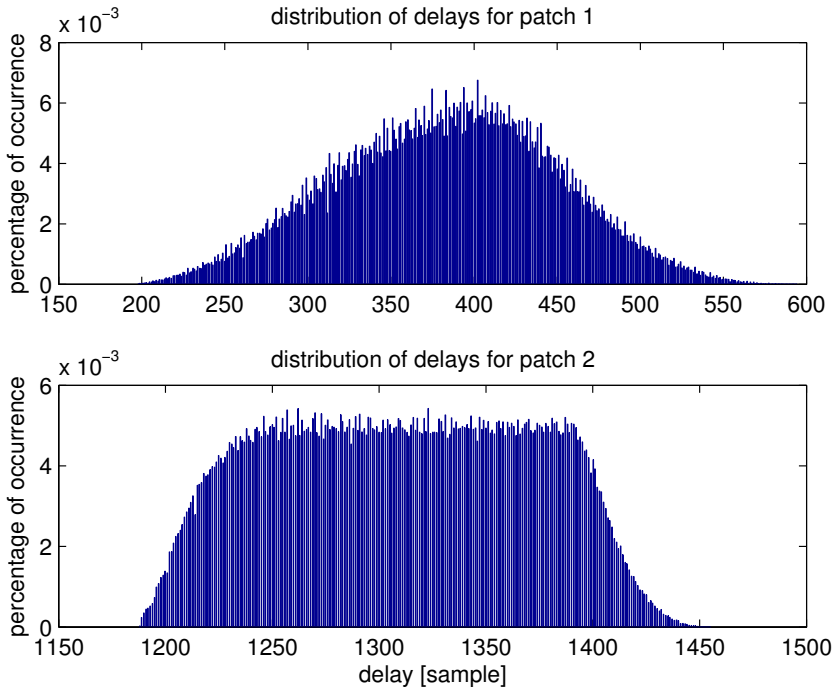


Figure 4.3: [Top] Distribution of delays between the reference patch (marked in red in Figure 4.2) and patch 1. [Bottom] Distribution of delays between the reference patch and patch 2. The delays are calculated in samples, using a sampling rate of 44.1 kHz.

$$B_i(t) = B_i^0(t) + (1-\alpha_i) \sum_{j=1, j\neq i}^{M} F_{i,j} B_j(t - \frac{\mathbf{r}_{i,j}^{(t)}}{c}), \tag{4.2}$$

where $\mathbf{r}_{i,j}^{(t)}$ varies according to the delay distribution $\mathcal{P}_{i,j}$ between patch i and j:

$$\mathbf{r}_{i,j}^{(t)} \sim \mathcal{P}_{i,j}. \tag{4.3}$$

In the progressive expression of the radiance transfer model, the time-varying delay term $\mathbf{r}_{i,j}^{(t)}$ takes a different realisation $\mathbf{r}_{i,j}^{(n)}$ at each iteration. The iterative radiance transfer method is expressed as:

$$B_i^{(n)}(t) = B_i^{(n-1)}(t) + (1-\alpha_i) \sum_{j=1, j\neq i}^{M} F_{i,j} B_j^{(n-1)}(t - \frac{\mathbf{r}_{i,j}^{(n)}}{c}). \tag{4.4}$$

4.1.2 Modification to the computation procedure

Introducing time-varying delays does not invoke much changes to the computational procedure of the radiance transfer method. The only difference compared with the traditional radiance transfer method is to randomly generate a delay length according to the delay distribution during each iteration of the computational procedure.

During the pre-computation of the traditional radiance transfer method, the order-1 patch-to-patch responses are pre-computed. The high order patch-to-patch responses are obtained by convolving the order-1 patch-to-patch responses. In contrast, by introducing the time-varying delays, the high order responses can not be obtained in the same way since the order-1 patch-to-patch responses are different in each iteration due to the time-varying delays. Thus for the delay-varying radiance transfer method, the high-order patch-to-patch responses are pre-computed in a progressive manner. First, the patch-to-patch form factors are calculated and are kept invariant during all the iterations. Then assume that the computation procedure is currently in the n^{th} iteration and the previous order-$(n-1)$ patch-to-patch responses are already computed. For the n^{th} iteration, a delay length is randomly generated, according to the delay distribution between the patches considered. Gaussian and uniform distributions can be used to approximate the delay distribution between patches. Together with the form factor, the order-1 patch-to-patch response is formed. The newly generated order-1 patch-to-patch response is convolved with the previous order-$(n-1)$ patch-to-patch responses to form the order-n patch-to-patch responses. The progression continues until the desired number of iterations are reached. The high-order patch-to-patch responses are stored for run-time simulation.

The run-time computation, including the initial-shooting and final-gathering stages, is the same as the traditional radiance transfer method.

4.2 Feedback Delay Networks Using Variable Delays

In Chapter 3, the delay lines of the feedback delay networks are linked with the radiance transfer methods by using one delay line to represent a group of patch-to-patch interactions. The groups of the patch-to-patch interactions are designed by the uniform energy grouping scheme in order to obtain an unitary feedback matrix. Different from using the uniform energy grouping scheme, in this section, we divide the groups based on nearest delay lengths principle. Variable delay lengths are used to statistically represent the different delay lengths in each group.

4.2.1 Delay distribution of a room

The delays of the reflection paths between surfaces of a room varies in a wide range. For diffuse reflections, this can be analyzed using the radiance transfer model, with the patch size approaching to infinitesimal. The one-order reflection delays of the room shown in Figure 4.2 are analyzed using the radiance transfer model with the patch size of 0.1×0.1m. The distribution of the delays is shown by the upper plot of Figure 4.4.

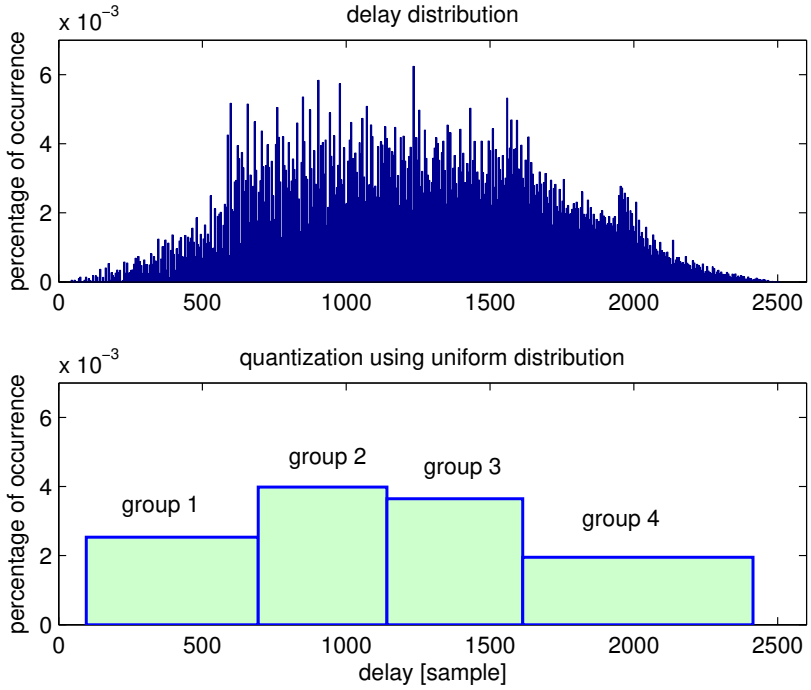


Figure 4.4: Distribution of delays between the patches for the room in Figure 4.2. The delay lengths are divided into four uniformly distributed groups.

The delays are distributed over a wide range. The delays start shortly after 0, which mainly come from the reflections between neighboring patches. The delays are concentrated between the range of $[598, 1594]$ samples, where 598 is approximately the sample delay of 4.5m, and 1598 is approximately the sample delay of 12m. There is another peak in the distribution around 2008 sample delays, which corresponding to the sample delay of 15m. The statistical distribution of the delay reflects to a certain degree the dimensions of the room.

In the case of diffuse late reverberation modeling, the precise modeling of the delays and amplitudes of each pulses is often difficult and unnecessary. We thus focus on modeling statistically the global energy exchange between the surface. We approximate the radiance transfer model,

grouping reflection paths with similar delays. As an example shown in Figure 4.4, the probability distribution of the delay lengths of the room are divided into four groups based on nearest delays principle, each of them containing equal accumulated probability. Each group is then approximated by an uniform distribution. The uniform distribution boundaries of each group are denoted as $[B_i^L, B_i^H]$, where i is the index of the group. By neglecting the very small and very large delay lengths since they have small probabilities, the delay distribution is divided into four uniformly distributed groups, with grouping boundaries as $[96, 694, 1141, 1614, 2416]$ samples.

In the rest of the chapter, uniform distribution is applied for the reason of simplicity. However, other distribution types can also be used, such as the Gaussian distribution illustrated in Figure 4.5. We use uniform and Gaussian distributions because it is easy to draw random samples from them. Drawing random samples from arbitrary distributions can be difficult.

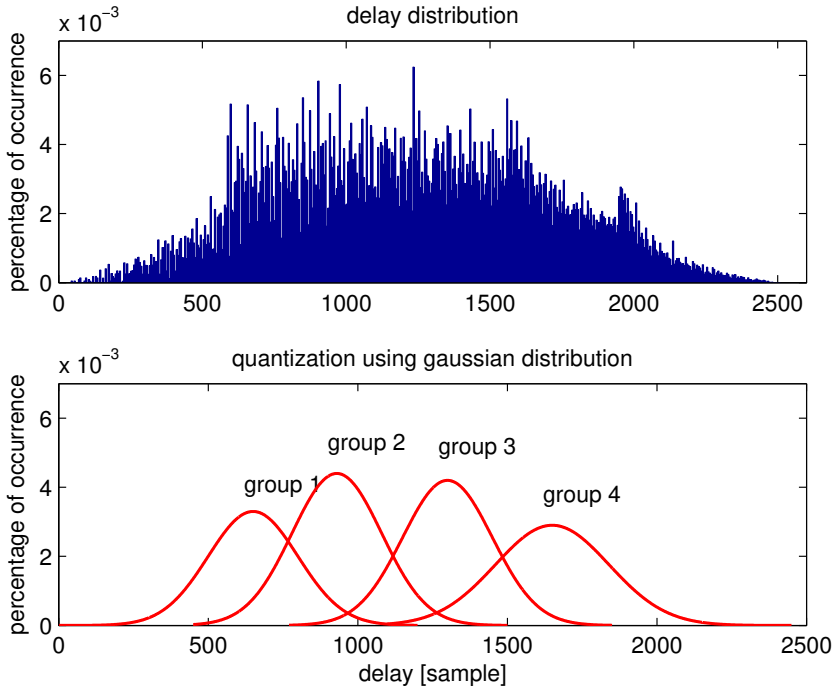


Figure 4.5: The delay length distribution of Figure 4.4 is divided into four groups, and each of them is represented by a Gaussian distribution.

4.2.2 Delay-varying comb filter

The building block of the various artificial reverberators are recursive feedback delay lines. The simplest recursive feedback can be represented by a comb filter [Schroeder, 1962], as illustrated

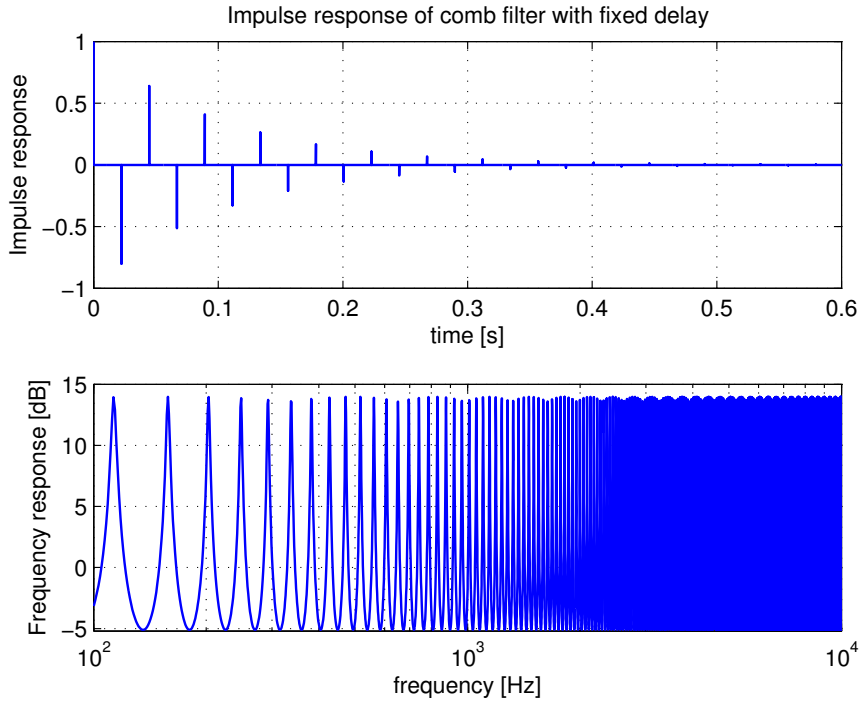


Figure 4.6: The impulse response and the frequency response of comb filter with fixed delay length of 983 samples.

in Figure 2.4 in chapter 2. It is well known that the comb filter with fixed delay length produces signals with metallic sound, which is explained, in Figure 4.6, by the comb-shape of its frequency response. The human perception system can not hear the modal frequencies when the modal density is high enough. Since the delay length is proportional to the model density, they use large delay lengths to reduce the distances between the peaks in the frequency responses. However, long delays result in sparse echoes, which is undesirable for late reverberation. Multiple comb filters are also used which are connected in parallel to increase the density of the peaks and to make the dense peaks inaudible by our perception system.

The comb-shaped peaks of frequency responses come from the regular repetition of the pulses with the same delays. By introducing randomness of the delay lengths, the comb-shape of the frequency responses can be removed. Thus we use a time-varying delay unit to replace the delay unit of the comb filter, whose delay length $d(m)$ changes randomly over a predefined range, as shown in Figure 4.7. By removing the fixed delay taps, the regular comb-shape in its frequency response is smoothed, as shown in Figure 4.8. The delay length varies within an uniform distribution, with its mean at 983 samples and the varying interval $[983 - 500, 983 + 500]$ samples. The metallic sound is greatly reduced while the decaying speed is statistically the

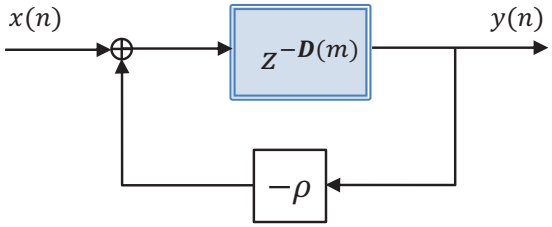


Figure 4.7: The comb filter with variable delay length.

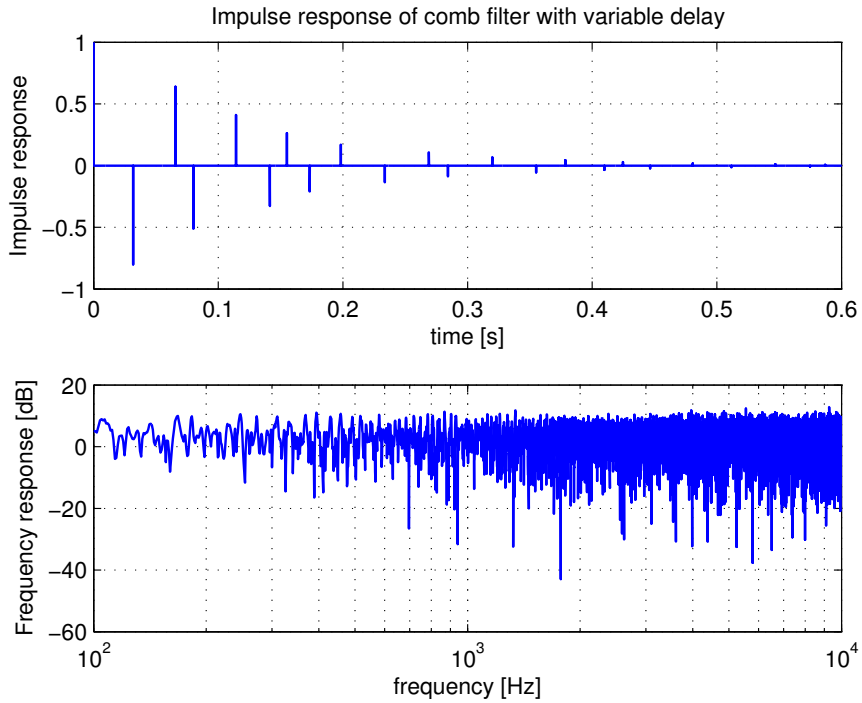


Figure 4.8: The impulse response and the frequency response of comb filter with variable delay length.

same as with the traditional comb filter with a fixed delay by adjusting the attenuation factors accordingly.

4.2.3 Recursive networks using variable delay lengths

Using delay-varying comb filters, each group of the reflection path distribution in Figure 4.4 can be regarded as a delay line in the FDNs structure, whose delay lengths are set according to the pre-defined delay distribution. In Figure 4.9 we illustrate the delay-varying FDNs structure using four delay lines. This structure is similar to the feedback delay networks in Chapter 3 except

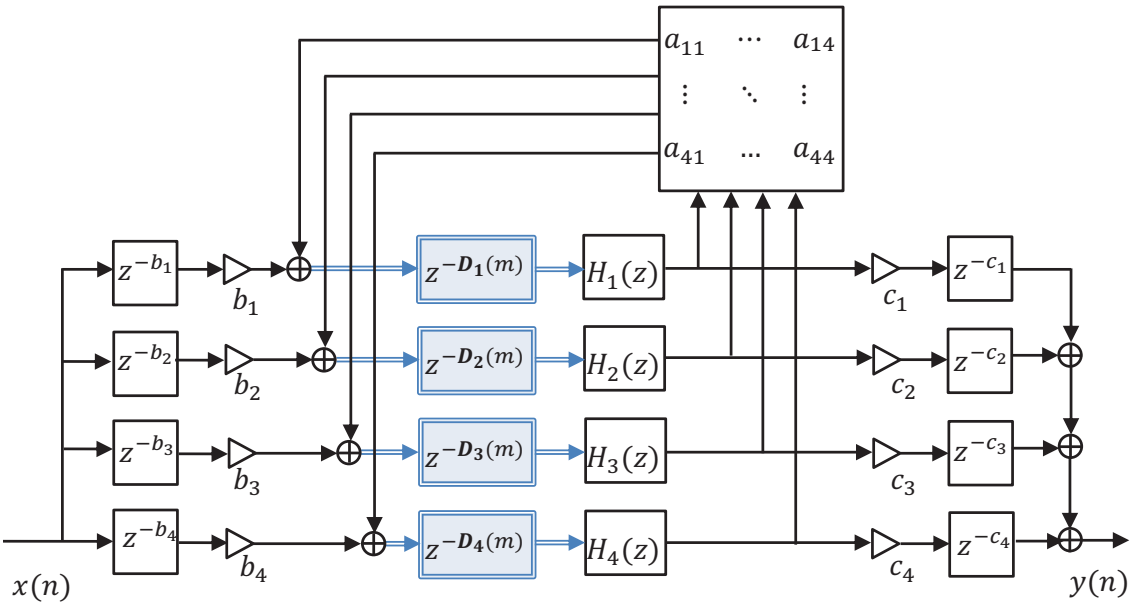


Figure 4.9: The system structure of the FDNs using variable delay lengths. Four delay lines are used in this example).

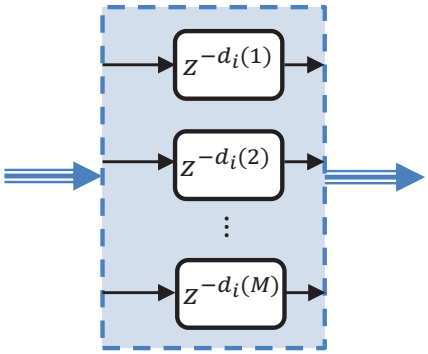


Figure 4.10: Deterministic implementation of the time-varying delay unit.

that each delay unit has a time-varying delay length which changes according to the uniform distribution within the grouping boundaries $[B_i^L, B_i^H]$. The delays of the impulse response produced by the varying-length feedback have uniform distributions in each group, which is a rough approximation of the distribution estimated by the radiance transfer modeled for the room. In order to reduce the quantization error of the distribution, more delay lines can be used. Thus the delay-varying feedback delay networks can be regarded as a simplified implementation of the radiance transfer method to produce the diffuse reflections.

Such varying-delay implementation can be used to produce delay-varying taps for off-line computation of the impulse response. However, it can not be used for on-line rendering of the reverberant sound as it is *not* a linear time-invariant (LTI) system. The input $x(n)$ of the networks is re-ordered and distorted by the time-varying delays. For example, if the input are five letters "h-e-l-l-o $[he - loh]$", by the randomly generated delays, the output signal can be distorted in the order, which may become "e-o-l-h-l", "e-l-o-l-h", or any order which is permitted by the maximum variation range. Thus it can be easily imagined that if the input is sampled music or speech signal, these samples can be randomly re-ordered, and the output, not surprisingly, may sound like random noise.

Thus for on-line rendering, the randomly generated delay lengths need to be replaced by deterministic values. We propose a deterministic implementation where each time-varying delay unit $z^{-\mathbf{D}_i(m)}$ is replaced by M basic delay units $[z^{-d_i(1)}, z^{-d_i(2)}, \cdots, z^{-d_i(M)}]$, connected in parallel, as in Figure 4.10. The delay lengths of the basic delay units $d_i(k)$ are taken as the instances from the delay length intervals $[B_i^L, B_i^H]$, following their probability distributions. The delay lengths of these basic delay units are unchanged during the rendering and thus guarantee the system to be LTI. During each feedback iteration, the basic delay units in delay group $[z^{-d_i(1)}, z^{-d_i(2)}, \cdots, z^{-d_i(M)}]$ are selected in the pre-defined order and their outputs are connected to other basic delay units within the same group and among other groups by the predefined connectivity matrix C.

An example showing the connectivity among delay units is given in Figure 4.11. The delay group 1 represents the variable delay line $z^{-\mathbf{D}_1(m)}$ and contains delays $[d_1(1), d_1(2), d_1(3), \cdots]$, and delay group 2 represents $z^{-\mathbf{D}_2(m)}$ and contains delays $[d_2(1), d_2(2), d_2(3), \cdots]$. In this figure, the delay line attenuations and the feedback matrix is ignored for conciseness.

The input dry signal $x(n)$ enters only into the input of one of the basic delay units in each group. We chose the 1^{st} basic delay unit of each group in the example. The input dry signal $x(n)$ passes the 1^{st} basic delay unit $d_1(1)$ in group 1. Then the output of the variable delay line $z^{-\mathbf{D}_1(m)}$ being fed back to itself in Figure 4.9 can be represented in a deterministic way, by the output of the 1^{st} basic delay unit $d_1(1)$ in group 1 being fed back to the input of one of the basic delay units of the same group. In this example, it is fed back to the 2^{nd} basic delay unit $d_1(2)$. The output of the variable delay line $z^{-\mathbf{D}_1(m)}$ being fed back to the variable delay line $z^{-\mathbf{D}_2(m)}$ in Figure 4.9 can be represented by the output of the 1^{st} basic delay unit $d_1(1)$ in group 1 being fed back to the input of the 2^{nd} basic delay unit $d_2(2)$ of group 2.

In the next iteration, a different basic delay unit can be chosen. For example, the output of the variable delay line $z^{-\mathbf{D}_1(m)}$ being fed back to itself is implemented by the output of the 2^{st} basic delay unit $d_1(2)$ in group 1 being fed back to the input of the 3^{rd} basic delay unit $d_1(3)$

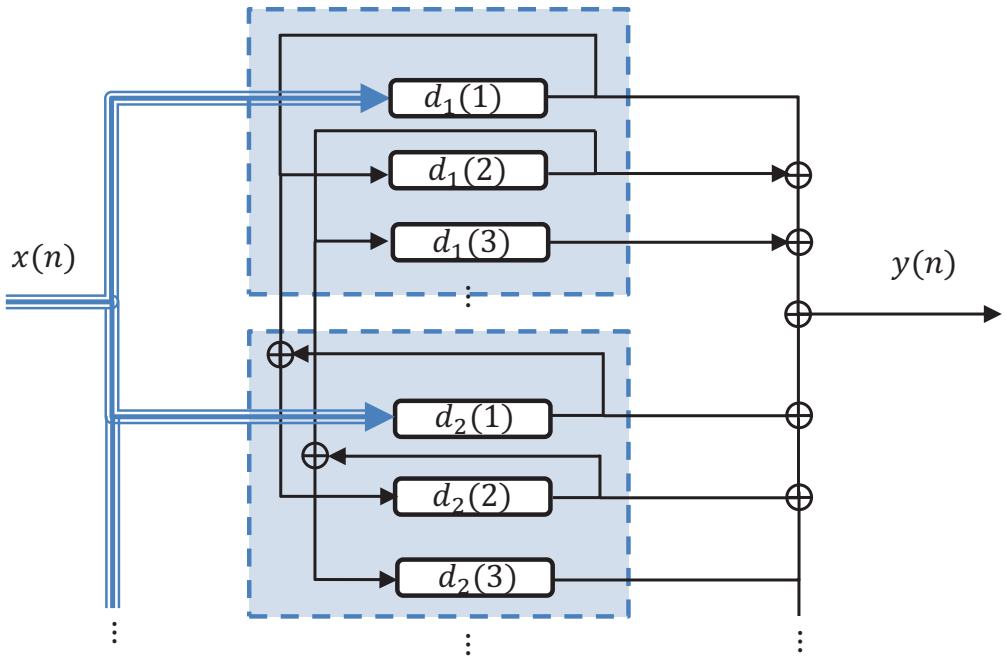


Figure 4.11: Deterministic implementation of the delay-varying feedback delay networks. In this figure, the delay line attenuations and the feedback matrix is ignored for conciseness.

of the same group. The output of the variable delay line $z^{-\mathbf{D}_1(m)}$ being fed back to the variable delay line $z^{-\mathbf{D}_2(m)}$ can also be implemented differently, by feeding back the output of the 2^{nd} basic delay unit $d_1(2)$ in group 1 to the input of the 3^{rd} basic delay unit $d_2(3)$ of group 2, etc. The output of all delay units are collected to form the output $y(n)$. The input delay units and the connectivity given in Figure 4.11 is only an example of simple implementation for the purpose of clear demonstration. The choice of basic delay units in each iteration and the connectivity between them can be designed randomly. But once designed, they are fixed during the process of auralization, which is the so-called deterministic implementation of the variable delay length.

The computational complexity of the feedback delay networks can be analyzed by the number of multiplications and additions required to produce one output sample. For the FDNs of order N, it requires $N \times N$ multiplications and βN additions, where β is a constant. The delay units of the original $N \times N$ FDNs are extended to variable length delay units, with each delay group containing M basic delay units. Since the connectivity between delay units are limited within and among delay groups, the total number of multiplications required is $M \times N^2$, which is much smaller than the $(M \times N)^2$ multiplications required by the traditional FDNs structure. The number of multiplications can be studied by the number of non-zeros of the feedback matrix where zeros denote non-connection between the corresponding delay units. The equivalent feedback

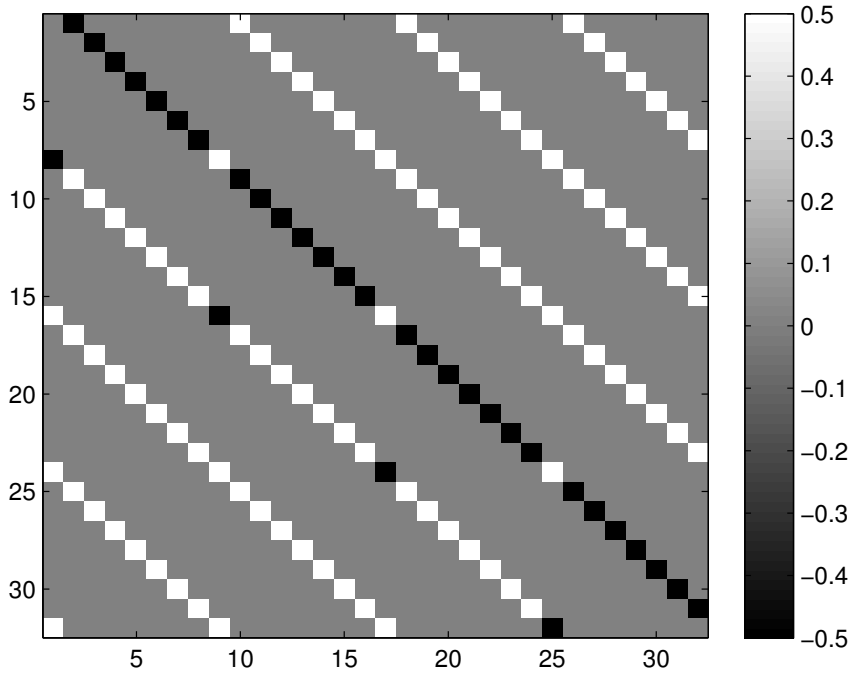


Figure 4.12: The feedback matrix of a 4×4 FDNs extended to 8 delay units in each delay line. The gray color denotes zeros in the figure

matrix of a 4×4 FDNs extended to 8 delay units in each delay line is shown in Figure 4.12. In this matrix, there are in total 128 non zeros elements, which is much smaller than the 1024 multiplications required by the full FDNs of order 16×16.

The randomly designed connectivity between basic delay units is equivalent to the introduction of a random permutation matrix P being multiplied with the feedback matrix C.

4.3 Experiments

4.3.1 Simulated RIRs

We evaluate the proposed delay-varying feedback delay networks using the geometry in Figure 4.2. Its delay distribution is given in Figure 4.4. We approximate the delays in each group using a uniform distribution. We simulate the reverberation using 4×4 FDNs as the reference. The delay lengths of the reference FDNs are chosen as prime numbers around the centroid of each group. The delay lengths are $[397, 907, 1373, 1861]$. Each delay group is extended to 4 delay units, uniformly distributed within the delay intervals. The impulse responses of the traditional 4×4 and 16×16 FDNs and their frequency response are given in Figures 4.13 and 4.14, and the impulse response and the frequency response of the extended $4 \times 4(4)$ delay-varying FDNs are given in Figure 4.15. Both generated impulse responses have a reverberation time of 2s.

It is shown that the modeled impulse response using the extended $4 \times 4(4)$ delay-varying FDNs has a smoother frequency response compared with that modeled using traditional 4×4 FDNs and 16×16 FDNs. The modal frequencies are denser for the $4 \times 4(4)$ delay-varying FDNs, which improves the synthesized sound quality and reduces the audibility of the metallic sound effects.

The $4 \times 4(4)$ delay-varying FDNs use in total 16 delays lines. The density of the pulses of the synthesized RIRs using $4 \times 4(4)$ delay-varying FDNs is higher than the traditional 4×4 FDNs but lower than the traditional 16×16 FDNs. The $4 \times 4(4)$ delay-varying FDNs use 4 input delay lines, which is the same as the traditional 4×4 FDNs, but collect the outputs of all the 16 delay lines, which is the case for the traditional 16×16 FDNs.

Extending the traditional FDNs using 8 delay units for each delay line, the statistics of the delay lengths of the whole delay-varying FDNs is obtained and compared with the delay distribution of the room. The delay distribution of the delay-varying FDNs is given in Figure 4.16. It is shown that the delays well approximate the designed uniform distribution in each group, and well represent the statistics of delays analyzed by the radiance transfer model.

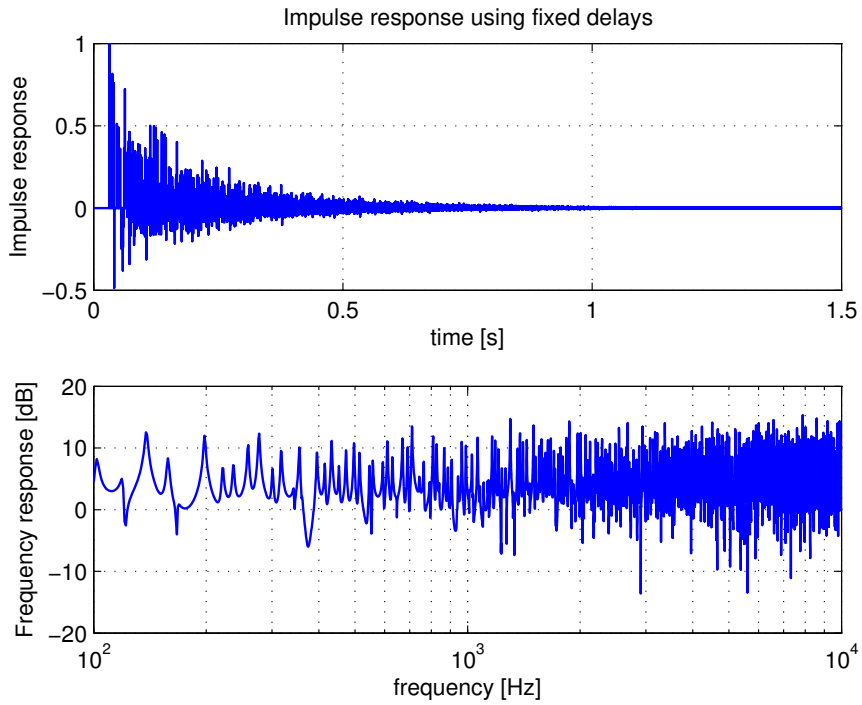


Figure 4.13: The impulse response and the frequency response of the traditional 4×4 FDNs.

4.3.2 Computational comparison

Theoretically, by extending each delay line using M deterministic basic delay units, the $4 \times 4(M)$ delay-varying FDNs requires M times more multiplications than the traditional 4×4 FDNs. In practice, using Matlab implementation, the computational time is slightly different. The computation time for rendering a one second signal using the traditional 4×4 FDNs is 0.083s, and that using the extended $4 \times 4(4)$ delay-varying FDNs is 0.197s, which is less than the theoretical analysis. The reason might be that using Matlab implementation, some computation is consumed by reading and writing the data in the memory, thus the real computation time does not well reflect the theoretical computational complexity of the two algorithms. Further analysis of the computational complexity requires optimized implementation in C or C^{++} language.

A comparison of the computation time used for $4 \times 4(M)$ delay-varying FDNs and the traditional 4×4 FDNs is given in Figure 4.17. For the traditional FDNs, the number of delay units are used as a design parameter to determine the system complexity and sound quality. Generally speaking, the more delay units used in the FDNs, the better sound quality can be. In Figure 4.17, with the same number of delay units, the computation time required by the delay-varying FDNs is less than that required by the traditional FDNs. It is mainly due to the limited connectivity

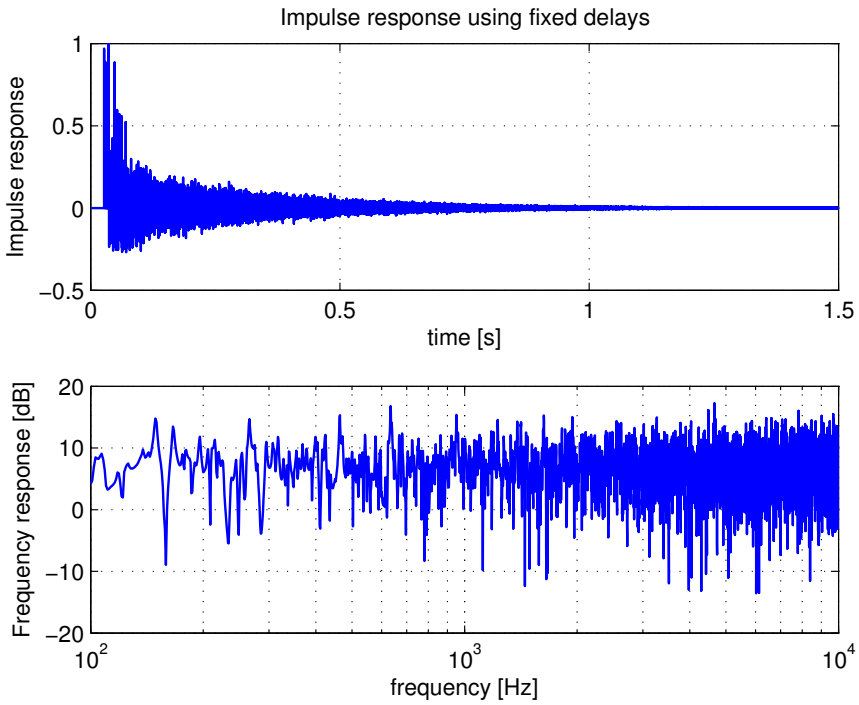


Figure 4.14: The impulse response and the frequency response of the traditional 16×16 FDNs.

between delay units, as reflected by the sparse feedback matrix in Figure 4.12.

4.3.3 Sound quality

Informal listening tests reveals that the sound quality using the delay-varying FDNs has great improvements over the traditional FDNs. The sound quality using $4 \times 4(4)$ delay-varying FDNs is much better than the traditional 4×4 FDNs where the metallic sound is almost not audible at all. Both using 16 delay lines, the sound quality using $4 \times 4(4)$ delay-varying FDNs is even better than the 16×16 FDNs in some cases, especially for speech signals that had a mettalic sound. The reason is that by extending each delay line to several basic delay units and the pre-define deterministic connectivity among them, regular repetition of the pulses is much reduced and a much smoother frequency response can be obtained. This can be explained using the example in Figure 4.11. For example, the input signal of the $1^{\mathbf{st}}$ basic delay unit $d_1(1)$ in group 1 is feedback to the $2^{\mathbf{nd}}$ basic delay unit of the same group, and then to the $3^{\mathbf{rd}}$ basic delay unit before going back again to the $1^{\mathbf{st}}$ basic delay unit. The total length of the delay path between two repetitive pulses is then $d = d_1(1) + d_1(2) + d_1(3)$. This actually increases the delay length, and thus increases the modal density of the frequency response.

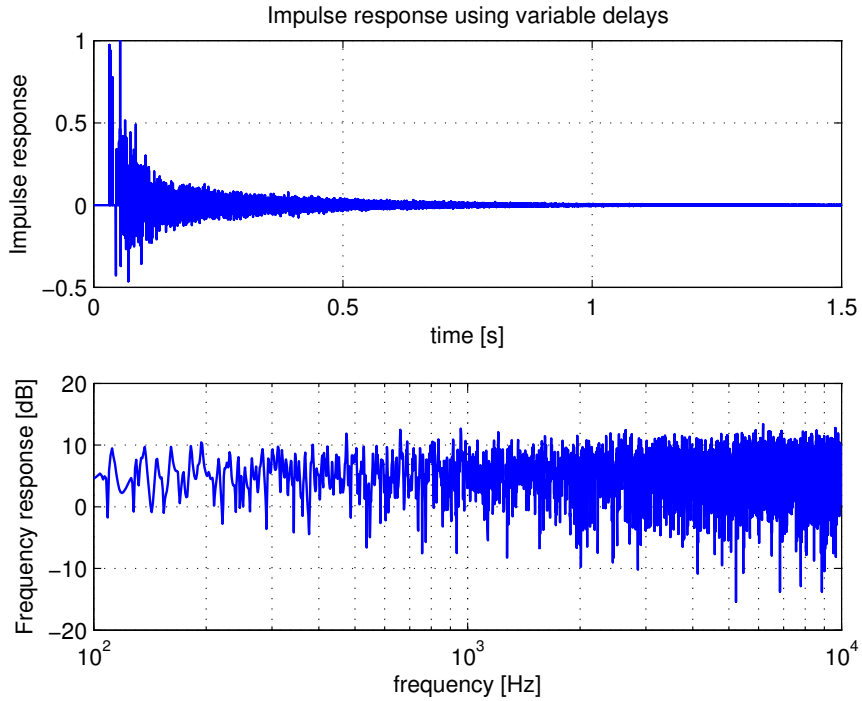


Figure 4.15: The impulse response and the frequency response of the extended $4 \times 4(4)$ delay-varying FDNs.

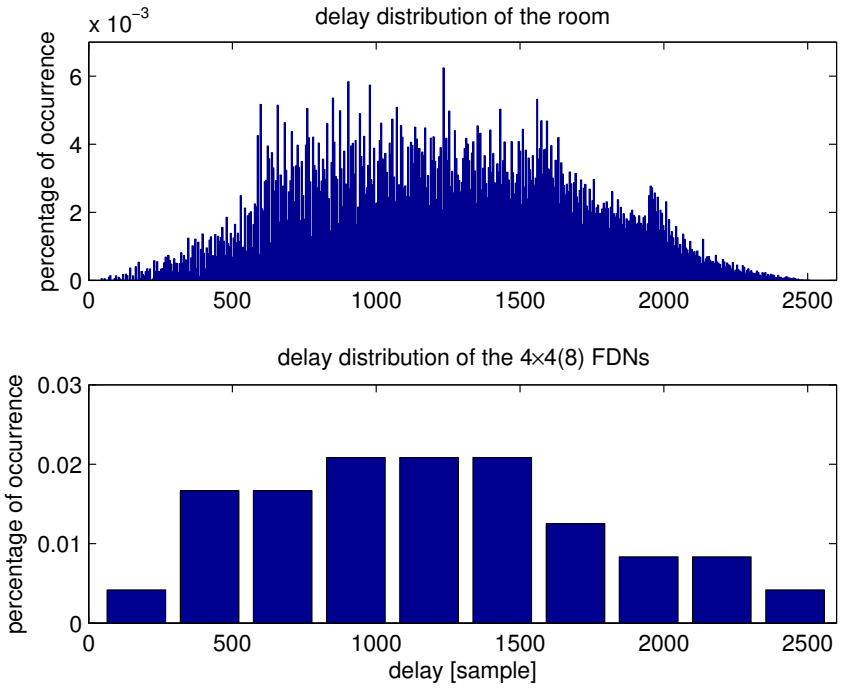


Figure 4.16: The distribution of the delays of the extended $4 \times 4(8)$ delay-varying FDNs with comparison of the delay distribution of the room.

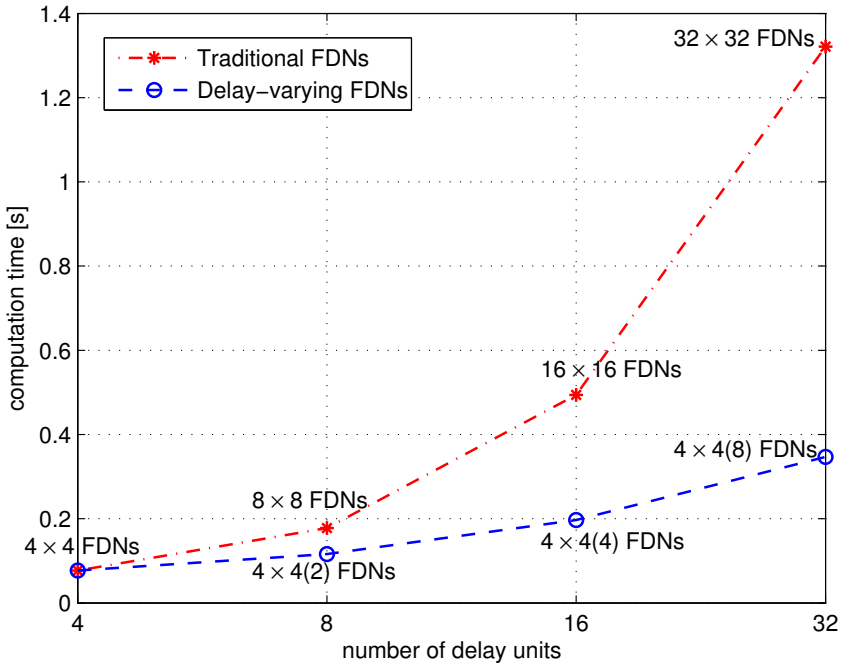


Figure 4.17: The comparison of the computation time to compute a RIR of duration 1 s consumed by the $4 \times 4(M)$ delay-varying FDNs and the traditional $N \times N$ FDNs. M is the number of basic delay units used for the extension of the deterministic delay-varying FDNs.

The delay paths between basic delay units can have multiple different combinations. For example, for a $4 \times 4(8)$ delay-varying FDNs where each of the delay line is extended to have 8 deterministic basic delay units, the total number of combination is:

$$N_p = M^N = 8^4. \tag{4.5}$$

If the average delay length is, in this example, 1327 samples, then the synthesized RIRs will only have a repetitive pulse every 1327×8 samples. It is in contrast of the traditional FDNs where the pulses repeat themselves on average every 1327 samples. For a RIR of duration 1 s with 44100 samples, the repetitions only occurs about 4 times during the whole RIR. Besides, the pulses are already attenuated a lot in their next repetitions. This is the explanation for the reduction of the metallic perception of the reverberant sound.

4.4 Conclusion

In this chapter, we studied the delay variation phenomenon of the radiance transfer method and the feedback delay networks. For the radiance transfer method, in order to better model the delay variation of the sound exchange between patches, a variable-delay radiance transfer method was proposed and the corresponding modification is made for the simulation procedure. For the feedback delay networks, the variable-delay feedback delay networks was proposed which can be used as approximated simple realization of the radiance transfer method for off-line computation of the IRs. For on-line rendering of the reverberant sound, a new feedback delay structure is proposed, which is a deterministic implementation of the variable delay units. Experimental results show that the proposed feedback delay structure greatly improves the sound quality of the rendered signal, and the delay distribution of the feedback delay structure approaches the delay distribution of the room.

Chapter 5

Modeling Early Reflections Using Acoustic Rendering Equation

In Chapter 3, Radiance Transfer Method (RTM) was investigated, which is an efficient tool to model diffuse reflections. In [Kiminki, 2005; Siltanen et al., 2007], an Acoustic Rendering Equation (ARE) was formulated, as an extension of the RTM, which considers both diffuse and non-diffuse reflections. Since the high-order reflections computation is pre-computed and decoupled from the run-time computation procedure, this method is well suited for dynamic applications with moving sources. A case study conducted by [Siltanen et al., 2007] has shown that the simulated RIR using this method can predict several acoustic parameters with acceptable accuracy, and thus gives a good estimation of the RIR decay curve, especially for the late reverberation.

However, the accuracy of the modeled early reflections is not fully discussed in the literature. Hybrid methods were proposed in [Antani et al., 2012b; Lewers, 1993] where ARE method is used to model only the diffuse reflections and the late reverberation while the geometric methods are used to model the more important early 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 the early and the late part. To that aim, we study in this chapter the modeling accuracy of the early echoes for the ARE method and modify the final gathering schemes using ray tracing methods which allows us to model the early part while keeping the advantages of the original method for late reverberation simulation.

This chapter is organised as follows. In section 5.1 we recall the basics of the ARE method and the main computation procedure. The formulation of the ARE presented in this section mainly comes from [Siltanen et al., 2007]. In section 5.2 we explain our computational procedure using ray tracing method. In section 5.3 the modeling accuracy of the ARE method for early echoes

is studied and two final-gathering schemes are compared. Some experimental comparison are given in section 5.4.

5.1 Acoustic Rendering Equation

The ARE model shares similar concepts with the RTM, but is constructed in more general conditions. In this chapter, the complete formulation of the ARE model will be presented, although some equations may seem similar with the RTM in Chapter 3. We emphasise their difference and the advantages of the ARE model over the RTM.

5.1.1 Kajiya's Rendering Equation

The formulation of the room acoustic rendering equation originates from the Kajiya's rendering equation in computer graphics [Kajiya, 1986]. Different from the room acoustic modeling, the propagation time of light rays can be neglected since the speed of light, in most practical cases, can be assumed as infinite. The time-independent of the Kajiya's rendering equation expressed in a recent form [Dutre et al., 2006] is:

$$L(\mathbf{x} \rightarrow \Omega) = L_0(\mathbf{x} \rightarrow \Omega) + \int_S \rho(\mathbf{x}, \Psi \rightarrow \Omega) L(\mathbf{x'} \rightarrow -\Psi) V(\mathbf{x}, \mathbf{x'}) G(\mathbf{x}, \mathbf{x'}) \mathrm{d}\mathbf{x'} \qquad (5.1)$$

where

- $\mathbf{x}$ and $\mathbf{x'}$ belong to $\mathbb{R}^3$, and are the points on the surfaces in 3D space.
- The symbol $\rightarrow$ means the "in the direction of". For example $(\mathbf{x} \rightarrow \Omega)$ means "leaving point $\mathbf{x}$ in the direction Ω", and $(\mathbf{x}, \Psi \rightarrow \Omega)$ means "incident on point $\mathbf{x}$ from the direction Ψ and leaving in the direction Ω".
- $L(\mathbf{x} \rightarrow \Omega)$ is the total outgoing radiance from point $\mathbf{x}$ in the direction Ω.
- $L_0(\mathbf{x} \rightarrow \Omega)$ is the current emitted radiance from point $\mathbf{x}$ in the direction Ω. It is zero if point $\mathbf{x}$ is not a light or sound source.
- $\rho(\mathbf{x}, \Psi \rightarrow \Omega)$ is the bidirectional reflectance distribution function (BRDF) which defines how the incoming radiance from the direction Ψ is reflected to the direction Ω at point $\mathbf{x}$. It is related to the material reflectance characteristics of the surface [Nicodemus et al., 1977].
- $L(\mathbf{x'} \rightarrow -\Psi)$ is the radiance at point $\mathbf{x'}$ in direction $-\Psi$, the direction toward $\mathbf{x}$.

- $V(\mathbf{x}, \mathbf{x'})$ is the visibility term between point **x** and **x'**.
- $G(\mathbf{x}, \mathbf{x'})$ is the geometric term.
- $\int_S \mathrm{d}\mathbf{x'}$ represent the integration of the set of all the points on surface of the enclosure S.

5.1.2 Temporal Propagation Operator

For the acoustic rendering equation, the sound energy exchange is described by the acoustic radiance, which contains directional information of radiation. The symbol for acoustic radiance $L(\mathbf{x}, \Omega)$ is extended to a time-dependent notation $L(\mathbf{x}, \Omega, t)$, where **x** is the position vector in 3D space and Ω is the direction of radiation. For the radiance transfer method in Chapter 3, the energy exchange between surface patches is described using radiation density, which is a position and time variant quantity, noted as $B(\mathbf{x}, t)$. Since the RTM only models diffuse reflections, which assumes uniform energy distribution in all directions, the radiation density only depend on the position **x** and the directional information Ω is thus not considered.

The temporal effects of acoustic wave propagation need to be taken into account for acoustic radiance modeling. The temporal operator which denotes the sound propagation over a linear absorptive medium is defined as:

$$S_r\{I(t)\} = e^{-\alpha r} I(t - \frac{r}{c}), \tag{5.2}$$

where $S_r\{\cdot\}$ is the temporal propagation operator that applies temporal delay and absorptive attenuation to the original sound intensity $I(t)$, c is the speed of sound, and α is the absorption coefficient of the medium.

The temporal operator $S_r\{\cdot\}$ has the linearity and additivity property. If an impulsive sound is emitted at $t = 0$ and denoted as a Dirac delta function $\delta(t)$, then the linearity and additivity property can be expressed as:

$$S_r\{\mu_1\delta(t) + \mu_2\delta(t)\} = \mu_1 S_r\{\delta(t)\} + \mu_2 S_r\{\delta(t)\} \tag{5.3}$$

$$S_{r_1}\{S_{r_2}\{\delta(t)\}\} = S_{r_1+r_2}\{\delta(t)\}. \tag{5.4}$$

The linearity property in Equation 5.3 allows the differentiation, and enables the temporal propagation operator to be applied to the irradiance and the radiance. Equations 5.3 and 5.4 justify the combination of the propagation operators in the case of multiple specular reflections.

The intensity, or the radiance, at a point can be computed as the sum of the direct path and several early reflection paths.

Consider sound source located at an arbitrary position in a room with emission intensity $I_0(t)$ varying over time. The received intensity at any other location is the summation of the propagation operators from the source to the receiver convolved with the original intensity. It is expressed as:

$$I(t) = \left[\sum_{i}^{N} \beta_i S_{r_i}\{\delta(t)\}\right] * I_0(t), \tag{5.5}$$

where β_i is the reflection attenuation coefficients, which take into accounts the various wall absorptions that their propagation paths encounter. r_i are the lengths of the reflection paths, and $I(t)$ is the detected intensity at the receiver. Note that the temporal delays of each reflection path $e^{-\alpha r_i}$ are hidden in the propagation operators S_{r_i}.

5.1.3 Reflection Kernel

The reflectance function, the geometric term and the visibility term together describe the reflection from point **x** via point **x'** into direction Ω, as in Figure 5.1, and is referred to as **reflection kernel**:

$$R(\mathbf{x}, \mathbf{x'}, \Omega) = V(\mathbf{x}, \mathbf{x'})\rho\left(\frac{\mathbf{x} - \mathbf{x'}}{|\mathbf{x} - \mathbf{x'}|}, \mathbf{x'} \rightarrow \Omega\right) g(\mathbf{x}, \mathbf{x'}), \tag{5.6}$$

where $V(\mathbf{x}, \mathbf{x'})$ is the visibility term, $\rho(\frac{\mathbf{x}-\mathbf{x'}}{|\mathbf{x}-\mathbf{x'}|}, \mathbf{x'} \rightarrow \Omega)$ is the reflectance function (or BRDF), and $g(\mathbf{x}, \mathbf{x'})$ is the geometric term.

The reflection kernel of the ARE uses three variables **x**, **x'** and Ω to define both the incident direction and the reflected direction. (The incident direction is defined by **x** and **x'**, the reflected direction is defined by Ω.) It defines how the incoming radiance from point **x** to point **x'** is reflected into direction Ω. In contrast, the equivalent reflection kernel of the RTM in Chapter 3, defines how the radiation density of patch i affects the radiation of patch j. Compared to the RTM, the bidirectional extension of the reflection kernel enables the ARE to model arbitrary reflection patterns including specular reflection, diffusion and even diffraction.

Visibility term

The visibility term describes if the line between two points **x** and **x'** is obstructed by an obstacle. This extension is made in order to model nonconvex geometries. The visibility term is defined

as follows:

$$V(\mathbf{x}, \mathbf{x'}) = \begin{cases} 1, & \text{when the line between } \mathbf{x} \text{ and } \mathbf{x'} \text{ is unobstructed;} \\ 0, & \text{otherwise.} \end{cases} \tag{5.7}$$

Geometric term

The geometric term is defined as:

$$g(\mathbf{x}, \mathbf{x'}) = \left(n(\mathbf{x}) \cdot \frac{\mathbf{x} - \mathbf{x'}}{|\mathbf{x} - \mathbf{x'}|} \right) \left(n(\mathbf{x'}) \cdot \frac{\mathbf{x'} - \mathbf{x}}{|\mathbf{x'} - \mathbf{x}|} \right) \frac{1}{|\mathbf{x} - \mathbf{x'}|^2}, \tag{5.8}$$

where $n(x)$ is the surface normal of the patch that contains point $\mathbf{x}$ (see Fgure 5.1). The term $\left(n(\mathbf{x}) \cdot \frac{\mathbf{x}-\mathbf{x'}}{|\mathbf{x}-\mathbf{x'}|} \right)$ is the dot product of the unit vector of outgoing direction $\frac{\mathbf{x}-\mathbf{x'}}{|\mathbf{x}-\mathbf{x'}|}$ and the surface normal $n(\mathbf{x})$. It can also be expressed as the cosine of the outgoing angle.

The geometric term g scales as the inverse square of the distance between $\mathbf{x}$ and $\mathbf{x'}$. The non-orthogonality divergence term reaches its maximum of 1 when the line between $\mathbf{x}$ and $\mathbf{x'}$ is perpendicular to the surface patches of both points.

Reflectance term

The reflectance term describes the fraction of energy reflected along a certain direction from an incident sound. It is also referred as Bidirectional Reflectance Distribution Funtion (BRDF). Suppose the incident acoustic radiance from point $\mathbf{x}$ to point $\mathbf{x'}$ is reflected along a solid angle Ω, the reflectance function is defined as the ratio between the differential of the incident radiance and the differential of the reflected radiance as:

$$\rho \left(\frac{\mathbf{x} - \mathbf{x'}}{|\mathbf{x} - \mathbf{x'}|}, \mathbf{x'} \to \Omega \right) = \frac{\mathrm{d}L(\mathbf{x'}, \Omega)}{\mathrm{d}L\left(\mathbf{x'}, \frac{\mathbf{x}-\mathbf{x'}}{|\mathbf{x}-\mathbf{x'}|} \right)}, \tag{5.9}$$

where $L\left(\mathbf{x'}, (\mathbf{x} - \mathbf{x'})/|\mathbf{x} - \mathbf{x'}|\right)$ is the incident radiance on point $\mathbf{x'}$ from point $\mathbf{x}$, and $L(\mathbf{x'}, \Omega)$ is the reflected radiance from point $\mathbf{x'}$ to solid angle Ω (see Figure 5.1).

5.1.4 Room Acoustic Rendering Equation

The acoustic rendering equation can be used to model room acoustics. The room acoustic rendering equation models the sound energy received at the listener's location by the contribution of the outgoing sound radiance on the surface in each direction. Let $L(x, \Omega, t)$ be the time-

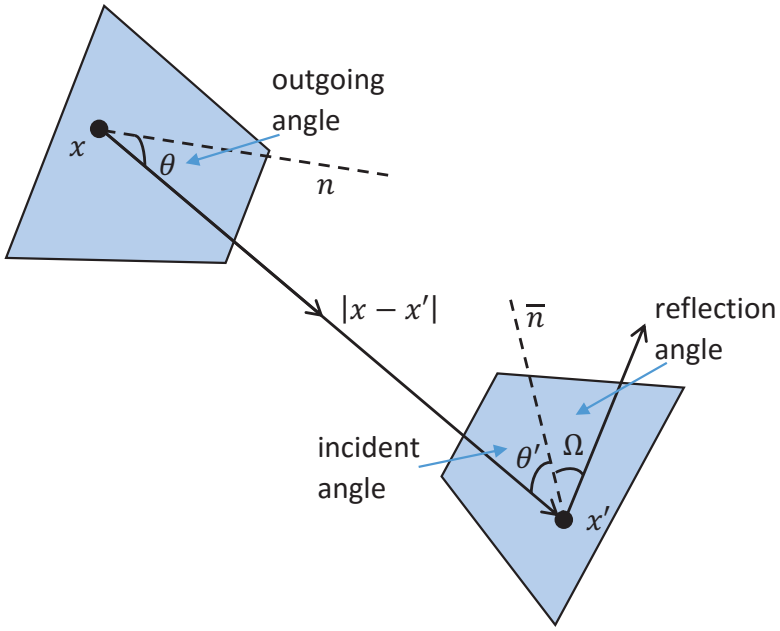


Figure 5.1: Notations of geometric term g.

dependent outgoing sound radiance at point **x** along direction Ω. Then the ARE can be described as

$$L(\mathbf{x}, \Omega, t) = L_0(\mathbf{x}, \Omega, t) + \int_S R(\mathbf{x}, \mathbf{x'}, \Omega, t) L\left(\mathbf{x'}, \frac{\mathbf{x'} - \mathbf{x}}{|\mathbf{x'} - \mathbf{x}|}, t\right) \mathrm{d}\mathbf{x'}, \tag{5.10}$$

where

- $L(\mathbf{x}, \Omega, t)$ is the total acoustic radiance from **x** along direction Ω.
- $L_0(\mathbf{x}, \Omega, t)$ is the initial emitted acoustic radiance from **x** along direction Ω. It only has values when **x** is a sound source.
- $R(x, x', \Omega, t)$ is the reflection kernel which describes how the outgoing energy from point x' influences the outgoing energy at point x in direction Ω, as described in Section 5.1.3.
- $L\left(\mathbf{x'}, (\mathbf{x} - \mathbf{x'})/|\mathbf{x} - \mathbf{x'}|, t\right)$ is the incoming, time-dependent radiance in direction $\mathbf{x'} - \mathbf{x}$ from **x'**.

Compared with Equation 3.3 in Chapter 3, each term in Equation 5.10 has an additional variable which denotes the outgoing direction, such as the radiation direction Ω in $L(\mathbf{x}, \Omega, t)$, $L_0(\mathbf{x}, \Omega, t)$

and $R(\mathbf{x}, \mathbf{x'}, \Omega, t)$, and the $\frac{\mathbf{x'}-\mathbf{x}}{|\mathbf{x'}-\mathbf{x}|}$ in the term of $L\left(\mathbf{x'}, \frac{\mathbf{x'}-\mathbf{x}}{|\mathbf{x'}-\mathbf{x}|}, t\right)$. Again, the introduction of the reflection direction, together with the incident direction, describes the sound energy exchange and propagation between room surfaces in a more general form, and enables the ARE to model the sound propagation phenomenons such as specular reflections and diffractions.

Equation 5.10 can be expressed in Neumann series presented in [Kajiya, 1986]. The outgoing energy resulting from the $(n+1)^{th}$ reflection can thus be expressed as

$$L_{n+1}(\mathbf{x}, \Omega, t) = \int_S R(\mathbf{x}, \mathbf{x'}, \Omega, t) L_n(\mathbf{x'}, \frac{\mathbf{x} - \mathbf{x'}}{|\mathbf{x} - \mathbf{x'}|}, t) \mathrm{d}\mathbf{x'}, \tag{5.11}$$

and the total outgoing energy is

$$L(\mathbf{x}, \Omega, t) = \sum_{n=0}^{\infty} L_n(\mathbf{x}, \Omega, t). \tag{5.12}$$

where $L_n(\mathbf{x}, \Omega, t)$ is the radiance at $\mathbf{x}$ after n reflections.

5.1.5 Discretized Room Acoustic Rendering Equation

In order to solve the Acoustic Rendering Equation by numerical simulation, the room surface S is discretized into N patches. The Bidirectional Reflectance Distribution Function (BRDF), which defines how the incoming sound flux is reflected at a surface, is discretized by dividing the surface-side hemisphere of the reflectance point into solid angles. The hemisphere is divided into m azimuth angles and n elevation angles.

This discretizing scheme results in $N \times m \times n$ radiance transfer elements. The sound energy intensity is approximated as constant over each solid angle. A further assumption is that the distance between points in the patches are approximated by the average distance between patches.

The outgoing energy resulting from the n^{th} reflection can be written in discretized form as

$$\hat{L}_n^{\mathbf{i}} = \sum_{\mathbf{j}=1}^{N \times m \times n} F_{\mathbf{j},\mathbf{i}} \hat{L}_{n-1}^{\mathbf{j}}. \tag{5.13}$$

with $\mathbf{i} = \{k, p, s\}$ denoting the radiance transfer element in patch k, azimuth p and elevation s, and $\mathbf{j} = \{l, q, t\}$ denoting the radiance transfer element in patch l, azimuth q and elevation t. $F_{\mathbf{j},\mathbf{i}}$ is an energy transfer response which models the reflected radiance from element $\mathbf{i}$ by the contribution of the reflected or emitted energy flux from element $\mathbf{j}$. $\hat{L}_{n-1}^{\mathbf{j}}$ is the outgoing energy for element $\mathbf{j}$ in the $n-1^{\text{th}}$ reflection, and $\hat{L}_n^{\mathbf{i}}$ is the outgoing energy for element $\mathbf{i}$ in the n^{th}

reflection.

In Chapter 3, we also use $F_{j,i}$ to denote the form factors, which define only the patch-to-patch interaction. In this chapter, $F_{\mathbf{j},\mathbf{i}}$ is an element-to-element interaction, where the elements **i** and **j** are three-dimensional variable containing the information of the patch, azimuth and elevation.

If the $\hat{L}_n^{\mathbf{i}}$ elements are put into vector form ϕ_n and the $F_{\mathbf{j},\mathbf{i}}$ into matrix F, such as:

$$\phi_n = \begin{pmatrix} \hat{L}_n^1 \\ \vdots \\ \hat{L}_n^N \end{pmatrix}, F = \begin{pmatrix} F_{1,1} & \cdots & F_{1,N} \\ \vdots & \ddots & \vdots \\ F_{N,1} & \cdots & F_{N,N} \end{pmatrix}, F_{\bar{i},\bar{i}} = 0, \tag{5.14}$$

then the $n-1^{\text{th}}$ order reflected energy is propagated into the reflected energy of the n^{th} order as:

$$\phi_n = F\phi_{n-1}. \tag{5.15}$$

The total outgoing energy from the surfaces can be written in discretized form as

$$\phi = \sum_{n=0}^{\infty} \phi_n = \phi_0 + \sum_{n=1}^{\infty} F^n \phi_0. \tag{5.16}$$

where F^n denotes the n^{th} convolution of the corresponding elements (energy responses) in the transfer matrix F if computed in the time domain, or the n^{th} power if computed in the frequency domain.

Utilising the property of geometric series, and assuming that the series in Equation 5.16 converge, the summation can be rewritten as:

$$\phi = \phi_0 + (I - F)^{-1}\phi_0. \tag{5.17}$$

5.1.6 Relation to the Radiance Transfer Method

Compared to the RTM in Chapter 3, the ARE has several major differences which have considerable impact on the modeling performance.

- The RTM only discretizes the room surfaces into patches, while the ARE also discretizes the directional space of each patch into solid angles.
- The RTM model assumes direction-independent diffuse reflections, while ARE model in-

corporates the directional information and is able to model other acoustic propagation phenomenons such as specular reflections and diffractions.

- The reflection kernel of the ARE contains the information of the incoming patch, the receiving patch and the reflection direction, while the reflection kernel of the RTM only needs the radiation density of the incoming and the receiving patches. The bidirectional extension makes the ARE more flexible in modeling various sound interactions.
- For the geometries with similar complexity, the computation and memory resources required by the ARE are much higher than the RTM. By discretizing the directional space of each patch into solid angles, the dimension of the transfer matrix F is much larger than the RTM, where no directional information is considered.

5.2 Numerical Simulation Using Ray-Tracing

[Siltanen et al., 2007] uses a progressive radiance transfer solution. The sound energy is shot from the source to all the surface patches. The patches receiving sound energy radiate to other patches following the acoustic rendering equation theory. The process continues until a termination criteria (time limit, iteration limit, or energy level threshold) is met.

Different from the work in [Siltanen et al., 2007], we first pre-compute the element-to-element responses $F_{\mathbf{i},\mathbf{j}}$ and the high-order element-to-element response matrix $\sum_{n=1}^{N_T} F^n$. The source-to-patch responses and the patch-to-receiver responses are computed during run time and are convolved with the high-order element-to-element responses to form the whole source-to-receiver response. In this way, the source and receiver is decoupled from the whole computational procedure. In addition, we use ray tracing method to compute the acoustic rendering equation.

5.2.1 Pre-computation

The numerical simulation can be divided into an off-line computation and a run-time computation, as shown in Figure 5.2. During the off-line computation, the room geometry is discretized. The surfaces are divided into patches, and the hemisphere of each patch is discretized into solid angles. The transfer matrix F is computed and the high-order element-to-element transfer matrix is obtained.

We use energy-based ray tracing to compute the element-to-element responses $F_{\mathbf{i},\mathbf{j}}$. We assume unit energy is emitted from each element. Rays are emitted uniformly within the solid angle of element i over the patch. Each ray carries a amount of energy $e_R = \frac{1}{N_R}$, where N_R is the number of rays emitted from each solid angle. When a ray encounters a geometric primitive,

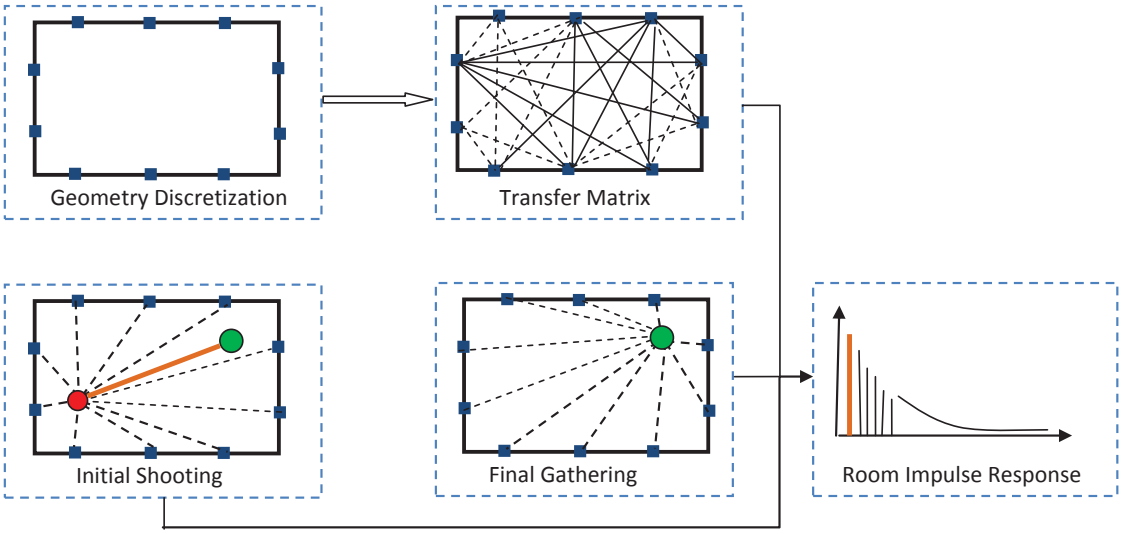


Figure 5.2: One-pass numerical simulation. The red spot denotes the sound source and the green spot denotes the listener. The orange line indicates the direct sound.

it is specularly reflected, according to the pre-defined BRDF. The reflected energy flux in each outgoing direction is calculated to form the one-order element to element response $F_{\mathbf{i,j}}$.

The elements (energy responses) in the transfer matrix F are convolved N_T times to get high order responses. N_T is chosen as a limited number since after several reflections, most energy is absorbed by the reflection materials. The remaining energy is below a desired threshold and thus can be neglected. The high order element-to-element transfer matrix is then formed by $\sum_{n=1}^{N_T} F^n$.

5.2.2 Run-time computation

Run-time computation can be divided into three stages: initial shooting, energy convolution and final gathering.

Initial shooting

The items in the initial shooting matrix contain the received energy and time delays from the source to each discretized solid angle of each patch. At the initial shooting stage, rays are emitted in uniform directions from the sound source. The energy carried by each ray can vary according to the directivity of the sound. These rays are reflected and collected in each reflected angle to form the initial condition ϕ_0.

Algorithm 1: Algorithm of pre-computing the element-to-element response

Data: Discretized room geometry S
Reflectance funtions BRDF ρ of each patch
Result: One-order element-to-element response $F_{\mathbf{i,j}}$
foreach *Patch* k **do**
  foreach *Solid angle* Ω, *(*$\boldsymbol{i} = [k, \Omega]$*)* **do**
    foreach *Ray* R *with energy* e_R **do**
      $x_R \leftarrow$ UniformSamplePoint()
      $d_R \leftarrow$ UniformSampleDirection()
      $y_I \leftarrow$ GetWallIntersection()
      SpecularReflection(d_R, y_I)
      $d_I \leftarrow$ GetReflectionDirection()
      $\mathbf{j} \leftarrow$ GetReflectionIndex(y_I,d_I,S)
      $F_{\mathbf{i,j}} \leftarrow F_{\mathbf{i,j}} + e_R \delta(t - \frac{d_{\mathbf{i,j}}}{c})$
    end
  end
end

Algorithm 2: Initial shooting algorithm

Data: Source location x_S and directivity
Discretized room geometry S
Reflectance funtions BRDF ρ of each patch
Result: Initial shooting matrix $\phi_0(i)$
foreach *Ray* R *with energy* e_R **do**
  $d_R \leftarrow$ UniformSampleDirection()
  $y_I \leftarrow$ GetWallIntersection()
  SpecularReflection(d_R, y_I)
  $d_I \leftarrow$ GetReflectionDirection()
  $i \leftarrow$ GetReflectionIndex(y_I,d_I,S)
  $\phi_0(\mathbf{i}) \leftarrow \phi_0(i) + e_R \delta(t - \frac{d_{S,\mathbf{i}}}{c})$
end

Final gathering

During the final gathering stage, the energy responses from all visible radiance transfer elements are accumulated at the listener's location. As for the initial shooting stage, we use ray tracing to calculate the amount of energy received by the listener from each element. The high-order element-to-element energy responses in the transfer matrix are convolved with the elements in the initial shooting matrix and the final gathering matrix to form the source-to-receiver response. The final gathering scheme will be discussed in detail in Section 5.3.

It is worth noting that in dynamic scene applications where both the sound sources and listeners move, only the initial shooting and final gathering matrices need to be re-calculated. The high-order reflection response matrix is independent from the sound sources and listeners positions. Thus this method has the potential to be applied in dynamic scene applications where both the source and the receiver move.

5.3 Final Gathering Schemes

In [Antani et al., 2012a,b], rays are uniformly sent from the listener and collected at each patch to calculate the energy response at the final gathering stage, using a reciprocity principle. Since these works only use ARE method to model diffuse reflections for the late reverberation, the effect of the directional discretization in the final gathering stage was not discussed. For simulation of early echoes, it is important to accurately model the reflection angle and the energy amplitude for each solid angle. Using such a final gathering method tends to result in inaccurate early echoes. Discretizing the geometry and directions introduces errors in two ways. Firstly, the outgoing energy flux in a certain direction of one patch is considered constant. In fact, it can have considerable variations, especially for the patches close to the source. Secondly, because the sound sources and listeners are free to move in the acoustic space, there exist positions for listeners where the final gathering energy responses are largely under- or over- estimated.

A more natural way is to send rays from the center of each patch and to accumulate them at the listener's position, similar to what they did for modeling diffuse radiosity. The rays are collected by a receiving volume at the listener's position. The receiving volume is adjustable as in the ray tracing method. The received energy responses are convolved with the corresponding elements (energy responses) in the initial shooting matrix and the transfer matrix to get the early echoes as well as the late reverberation. The direct sound is calculated at the initial shooting stage. Together with the direct sound, the whole RIR is generated in a single computation pass, instead of splitting the RIR into ER and LR and modeling them separately. However, this method still

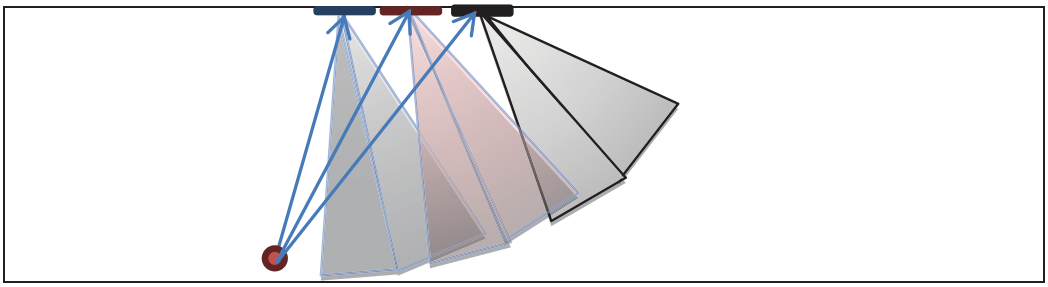


Figure 5.3: Original scheme: final gathering scheme by emitting rays from the centers of patches.

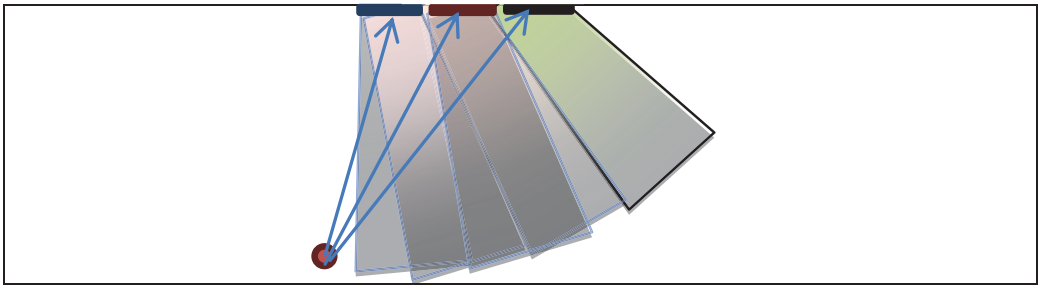


Figure 5.4: Proposed scheme: final gathering scheme by emitting rays uniformly over the patch.

leads to large areas where the listener receives no sound, and some other areas where he receives overlapped sounds from adjacent patches, as one can see on Figure 5.3.

To make the reflected sound more uniformly distributed and to better cover the acoustic space, we propose to emit rays uniformly over each patch along the solid angle, instead of emitting all rays from the center of the patch. For each solid angle of a patch, we uniformly choose K points on that patch. We send N_R rays from each point along uniform directions within the solid angle. Each ray carries energy of $e_R = \frac{1}{K \times N_R}$. An non-uniform receiver directivity can also be taken into account using a varying ray energy distribution. These rays are accumulated at the receiving volume to generate the final gathering responses. The uniform final gather scheme is illustrated in Figure 5.4.

5.4 Experimental Results

In this section, we compare the performance for early reflections simulation of the ARE method using the proposed uniform final gathering scheme (called as ARE-proposed when presenting the experimental results) to the ARE-traditional method.

Algorithm 3: Algorithm of uniform final gathering scheme

Data: Listener's location x_L and directivity
Diameter of the receiving volume ball ε_L
Discretized room geometry S
Result: Final gathering matrix $G(\mathbf{i})$
foreach *Patch* k **do**
  foreach *Solid angle* Ω, *(*$\boldsymbol{i} = [k, \Omega]$*)* **do**
    $x_p \leftarrow$ GetKEmissionPoint()
    foreach *Emission Point* x_p **do**
      foreach *Ray* R *with energy* e_R **do**
        $d_R \leftarrow$ UniformSampleDirection()
        $r_I \leftarrow$ GetNearestIntersectionWithListener()
        if $r_I \leq \varepsilon_L$ **then**
          $G(\mathbf{i}) \leftarrow G(\mathbf{i}) + e_R \delta(t - \frac{d_{i,L}}{c})$
        end
      end
    end
  end
end

5.4.1 Experiment

The test environment is a rectangular room with dimensions $4m \times 3m \times 2m$, as shown in Figure 5.5. The room surface is divided into 104 triangles. For each patch, the hemisphere is divided into 144 directions, 24 in azimuth and 6 in elevation, with 15^o resolution in each dimension. The sound source is fixed at $(1, 2, 1)$, while 40 listeners' positions are uniformly selected on the plane at height $z = 1$. Energy responses are generated using each method and are evaluated by comparison with reference signals obtained by ray tracing. The evaluation of the energy responses at these 40 locations is presented on Figure 5.8 as a 2D map. The energy responses at two locations, one picked at the central of the room and the other near the surface, are given on Figure 5.6 and Figure 5.7 to show the reconstruction details.

In another set of experiments we study the effect of discretization of the room surface and directions. The number of patches varies from 32 to 416, and the solid angle resolution varies from 15^o to 30^o. Their combined patterns are shown in Table 5.1. For each discretization pattern, 40 energy responses are simulated at the locations designed above, and their reconstruction accuracy as well as the average computation time are shown in Table 5.1.

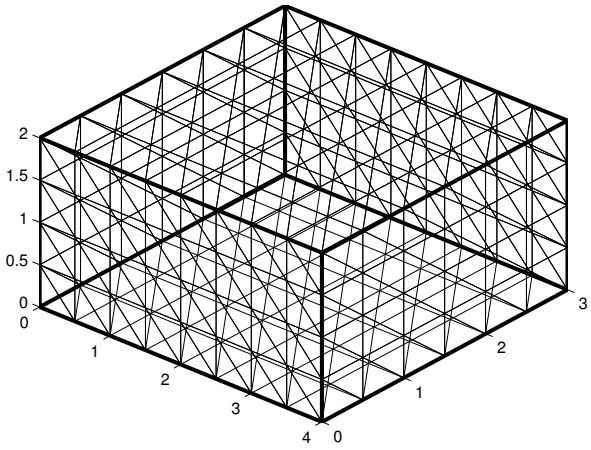


Figure 5.5: Test geometry: a rectangular room with dimensions $4m \times 3m \times 2m$. The surfaces are discretized into 416 triangle patches.

5.4.2 Evaluation Criterion

The ray tracing method is used as the reference method to evaluate the accuracy of both methods. Since the performance of the ARE method in the reconstruction of diffuse reflections and the late reverberation has been studied in [Antani et al., 2012b; Siltanen et al., 2007], we mainly focus on the accuracy in early echoes simulation. We thus consider the simulated energy responses up to the fourth order reflections.

As a first evaluation criterion, we use normalized cross-correlation to assess the distance between the simulated energy impulse responses simulated with the different ARE methods, and the baseline ray tracing method. The formula is given in (Equation 5.18) as

$$\Delta = \frac{f(t) \star g(t)}{||f(t)||_2 ||g(t)||_2} \tag{5.18}$$

where $f(t)$ and $g(t)$ are the considered signals, $\star$ denotes correlation, and $||\cdot||_2$ denotes the l_2 norm.

Since the individual impulses in the reconstructed energy response using ARE methods can have slight time shift due to surface discretization, the normalized cross-correlation is not sufficiently informative to assess the results. Thus, the difference between the accumulated energy is also used to evaluate the performance.

Let $G(t) = \int_0^t g(t)\,\mathrm{d}t$ and $F(t) = \int_0^t f(t)\,\mathrm{d}t$, then the error between the accumulated energy

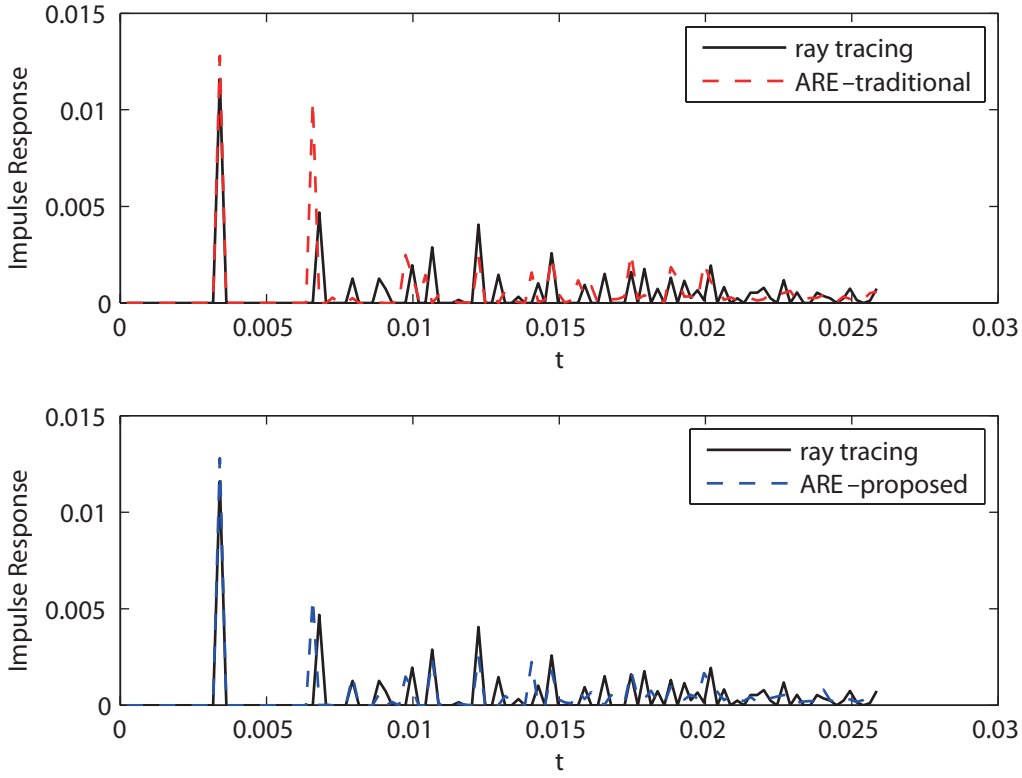


Figure 5.6: Case A: energy response obtained by the ARE methods and the ray tracing method. The sound source is at $(1, 2, 1)$ and the listener is at $(2, 1.5, 1)$.

is

$$\varepsilon = \int_0^{T_M} |G(t) - F(t)| \, \mathrm{d}t \tag{5.19}$$

where $[0, T_M]$ is the time interval considered which in our case roughly corresponds to including the first four orders reflections.

5.4.3 Results

Figure 5.6 and Figure 5.7 display the reconstructed early echoes at listener's location $(2, 1.5, 1)$ and $(1, 0.5, 1)$ (case A and B in Figure 5.8). In Figure 5.6, the proposed method shows a better reconstruction. Note that the first echo is largely over-estimated by the ARE-traditional method, but more accurately modelled by the proposed method. In Figure 5.7 the first echo is totally missed by the ARE-traditional method, but is well preserved using the proposed method. The improvement is due to the averaging effect of the uniform distribution of the emission points

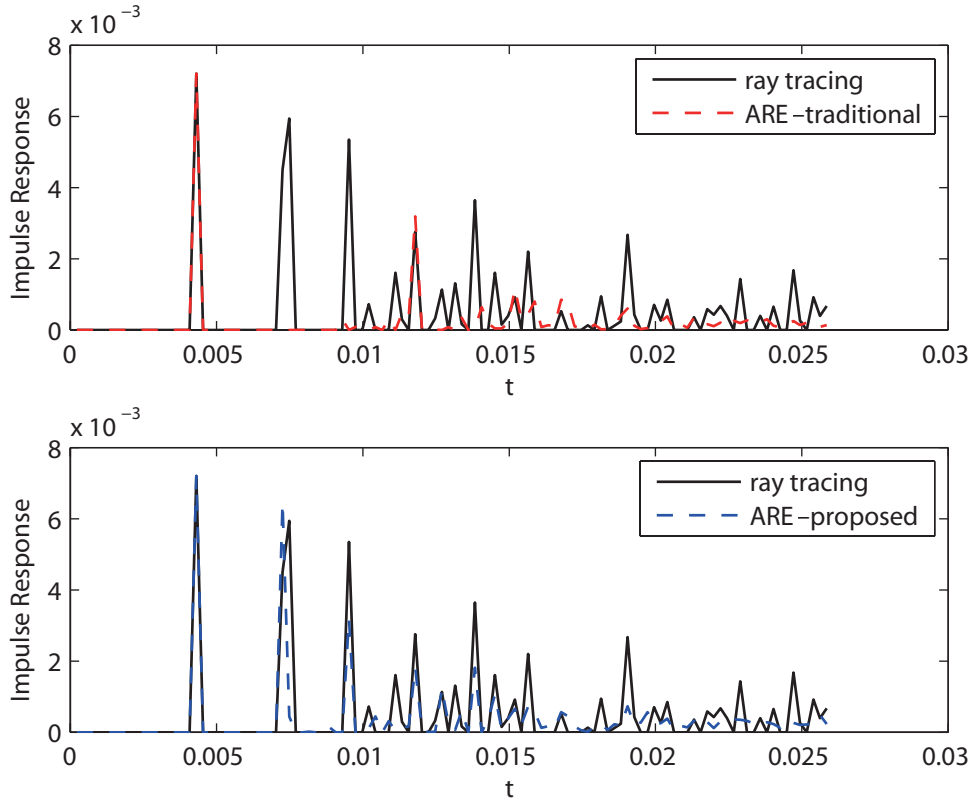


Figure 5.7: Case B: energy response obtained by the ARE methods and the ray tracing method. The sound source is at $(1, 2, 1)$ and the listener is at $(1, 0.5, 1)$.

over each patch. Note that, as shown on Figure 5.7, there still exist a few impulses that are not well modeled by any of the two methods. In this case, the listener is very close to the surface, which leads to strong surface quantization effects.

Figure 5.8 shows how the error of accumulated energy ε varies at different locations in the room. The darker color in the error map indicates smaller accumulated energy error, and thus has better reconstruction accuracy. The error map shows that the reconstructed responses using ARE-proposed are in general more accurate than the traditional method (noted ARE-traditional) at almost every position. The errors at some locations such as at $(1, 0.5, 1)$ and $(2.5, 2, 1)$ are effectively reduced by the proposed method.

Combined with high order reflections, the proposed method is able to effectively model the whole RIR. In order to check that the proposed method did not affect the late reverberation, we computed the reverberation time (T_{30}) of both methods, and found no significant difference. The comparison using cross-correlation as criteria gives similar conclusion, and hence is not

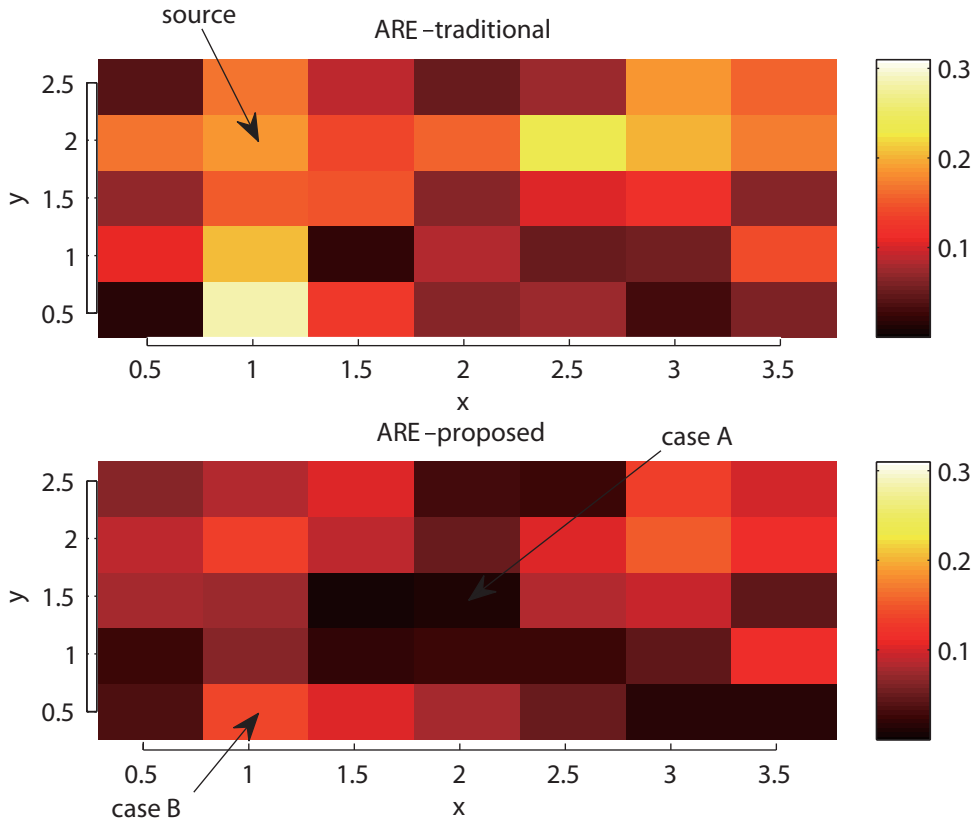


Figure 5.8: Error map of accumulated energy (lower value indicates better result). The sound source is at $(1, 2, 1)$ and the listeners are located uniformly on the plane at height $z = 1$.

Table 5.1: Performance comparison for the traditional and proposed methods under different discretization patterns.

N	m	n	Algorithm	$\bar{\varepsilon}$	$max(\varepsilon)$	$\bar{\Delta}$	$min(\Delta)$	$t\ [min]$
32	12	3	Traditional	0.11	0.37	0.67	0.38	0.56
			Proposed	0.09	0.18	0.69	0.43	0.52
32	24	6	Traditional	0.13	0.67	0.58	0.25	0.77
			Proposed	0.08	0.15	0.65	0.46	0.91
104	12	3	Traditional	0.11	0.26	0.73	0.59	1.99
			Proposed	0.07	0.15	0.79	0.64	1.95
104	24	6	Traditional	0.12	0.28	0.71	0.55	2.53
			Proposed	0.07	0.18	0.79	0.65	2.62
416	24	6	Traditional	0.08	0.22	0.78	0.68	15.1
			Proposed	0.05	0.12	0.81	0.71	15.6

presented here.

Table 5.1 compares the performance of the two methods under different discretization patterns. In this table, N denotes the number of patches, m the number of discretization in azimuth, and n the number of discretization in elevation. We use $\bar{\varepsilon}$ for the average error of accumulated energy, $max(\varepsilon)$ for the maximum error of accumulated energy, $\bar{\Delta}$ for the average correlation, and $min(\Delta)$ for the minimum correlation. t is the computation time in minutes.

It is interesting to observe from Table 5.1 that dividing the hemisphere finer does not necessarily bring improvement to the accuracy of the early echoes. On the contrary, the performance is slightly degraded in some tests, although not in a systematic way. However, this observation indicates that when more computation power and memory is available, first trying to divide the surface into finer patches can be a better option.

Comparing the two methods under the same discretization pattern, it is clear that the proposed ARE technique outperforms the traditional ARE method, at a very mild cost in terms of computational requirements. As seen in the average values in the last two discretization patterns, the proposed method can sometimes achieve a performance similar to the traditional method used with finer discretization. In other words, the proposed method is a competitive alternative to using smaller patches, when high precision is needed with limited computational resources.

5.5 Conclusion

In this chapter we investigated the use of the acoustic rendering equation in modeling early reflections of room impulse responses. We utilised a different energy final gathering method compared to the traditional scheme. Using the new scheme, a single generic method can efficiently model both the early and the late reverberation of RIRs. Although acoustic rendering equation is not guaranteed to reconstruct the RIRs as accurately as ray tracing and image source methods, the fact that they decouple the source and listener positions from the bulk of computation makes them an appealing choice for real-time acoustic rendering.

Chapter 6

Acoustic Rendering Networks

Feedback delay networks (FDN) have been widely used to model late reverberation. In the ARE model in Chapter 5, 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. In this chapter, we present the acoustic rendering networks (ARN), an approach which inherits the computational efficiency of the FDN structure, but models the specular/diffuse reflections by using the directional discretization of the ARE method. This study is an extension of the geometry-based reverberator presented in Chapter 3 [Bai et al., 2015b] and the modeling scheme of early reflections using ARE method presented in Chapter 5 [Bai et al., 2013].

This chapter is organised as follows. In section 6.1 the pros and cons of the feedback delay networks are briefly summarized. The idea of joining the ARE model with FDNs is then proposed to overcome one of the main drawbacks of the later. In section 6.2 we recall the basics of the ARE model and the modification that we made in order to incorporate it with the feedback delay structure. In section 6.3 we describe the new acoustic rendering networks and how the parameters of the networks are related to the geometry of the modeled room. Some simulation results are given in section 6.4.

6.1 Feedback Delay Network for Artificial Reverberation

Feedback delay networks have been a widely used perception-based method to model the late reverberation. It uses the iterative feedback structure to produce dense and exponentially decaying pulses for the late reverberation. The feedback delay networks are used for the following advantages:

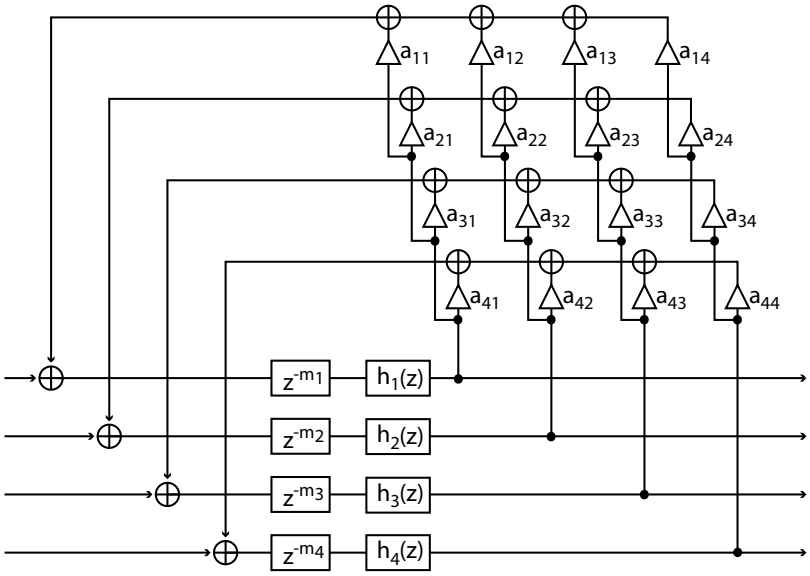


Figure 6.1: Feedback delay structure for producing exponentially decaying dense pulses for the late reverberation.

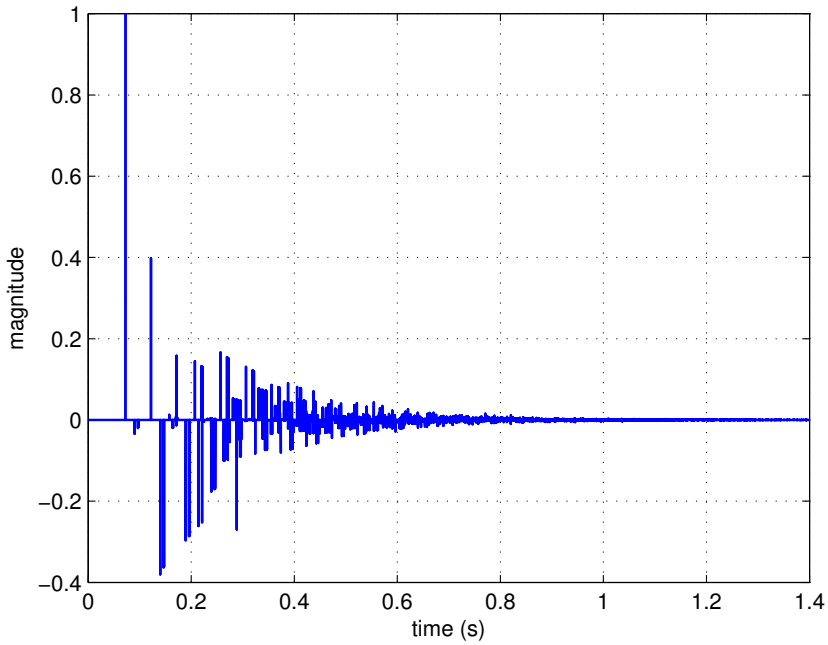


Figure 6.2: Noise-like pulses produced by FDNs with four delay lines. The delay lengths are [0.0497, 0.0674, 0.0738, 0.0750]s.

- The feedback delay structure is computationally efficient. Each sample of the input dry signal requires $N \times N$ multiplications and a few additions, where N is the number of delay lines. The output is the reverberant signal.
- Its parameters are flexible to set in order to model different reverberations. The delay lengths ($[m_1, \cdots, m_i, \cdots]$ in Figure 6.2) and the delay line attenuation filters ($[H_1(z), \cdots, H_i(z), \cdots]$ in Figure 6.2) can be used to adjust the decaying trend and the reverberation time. The frequency-dependent attenuation can also be flexibly parameterized in order to model the different absorptive characteristics of the medium and wall surfaces.
- The feedback matrix enables the mixing of the echoes among the delay lines, which produces more and more sub-echoes in the late part of the reverberation.

On the other hand, it also has its drawbacks and it needs to rely on other techniques to tailor the reverberation for a given room (for example, using the measured impulse responses to estimate the reverberation time, using geometric methods to model early reflections, etc). Its disadvantages include:

- Setting the parameters of a given feedback structure, including the delay lengths, attenuation factors and feedback matrix, can be a rather tedious and empirical task [Rocchesso and Smith, 1997a].
- The feedback delay networks do not model well the early part of the reverberation, leading to hybrid methods where geometric methods can be used to model the early echoes.
- The reverberation modeled by the feedback delay networks tries to meet the perceptual characteristics of the acoustic response in a real environment. Its modeling scheme is not based on the geometry or the materials' acoustical characteristics of the room.

In Chapter 5, the acoustic rendering equation was investigated and extended to model both specular and diffuse reflections. By adopting uniform final gathering scheme and increasing the modeling accuracy of the early echoes, the ARE model offers a single generic approach to model both early reflections and late reverberations of the RIRs. In Chapter 3, using the radiance transfer model, the diffuse reflections between surface patches are linked to the feedback structure of the feedback delay networks. It offers a guideline of how to relate the room geometries to the FDNs and how to estimate the parameters of the FDNs.

It can be noticed that, in the ARE model, the sound reflections between surface patches along the discretized direction are also analogous to the feedback of signals in the FDNs. Thus if we

represent each discretized direction by a delay line, the acoustic modeling using ARE can be replaced by a structure similar to FDNs. In other words, the new reverberator is able to relate the FDNs to the physical room layout. The improvement, compared to the geometric-based late reverberator in Chapter 3, is that it is able to model not only diffuse reflections but also early and specular echoes. It is also the main improvement from the radiance transfer method to the acoustic rendering equation. The new hybrid model also alleviates the shortcomings of the FDNs mentioned above. Before we can easily combine these two methods, some modifications or assumptions need to be made to the acoustic rendering equation.

6.2 Modification of the Acoustic Rendering Equation

As described in Section 5.1, the ARE models the sound energy received at the listener's location as 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 how the outgoing energy from point x' influences the outgoing energy at point x in direction Ω.

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.

6.2.1 Discretization of directions

In [Siltanen et al., 2007], the directions in the hemisphere of a patch, are divided into m uniform azimuth angles and n uniform elevation angles, as shown in Figure 6.4. This results in that a given discretized solid angle from a patch may cover multiple other patches. Conversely, other patches (especially the patches close to each other) have multiple discretized solid angle that over a single receiving patch.

Unlike the directional discretization scheme used in [Siltanen et al., 2007] 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 shown in Figure 6.3. As an example,

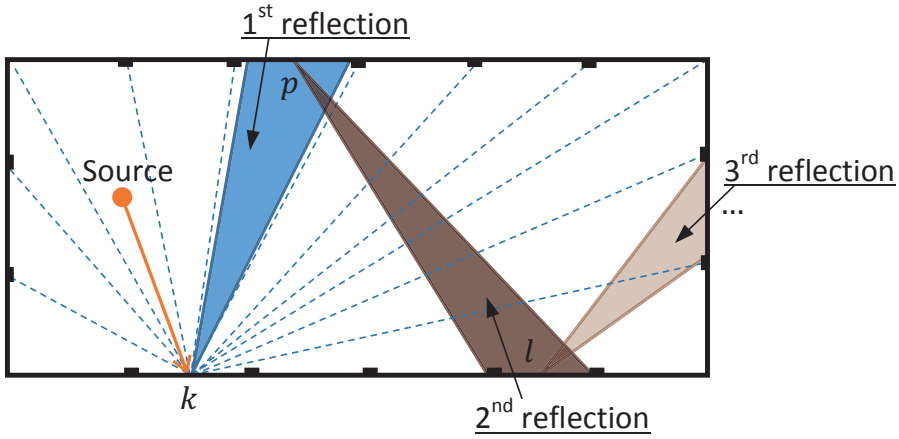


Figure 6.3: Illustration of discretized acoustic rendering equation for specular reflections.

the discretized direction for patch k is illustrated by the dotted lines in Figure 6.3. 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.

Assume that from patch k there are N_k visible patches. Then the total number of discretized direction is:

$$N = \sum_{k=1}^{M} N_k, \tag{6.1}$$

where M is the number of patches.

6.2.2 Simplification of the reflection kernel

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 shown in Figure 6.3. This can be done by choosing patch l as the direction which receives the most reflected energy. This is a rough approximation, which can result in inaccuracy of the amplitude of the modeled early echoes. However, for auralization purpose for real time application, we assume that this is of secondary priority. This approximation results in that the $\mathbf{i}^{th}$ row of the transfer matrix F in Section 5.1, where $\mathbf{i} = \{k, p\}$, has only one specular reflection

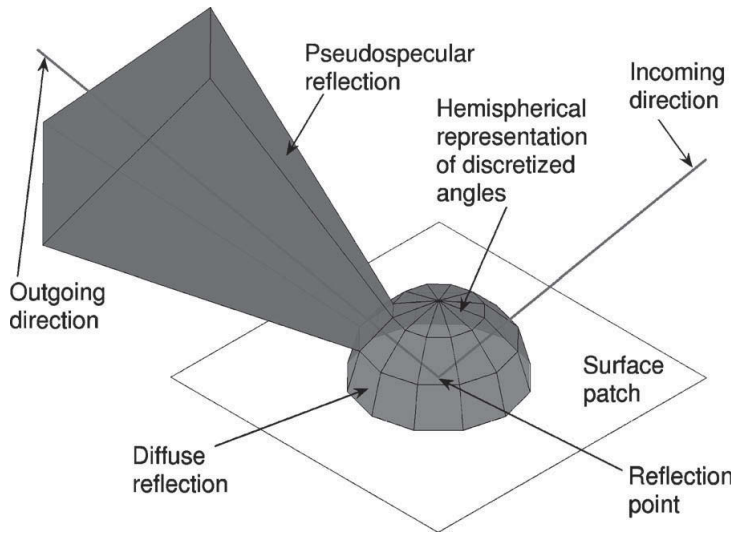


Figure 6.4: Traditional directional discretization scheme [Siltanen et al., 2007]. Hemisphere is divided into solid angles with 12 uniform azimuth angles and 3 elevation angles.

component. The main reason of such a simplification is that, when adopting the ARE to the feedback structure, the feedback matrix can be designed to be stable and unitary. This will be further discussed in Section 6.3.

The discretized acoustic rendering equation and the computational procedures, including initial shooting and final gathering, are similar to the traditional acoustic rendering equation described in Section 5.1 and 5.2.

The computation procedure of the modified acoustic rendering equation can be mapped to the elements in the feedback delay networks. Consider the specular reflection paths shown in Figure 6.3. One reflection path is that the sound emitted from the source is received by patch k. It is then reflected to patch p along the discretized direction defined by the patch pair k and p. It is then, according to the discretized reflection kernel, reflected to patch l to continue the sound propagation. If each patch-to-patch discretized direction is represented by a delay line, where the parameters of the delay line are set according to the acoustic properties of the corresponding discretized direction, the process of sound reflections within the enclosure can be reformulated by a feedback delay structure, as shown in Figure 6.7.

In Figure 6.3, the discretized direction from patch k to patch p is represented by a delay line $L_{k,p}$ in Figure 6.7, with a delay length τ_1 and appropriate feedback coefficient. Similarly, the discretized direction from patch p to patch l is represented by delay line $L_{p,l}$ with delay length τ_2, etc. Then the sound reflections from patch k to patch p and then to patch l are reformulated using the feedback delay lines. The output of the delay line $L_{k,p}$ is fed into the input of the delay

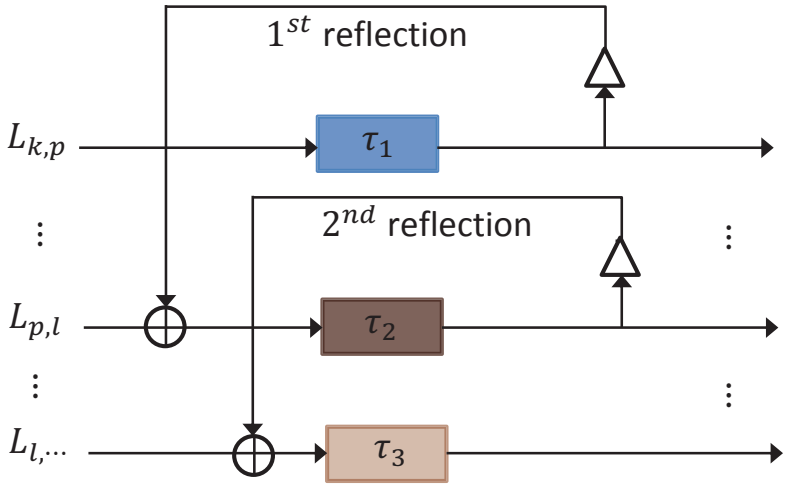


Figure 6.5: Illustration of the feedback of the signals described by the discretized acoustic rendering equation for specular reflections.

line $L_{p,l}$, the output of the delay line $L_{p,l}$ is fed into the input of $L_{l,\dots}$, and so on, as seen on Figure 6.7. If all the patch-to-patch directions are represented by delay lines, the ARE can be formulated by a specific form of feedback delay networks. In the next section, we explain the relationship between the ARE model and the parameters of the delay lines in the feedback delay networks.

6.3 Acoustic Rendering Networks

In this section, we relate the ARE model to the components and parameters of the delay lines in the feedback delay networks.

6.3.1 Design overview

The proposed acoustic rendering networks based reverberator uses one delay line to represent a discretized direction between patches. The acoustic parameters of the patch-to-patch directions are modeled by unidirectional delay units $D_f(z)$ and absorptive filters $H(z)$. The delay lines are inter-connected to each other through a feedback matrix A. This feedback matrix basically defines which delay lines are connected via specular or diffuse reflections and the amount of energy exchanged between them. In addition, source initial-shooting delays $D_S(z)$ and amplitude G_S and microphone final-gathering delays $D_R(z)$ and amplitude G_R are connected to each delay line using unidirectional delay lines.

We name this reverberator an acoustic rendering network because it is a combination of the

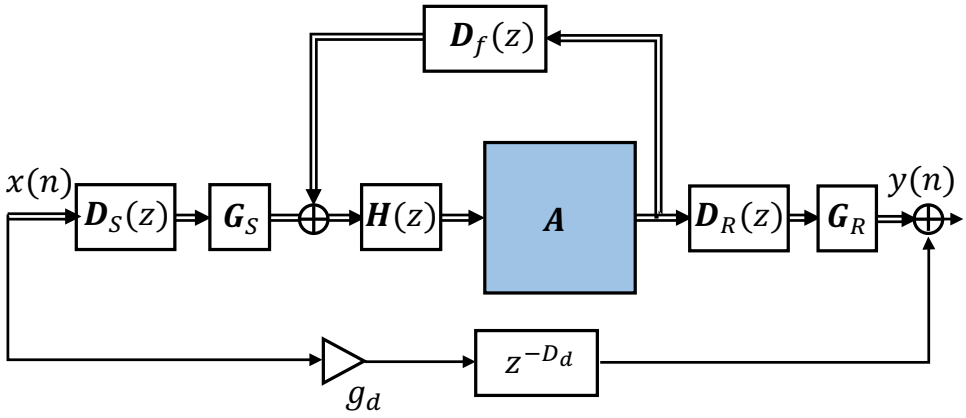


Figure 6.6: System structure of the acoustic rendering network.

acoustic rendering method and the feedback delay network structure. Figure 6.6 shows a conceptual scheme of the ARN reverberator.

6.3.2 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:

$$N = \sum_{k=1}^{M} N_k, \tag{6.2}$$

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:

$$[1 \to \rho_1, \cdots, k \to \rho_k, \cdots, M \to \rho_M], \tag{6.3}$$

where ρ_k denotes the set of visible patches from patch k.

Delay unit

The length of the delay between patch k and p is calculated as:

$$D_{k \to p} = \lfloor F_s \|\mathbf{x}_k - \mathbf{x}_p\| / c \rceil, \tag{6.4}$$

where c is the speed of sound, F_s is the sampling rate, $\mathbf{x}_k$ and $\mathbf{x}_p$ are the positions of the central

point of patch k and p respectively. The feedback delay vector is:

$$D_f(z) = [D_{1\to\rho_1}(z), D_{2\to\rho_2}(z), \cdots, D_{M\to\rho_M}(z)]. \tag{6.5}$$

The feedback attenuation is caused by the wall material absorption from the emitting patch. The air absorption can be easily included as well. The attenuation element can be a fixed attenuation factor or a frequency-dependent reflection filter. Using the same order as we place the delay lines in Equation 6.3, the feedback attenuation vector is formed as:

$$H(z) = [\underbrace{H_1(z), \cdots, H_1(z)}_{N_1}, \cdots, \underbrace{H_M(z), \cdots, H_M(z)}_{N_M}], \tag{6.6}$$

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

Feedback matrix

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 i to patch k is specularly reflected to the direction from patch k to patch j. Then, in the feedback matrix as shown in Figure 6.7, the intersection of the row corresponding to the direction from i to k and column corresponding to the direction from k to j is set to 1, and all the other elements are 0s. The row of the feedback matrix corresponding to the delay line $L_{i,k}$ is:

$$[\underbrace{0, \cdots, 0}_{N_1}, \cdots, \underbrace{0, \cdots, 1, \cdots, 0}_{N_k}, \cdots, \underbrace{0, \cdots, 0}_{N_M}], \tag{6.7}$$

where the "1" is at the corresponding index **s** for the direction pair $\{k, j\}$.

Because of the reflection kernel simplification in Section 6.2.2, 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. The permutation matrix is stable and lossless and does not alter the energy divergence of the system.

If diffuse reflections are considered, the row of the feedback matrix corresponding to the delay line $L_{i,k}$ is modified as:

$$[\underbrace{0, \cdots, 0}_{N_1}, \cdots, \underbrace{\frac{2}{N_k}, \cdots, \frac{2-N_k}{N_k}, \cdots, \frac{2}{N_k}}_{N_k}, \cdots, \underbrace{0, \cdots, 0}_{N_M}], \tag{6.8}$$

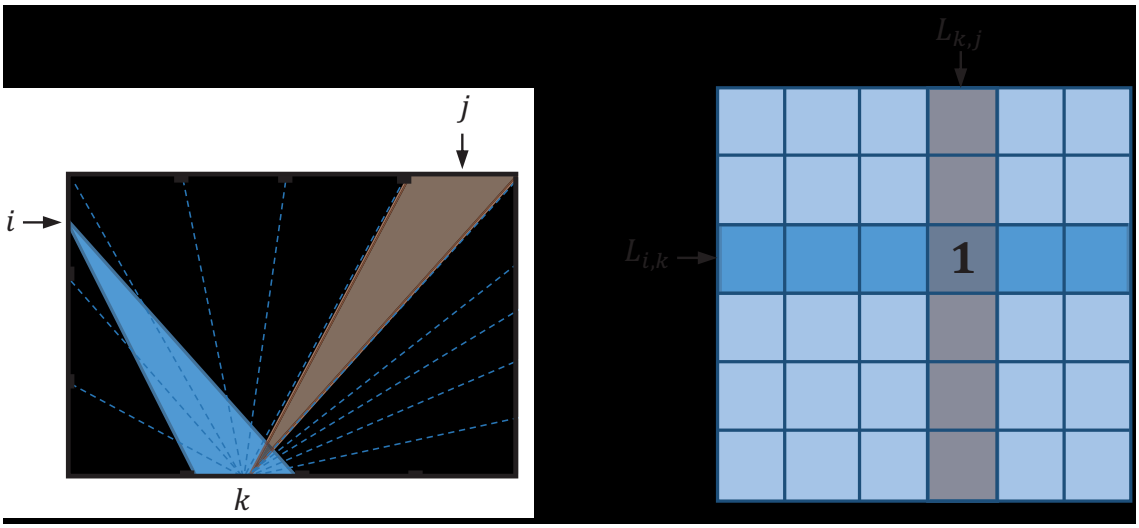


Figure 6.7: Illustration of setting the feedback matrix.

where the specular reflection contains energy:

$$E_{\mathrm{s}} = 1 - \frac{4(N_k - 1)}{{N_k}^2} \tag{6.9}$$

and the diffuse reflections contain energy:

$$E_{\mathrm{d}} = \frac{4(N_k - 1)}{{N_k}^2}. \tag{6.10}$$

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

$$A = \begin{pmatrix} \Lambda_{N_1} & & & 0 \\ & \Lambda_{N_2} & & \\ & & \ddots & \\ 0 & & & \Lambda_{N_M} \end{pmatrix} P, \tag{6.11}$$

where Λ_n is the Householder matrix of order n and P is a permutation matrix. Since the matrix diag
$([\Lambda_{N_1}, \Lambda_{N_2}, \cdots, \Lambda_{N_M}])$ is composed of householder matrices, it is thus 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

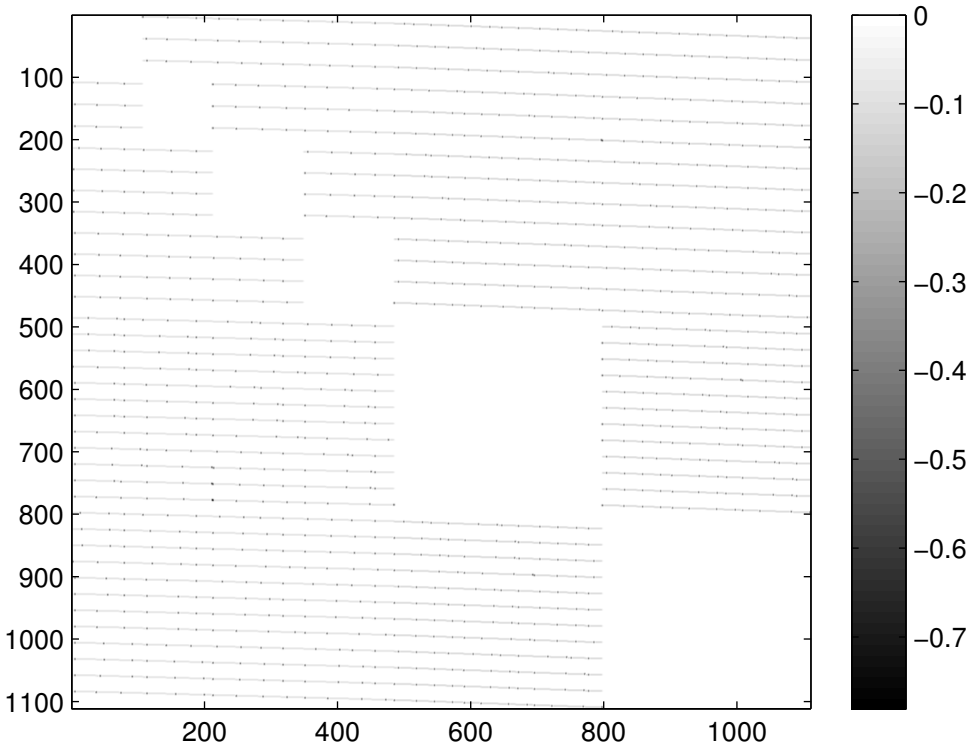


Figure 6.8: Illustration of the sparsity of the feedback matrix of dimension 1106×1106.

makes it possible to link the frequency-dependent absorption property of the material with the attenuation filter of each delay line.

Besides, since the visibility between the discretized directions in the ARE model is limited, the feedback matrix is very sparse. Although it is of dimension $N \times N$, where $N = \sum_{k=1}^{M} N_k$, it has only N specular components. The total number of diffuse components are:

$$N_{\mathrm{d}} = \sum_{k=1}^{M} N_k(N_k - 1). \tag{6.12}$$

However, the diffuse components of each row are of the same amplitude, which requires only one multiplication for the diffuse components instead of the N_k multiplications for each of them.

The sparsity of the feedback matrix for acoustic rendering networks is illustrated in Figure 6.8. The colormap of the figure is scaled and inverted in order to make the color display of the

non-zeros elements more visible. In Figure 6.8, white color denotes zero elements, dark gray dots (with amplitude around -0.7 with reflection coefficient of 0.75) denote specular components and light gray dots (with amplitude below 0.1) denote the diffuse components. (The light gray dots are so densely placed such that they look like light gray lines.) It can be noted that for a feedback matrix of dimension 1106×1106, the number of non-zero elements is relatively small. There is only one dark gray dot in each row. The square-shaped white areas at the diagonal of the feedback matrix mainly represent the in-visibility of the discretized directions to those patches who lie on the same surface.

6.3.3 Initial and final components

Initial components

The input signal first passes through the initial-shooting units and then connects to the delay lines.

The length of the delay between the source at $\mathbf{x}_S$ and the patch k at $\mathbf{x}_k$ is determined by the propagation delay to the central point of the patch:

$$D_{Si} = \lfloor F_s \|\mathbf{x}_S - \mathbf{x}_k\|/c \rceil. \tag{6.13}$$

Thus the initial-shooting delay vector is:

$$D_S(z) = [\underbrace{D_{S1}(z), \cdots, D_{S1}(z)}_{N_1}, \cdots, \underbrace{D_{SM}(z), \cdots, D_{SM}(z)}_{N_M}]. \tag{6.14}$$

The initial-shooting amplitude vector is determined by the amount of energy received by each discretized angle from the source (Sections 5.1 and 5.2). Since the feedback delay networks simulate the reflections in the sense of sound pressure, the received energy is square-rooted to get G_S.

Final components

Like the initial shooting delays, the length of the delay line between the central point of patch k at $\mathbf{x}_k$ and the listener at $\mathbf{x}_R$ is determined by the propagation delay:

$$D_{iR} = \lfloor F_s \|\mathbf{x}_k - \mathbf{x}_R\|/c \rceil. \tag{6.15}$$

The final-gathering delay vector is

$$D_R^T(z) = [D_{\rho_1 R}(z), D_{\rho_2 R}(z), \cdots, D_{\rho_M R}(z)]^T. \tag{6.16}$$

The final-gathering amplitude vector G_R is the square-root of the received energy at the listener from each discretized angle (Sections 5.1 and 5.2).

6.3.4 Relation to previous works

After Jot's reverberator [Jot and Chaigne, 1991], numerous extensions have been reported. However, among them, only a few have been devoted to model early/specular reflections using feedback networks [De Sena et al., 2011; Menzer, 2010]. In Chapter 3 [Bai et al., 2015b] , 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 Chapter 3 relates the geometric acoustic model to 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 Chapter 3 only includes diffuse reflections, where the ARN model incorporates specular reflections.

In [Menzer, 2010], 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.

[De Sena et al., 2011] 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.

6.4 Numerical Evaluation

In this section, RIRs of two rooms are generated by the ARN reverberator, and shown along with their energy decay curves. Early echoes are compared to ISM simulations, taken as reference. An evaluation of the decay rate and the associated reverberation time is then carried out and is compared to that estimated by Sabine formula. Finally an evaluation of the impact of frequency-dependent wall absorption on the reverberation time is given.

6.4.1 Early reflections

Two examples are given to illustrate the simulated early reflections.

Room 1

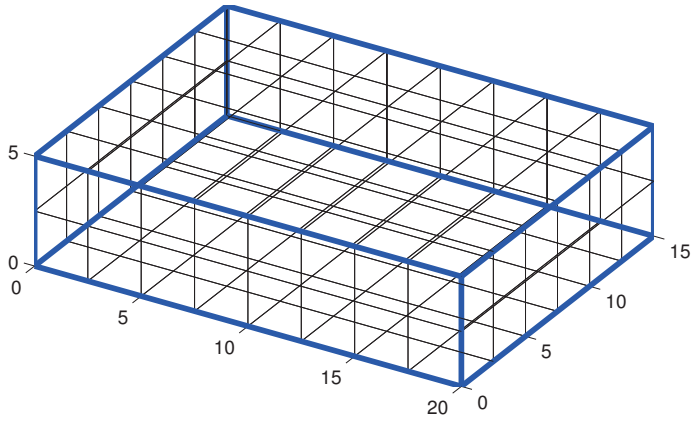


Figure 6.9: Geometry and discretization of Room 1.

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 coefficient $\beta = \sqrt{1-\alpha}$.

The simulation results are shown in Figure 6.10.

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 to sound propagation properties. Carefully observing the zoomed-in early echoes, we can see that most of the early reflections are modelled with

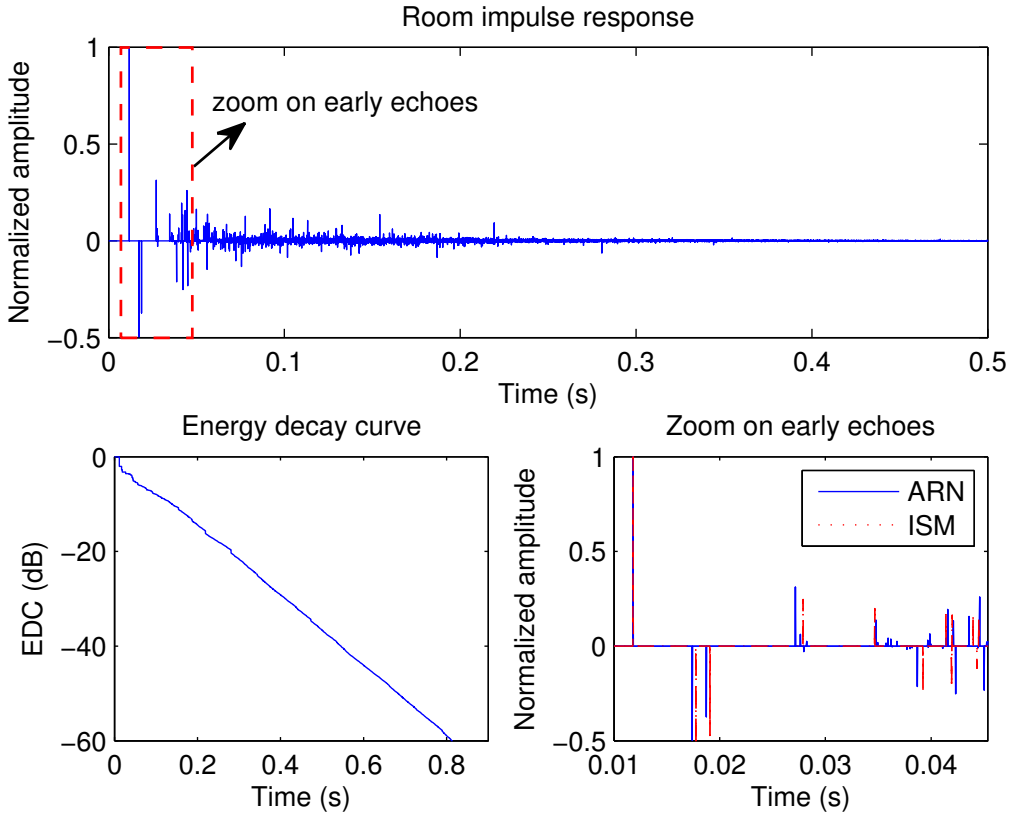


Figure 6.10: Simulation results of Room 1 with absorption $\alpha = 0.3$.

good accuracy. Small time shifts, compared to the reference ISM method, are 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 usually hard to perceive for human ears.

Room 2

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 6.12.

In general, the early echoes are well modeled and the pulses of the late reverberation is dense and exponentially decaying. However, there are a few, at least one first order reflection and one second order reflection, that are largely under-estimated. Compared with the simulation results of Room 1, it is a larger error.

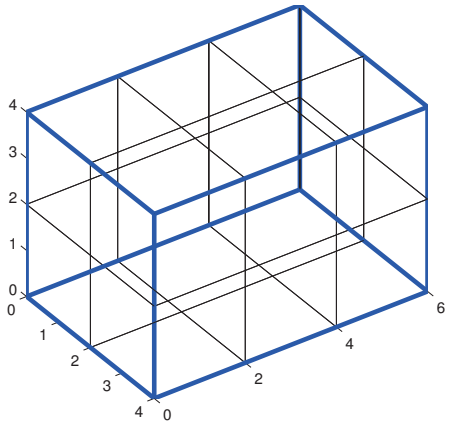


Figure 6.11: Geometry and discretization of Room 2.

The main reason of the errors of the amplitudes of the modeled early reflections is from the coarse discretization. Recall that for room 2 we use a patch size of 2×2m which is comparable to the patch size of 2.5×2.5m for room 1. But room 2 is a much smaller room than room 1. Besides, the simplification of the ARE model also plays a role. Besides using a denser discretization, improvement of the initial shooting and final gather schemes can also increase the modeling accuracy of the early reflections. The beam tracing methods [Antonacci et al., 2008; Funkhouser et al., 1998; Laine et al., 2009; Lauterbach et al., 2007] can also be used to optimise the discretization scheme when the source/receiver is known.

6.4.2 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 [Kuttruff, 2009].

The absorption coefficient varies from $\alpha = 0.2$ to $\alpha = 0.6$.

Table 6.1 shows the predicted T_{60} using the ARN model. The prediction results in general agree with the Sabine predictions. The results are good for Room 1 where the highest prediction error is 7.6%. It is again room 2 which has higher prediction error. This mainly comes from the coarse discretization of the geometry. When we use the distance between the central points of two patches as the delay length of the delay line corresponding to the direction, it is different from the real reflection paths. Possible improvement includes using the average distance between the points on both patches as the delay length.

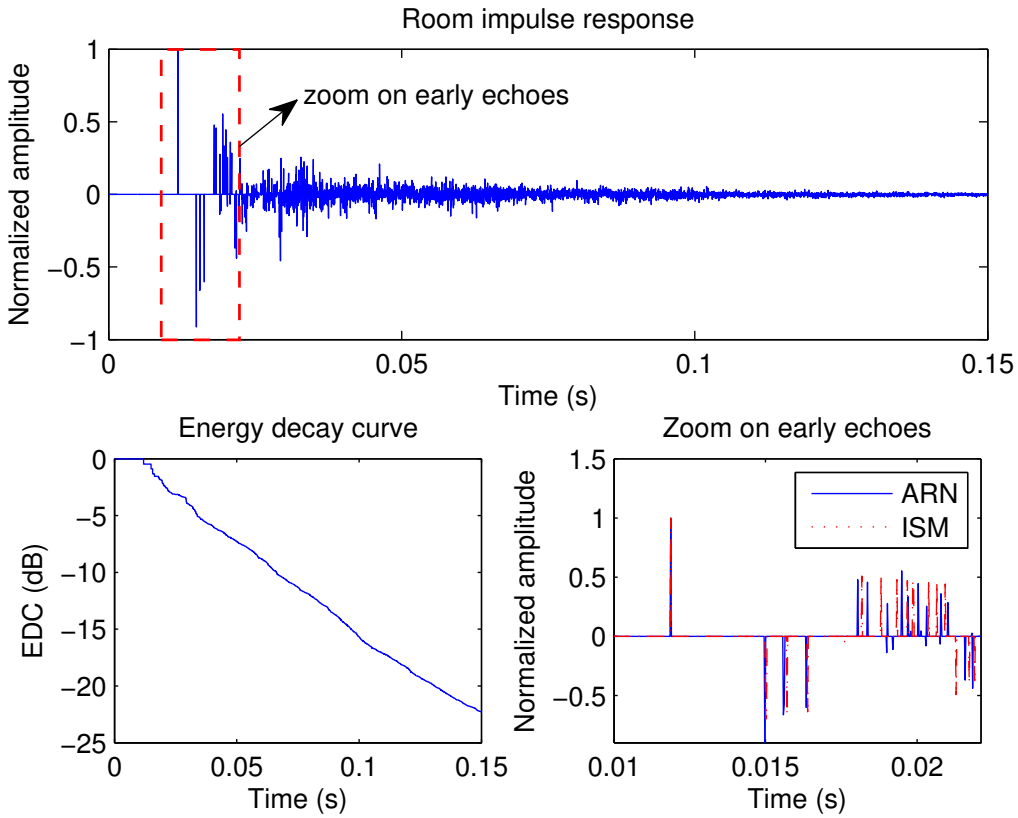


Figure 6.12: Simulation results of Room 2 with absorption $\alpha = 0.25$.

6.4.3 Sound quality evaluation

We compare the sound quality by informal listening tests of the reverberant sound modeled by the ARN with the pure late reverberation modeled by FDNs and the signal modeled by hybrid methods where the early echoes are modeled by the ISM and the late reverberation is modeled by FDNs.

Informal listening tests reveal that the sound quality modeled using ARN has considerable improvement compared to the pure FDNs. The early reflections are noticeable which improves the clarity of the reverberant sound. The sound quality is comparable with that of the hybrid method and can be used as an alternative to the hybrid method.

Besides, due to the high number of delay lines and the "random" permutation between the feedback delay line, the late reverberation is of satisfying sound quality with smooth reverberant sound. If the number of the delay lines is reduced, such a feedback structure with random permutation and sparse feedback matrix can also be used as an alternative artificial reverberator

Table 6.1: 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

even only to model the late reverberation.

6.4.4 Computation performance

The large number of delay lines of the ARN can result in huge computation load. However, the feedback matrix is very sparse. Thus the majority of computation can be saved by ignoring the zero elements.

The simulation uses Matlab on a standard computer (2.8 GHz CPU) to synthesize a one-second RIR. In our experiment using un-optimised Matlab code, to synthesize a one-second RIR, Room 1 simulation consumed 429s, and Room 2 simulation consumed 16.5s. This is not yet ready for realtime simulation, but potential optimization is still possible.

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.

6.5 Conclusion

In this chapter we presented 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 show that the extension to specular reflections does not degrade its performance in late reverberation modeling. Coarser discretization 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. A formal listening test can be of interest to evaluate to which

degree the difference of the delays and amplitudes of the early echoes can affect our perception.

Chapter 7

Applications

Virtual reality immersion is a technique to allow individuals that are geographically apart to interact naturally with each other in a shared virtual environment. In order to enhance the sense of presence of a sound source located somewhere around a listener, a high-quality virtual spatial audio scene environment needs to be created. This chapter discusses the design of virtual auditory environment systems, and links the techniques discussed in the previous chapters to their potential applications. REVERIE[1] is a platform enabling immersive and realistic social networking by utilising virtual reality technologies. The second section of this chapter focuses on the audio module of the REVERIE platform, which captures, transmits and renders 3D audio signals over a headphone to create an interactive virtual reality environment. The techniques proposed in the previous chapters, such as the geometric-based later reverberator of Chapter 3, are applied to generate room effects for virtual environments.

7.1 Virtual Auditory Environments

The creation of virtual environments is associated with the generation of realistic three-dimensional visual and audio scenes. Although auditory cues are often assumed less important than visual cues, a high quality 3D sound can significantly increase the immersiveness and realism of the experience in the virtual world. Besides, events which take place out of the sight of the subject, can only be perceived from the auditory cues [Begault, 1994; Krebber et al., 2000]. For example, the user is able to be aware of the approach of an action before actually seeing it by the distance cues from the sound source. The user can also experience the actions taking place behind him by placing virtual sound sources in his back. Besides, the reverberation also carries some infor-

[1] http://www.reveriefp7.eu/

mation about the environment in which the events occur. This environmental information helps in communication and information retrieval in virtual world.

The virtual auditory environment system includes two main components, the auditory scene synthesis and the sound production. These are known challenges for virtual auditory environment systems, which have motivated various researches on the topics [Blauert et al., 2000; Savioja et al., 1999; Zhang and Xie, 2013].

7.1.1 Interactive auditory scene generation

Virtual auditory systems require a realistic synthesis of the acoustic effects of the environment, which involves the simulation of various acoustic propagation phenomenons. However, there are different considerations between room acoustic simulation and interactive auditory scene generation.

Room acoustic simulation emphasises on the accurate simulation of the physical phenomenons of sound propagation. The many physics-based techniques mentioned in Chapter 2, including the wave-based and ray-based methods, are within this category. They are based on the geometry of the room and on the physical principles of room acoustics. Recent advances include the modeling of diffusions and diffractions alongside with specular reflections. A sophisticated acoustic simulation requires from several seconds to days of computation, depending on the complexity of the room geometry and the simulation algorithm.

On the contrary, auditory scene generation for virtual reality applications requires that the room acoustic effect simulation is done in real-time. Sound sources and listeners are moving in the virtual scene, which further imposes interactivity requirement for the algorithm, since any delay in the scene rendering reduces the impression of immersiveness in the virtual world. To achieve smooth spatial transitions, the location and the head orientation, as well as the room acoustic effects and the rendering components, need to be updated at least 30 times per second [Krebber et al., 2000].

To this aim, artificial reverberators are used for the purpose of real-time synthesis and rendering of the late reverberation effects of the room [Välimäki et al., 2012]. A limited number of early reflections are calculated in real-time as the source and the receiver moves to include the early reflection impression. The image source method and ray tracing are used to calculate the first few reflections. Radiance transfer methods, as a pre-computed approach, are also used to model the diffuse reflections, can be optimised for real-time rendering [Antani et al., 2012a].

Besides real-time requirements, latency is also an important consideration of algorithm design. A critical value is that the latency should not exceed 100 ms [Sandvad, 1996]. Despite

the computation time of room simulation, the convolution of the anechoic input signal with the room impulse response in the time domain alone takes considerable computational resources. Convolution in the frequency domain using block-based FFT is a more efficient implementation. However, the block-based computation introduces extra latency between the input and output. Low latency convolution techniques can be used [Hurchalla, 2010]. Alternatively, artificial reverberators, by modeling the input-output delay and attenuation effects directly from their recursive feedback structures, reduce the input-output latency.

The artificial reverberators have been the focus of the research in this dissertation. In particular we have studied how the parameters of the artificial reverberators can be related with the physical principles of acoustics and the layout of the room geometry. As being computational efficient and with low latency, the proposed approaches can be applied in acoustic effects rendering in virtual reality applications. The geometric-based late reverberator proposed in Chapter 3 can be used to estimate the parameters of the feedback delay networks which can synthesize late reverberations consistent with the virtual room geometry. The acoustic rendering networks proposed in Chapter 6, although not ready to be immediately implemented in virtual reality applications due to the heavy computational resources required by the large number of delay lines, still has the potential in further development since it has the capacity of modeling the early specular reflections, which most artificial reverberators cannot do.

7.1.2 Binaural auralization

Sound auralization addresses the problem of reproducing the ear signals through loudspeakers or headphones in order to create the desired hearing sensation.

Binaural auralization utilises headphones to create the spatial impression of the sound scene. The spatialization is achieved by exploring some physical parameters called "cues" to create a spatial effect using only two channels. There mainly exist three different so-called "localization cues": interaural level difference (ILD), interaural time difference (ITD) and direction-dependent filtering (DDF) [Blauert, 1997; Yost, 1994]. ITD is the difference of arrival time of a sound between the left and right ears, which plays an important role in identifying the direction of low-frequency sound sources. ILD represents the difference in intensity or loudness level between the ears and is important to localize high-frequency sounds greater than 1000 Hz. As ITD and ILD alone cannot distinguish between sound sources at an azimuth of 30° or of 150°, DDF is introduced to solve this issue. DDF is a selective filtering effect caused by the shape and the position of the pinna of the ear and works in a monaural way.

Binaural spatialization is achieved by the convolution with the head-related impulse response (HRIR) whose Fourier transform is called the head-related transfer function (HRTF). HRIR con-

tains the above three important "cues" and is used to model the filtering effect of the outer, head and torso of the subject.

Binaural auralization requires the least installation and parameterization process compared to other auralization systems, such as ambisonics and wavefield synthesis, and thus is suitable for portable devices. Besides, there is no crosstalk between the two channels of the binaural signals. However, due to the difference of the geometrical dimension of the acoustically relevant elements such as torso, head and ear pinna, the HRTFs differs from person to person. Using an averaged HRTF set results in inaccuracies in auralization, such as front-back confusion. To obtain the best localization effect, one would need to use personalized HRTFs.

A two-loudspeaker arrangement [Cooper and Bauck, 1989; Møller, 1989] usually uses crosstalk canceling techniques to reproduce the 3D sound field. In ideal conditions (anechoic room, correct positioning of the receiver), it is possible to achieve the same auditory results as using headphones. However, these ideal conditions are difficult to meet, and the reverberation effect of the listening room is difficult to eliminate. The two-speaker arrangement reproduces the sound with good front localization, but has difficulty to model the sound coming from the back.

7.1.3 Applications

Interactive virtual auditory systems are linked to many applications, ranging from entertainment such as video games and films, to distant education and airplane training. Here we list a few examples.

- *Communication.* In video conferencing systems, 3D auralization can enhance sound quality and speech intelligibility when several persons speak at the same time but at different locations (the so-called cocktail party effect). In speech communication, adding the environment acoustic effects can increase the situational awareness and can help deliver environmental information. The European project REVERIE focuses on the development of novel networking techniques for immersive communication in virtual environments. High quality 3D audiovisual virtual scenes are produced to increase the experience of the immersiveness [Mauro et al., 2013; Pasin et al.].

- *Medical applications.* VETIR is a European project focusing on the development of virtual reality system for the rehabilitation of patients. By utilising virtual acoustic techniques, [Sánchez and Lumbreras, 1999] constructed an interactive software used for learning, cognition, and entertainment purposes for blind people.

- *Perception and interaction.* The ESPRIT project SCATIS (ESPRIT 6358) emphasizes on the

auditory and multi-modal perception and interactions in a virtual environment.

- *Sound design.* The acoustical measurements from a specific environment are used in sound production in order to create the environmental impression. Off-line simulation can be used to create totally virtual acoustic effects in music and film sound design. While for interactive applications such as computer games, online simulation is required [Cowan and Kapralos, 2008; Raghuvanshi et al., 2010].

- *Military and aeronaut training [Xie, 2013].* It includes the virtual acoustic systems designed by NASA [Wenzel, 1991] and the cockpit aiding system by [Calhoun et al., 1987]. Besides, the military standard of the U.S.A., MIL-STD-1472F, also points out that virtual auditory displays should be preferred to provide immersion under some conditions.

7.2 3D Audio Rendering in REVERIE

7.2.1 The REVERIE research project

The project REVERIE (REal and Virtual Engagement in Realistic Immersive Environments), supported by the European Union FP7 program, aims at providing tools for building a mixed reality space. In the mixed reality space, real and virtual worlds engage and seamlessly interact in real-time, generating compelling and highly realistic immersive environments. A particular focus of REVERIE is to introduce a paradigm shift in the content of communication in social networks. The project involves thirteen research institutions and industrial players from Europe, including Télécom ParisTech, Fraunhofer HHI, Disney Research Zurich, STMicroelectronics, CTVC, Blitz Games Studios, Queen Mary College London, Alcatel Lucent Bell, TP Vision, Centrum voor Wiskunde en Informatica Netherlands, Microsoft Innovation Center Torino and WOOX Innovations Belgium.

According to the business plan analysis of REVERIE, the project is focusing on online social networking, providing a collaborative environment bringing together realistic inter-personal communication and interaction. REVERIE solution encompasses 3D and immersive audio-visual display technologies for enhancing user experience, introducing also autonomy into avatar behaviour, capturing, representing and recreating user emotional state as well as physical activity. In parallel, REVERIE introduces a new paradigm for immersive communications in social networks. It also considers sociological considerations including ethical aspects.

The cutting-edge technologies are integrated, including 3D data acquisition and processing, sound processing, autonomous avatars, networking, real-time rendering, and physical interaction and emotional engagement in virtual worlds.

Figure 7.1: Example of participants in a collaborative educational scenario (in European Parliament).

REVERIE offers two use cases which are closely related to education. The first use case, originally defined as "immersive social networks", provides strong integration of many people with emphasis on social networking, and refers to the use of the system by students and teachers to serve educational services. Two scenarios are considered: Scenario 1 – Tour and debate in 3D virtual European Parliament and Scenario 2 – Sharing experience in realistic virtual environments.

The second use case has been originally defined as "online master classes" and later transformed to "natural interaction in immersive environments". It focuses on high quality immersion (aiming at photo-realism and real-time 3D reconstruction), initially for a small set of users, based on advanced 3D capturing, processing, reconstruction, streaming and visualisation techniques. It includes two sub use cases: case 1 – REVERIE hangout [Pasin et al.] (i.e. "3D Skype") and case 2 – role playing games (e.g. table drop, pop quiz, Simon Says).

The audio module of the REVERIE platform utilises the research outcomes from this dissertation, notably the geometric-based reverberator of Chapter 3. The rest of this section gives a brief introduction of the audio module of the REVERIE platform.

7.2.2 3D audio rendering

The audio module of the REVERIE platform supports multi-user communications in virtual world. The multi-users physically apart from each other are connected by internet and are re-located in the same virtual room. Figure 7.2 shows the two-users communication case. The audio signals captured by local microphones are sent to the far-end for room effects enhancement and spatial auralization.

Figure 7.2: REVERIE platform for multi-user communication.

Figure 7.3 shows the 3D audio module of the REVERIE platform. A monophonic audio signal is first captured with microphones embedded in a headset of the local user. The local user's position and head rotation in the virtual world are tracked in real-time. Via the IP network, the audio signal spoken by the local user, together with its position and orientation in the virtual world coordinate, is transmitted from the local user to the far-end listener. At the listener side, the world coordinate of the virtual speaker is transformed to local coordinate from the viewpoint of the listener. The coordinate conversion is done with the aid of the "Scene information database" where the geometrical information of the scenes is stored. The originally received audio signal then passes the audio subsystem to add room reverberation and spatialization effects. The room reverberation effects, including the early reflections and the late reverberation are added to the original signal. Using the converted local coordinate, the reverberant signals are ready for binaural spatialization, and 3D binaural audio signals are generated.

The workflow diagram of 3D audio rendering is given in Figure 7.4. The algorithm implemented by the audio module includes the following steps:

1. obtain the positions and orientations of the local user and the remote user;
2. calculate the distance between the local user and remote user and then use it for sound volume control;
3. transform the Cartesian coordinate system to the interaural polar coordinate system used by the CIPIC HRTF database [Algazi et al., 2001] for the local user;
4. calculate the elevation and azimuth values between the local speaker and the far-end listener and find an index related to the HRTF database
5. pick up the corresponding set of HRTF coefficients from the database to obtain 3D binaurally rendered signals.

3D audio rendering also needs to include the dynamic synthesis of sound propagation. Except the distance and spatial effects between the sound sources and the listeners, sound reflections also

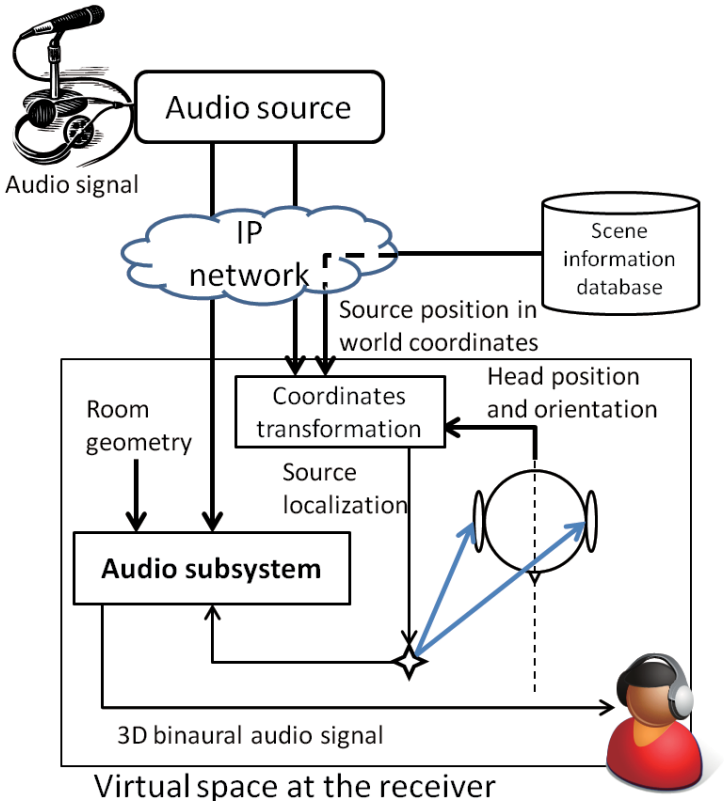


Figure 7.3: Overall REVERIE system architecture for virtual acoustic scene composition and rendering.

need to be included due to the geometry of the simulated room and the presence of obstacles. Due to the computational efficiency and low latency, the statistical method, geometric-based feedback delay network (FDN-RTM) proposed in Figure 3.10 in Chapter 3, is chosen to model the late diffuse reverberation. The geometry of the room is simplified and sound exchange behaviours are analysed by the use of the radiance transfer method. The parameters of the FDNs are estimated with an offline algorithm (Section 3.3) by studying the energy interaction of the radiance transfer method. The acoustic characteristics of the construction materials of the room are integrated to design the low-pass filters which model the frequency-dependent sound absorptions. The simplified geometry of the room is also used to calculate the early specular reflections by the aid of the image source method.

The early reflections and the reverberations together create a telepresence effects in the virtual world. By using the HRTF sets, the receiver is able to localize the audio source in the virtual environment. An interpolation mechanism is proposed to render the moving sources in the virtual environment. When the speakers or the listeners move or change their orientations,

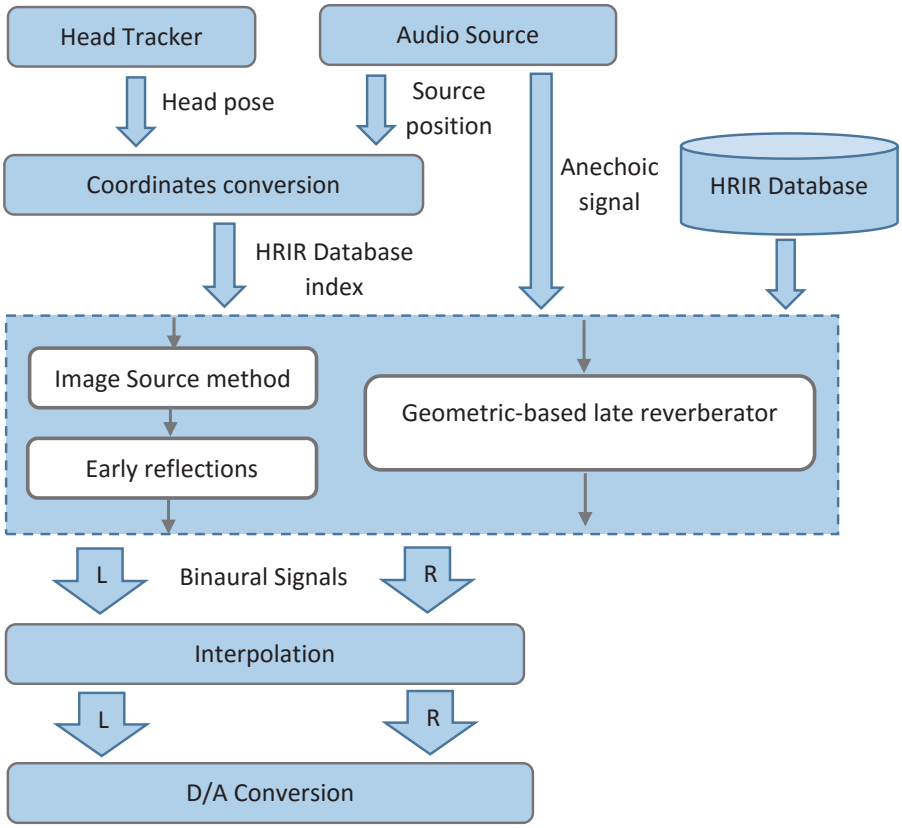


Figure 7.4: The workflow diagram of 3D audio rendering.

the amplitudes and the directions of the direct sound and early reflections are re-calculated and rendered in real-time to create the impression of a dynamic scene.

The current rendering system utilises the geometric-based late reverberator of Chapter 3 to model the diffuse reflections. By using fewer patches and optimizing the computation process, the acoustic rendering networks in Chapter 6 can also be applied to model the specular reflections and the early echoes.

Chapter 8

Conclusions and future work

8.1 Conclusions

The main research work of this dissertation are on the exploration of novel and hybrid methods to model room acoustical reverberations. Special interest has been put on the simulation and rendering methods which support real-time simulation and dynamic rendering in interactive scenarios. Two categories of methods became the focus of the dissertation, the radiance transfer method and the artificial reverberators. The radiance transfer method, together with its recently extended acoustic rendering equation, decouples the source and receiver's location and directivity from the complex sound propagation simulations, and uses the pre-computed energy exchange matrix between room surfaces to generate the source-to-receiver impulse response. They are thus advantageous for dynamic acoustic scene rendering. Artificial reverberators, being widely used in the reverberation industry, are computationally efficient and have low input-output latency. In this dissertation, these two categories of methods have been specifically studied. The links between the parameters of the artificial reverberators and the physical representations of the radiance transfer models are exploited. Parameter estimation schemes, novel artificial reverberation structures and hybrid structures for the acoustical reverberation simulation and rendering have been proposed and investigated in this thesis.

Chapters 1 and 2 introduced the basic knowledge on room acoustic and the state of the art techniques for room acoustic modeling. The pros and cons of these techniques with respect to the evaluation criteria for real-time rendering have been discussed. They motivated the hybridation of these methods in the following chapters.

Chapter 3 studied the radiance transfer methods in their performance of modeling diffuse reflections and presented a new method which implements a simplified radiance transfer method

using the feedback delay network structure. The hybrid method, from one perspective, inherits the accuracy of the physical model of the acoustic radiance transfer method, at a much lower computational load. From another perspective, it relates the parameters of the feedback delay networks with the acoustic properties of the geometrical layout of the room, and is thus a geometric-based artificial reverberator. The sound quality of the geometric-based reverberator approaches that of the radiance transfer method, with the increase of delay lines. This method can be used to simulate room acoustic characteristics, and synthesize virtual reverberation in real time, taking into account the physical characteristics of the simulated room. Combined with the image source method for generating the early echoes, this geometric-based reverberator was applied in a virtual reality application to synthesize the dynamic acoustic reverberation in real time.

After the recent generalization of the acoustic rendering equation to model specular reflections, in Chapter 5 we investigated the modeling accuracy of the early echoes using the acoustic rendering techniques. Two final gathering schemes were studied and compared. The results show that the uniform final gathering scheme reconstructs more accurately the early echoes and can be used to simulate early echoes for interactive virtual acoustic environment applications where interactivity and simulation rate are more important than the modeling accuracy. We further modified the reflection kernel of the acoustic rendering equation and incorporated it with the feedback delay networks in Chapter 6. We presented a novel recursive reverberation structure, which combined the modified acoustic rendering equation with the feedback delay networks. The new reverberator is capable of modeling both early/specular reflections and late/diffuse reverberations. Although discretization errors exist, the modeled early echoes have considerable accuracy for virtual acoustic applications.

In Chapter 7 we discussed the design criteria of audio modules for the virtual acoustic environment applications. The applicability of the geometric-based reverberator proposed in Chapter 3 is justified.

8.2 Contributions

The main contribution of this dissertation has been on the exploration of the geometric-based artificial reverberator. In most of the previous works on room acoustic modeling, geometrical methods and artificial reverberators are often investigated separately, or are used in hybrid way where each method is used to simulate a specific time-frequency region of the impulse response (see Figure 2.6). Thus very few emphasis has been put on the physical interpretation of the structures and parameters of the artificial reverberators, such as the feedback matrix, delays

units and attenuation filters. In this dissertation, we have been focusing on the study of the links between the physical models and the artificial reverberators. The physical meanings of the various parameters in the artificial reverberators are studied. Geometrical acoustic models, including the radiance transfer method and the acoustic rendering equation, are used to link both categories of methods, and have been incorporated in the recursive feedback structures for estimating their parameters. More specifically, the radiance transfer method has been used to estimate the parameters of the artificial reverberators to model the diffuse reflections, and the acoustic rendering equation further extends the recursive feedback structures and enables them to model both specular and diffuse reflections.

The contribution of this dissertation includes the following techniques:

- A parameter estimation scheme for the feedback delay networks and a geometric-based late reverberator (Chapter 3).
- A practical implementation of variable delays for artificial reverberators and a novel recursive feedback delay structure to improve sound quality of the traditional feedback delay networks (Chapter 4).
- An uniform final gathering scheme for the acoustic rendering equation (Chapter 5).
- An novel acoustic rendering network which generates both early/specular echoes and late/diffuse reflections (Chapter 6).
- A virtual acoustic rendering system for interactive social networking application (Chapter 7).

8.3 Future work

Based on the challenges of the topics and the accomplishment done in this dissertation, further research and development can be carried out in the future along several directions.

Generally speaking, for the techniques investigated in this dissertation, the modeling techniques of diffraction effects in the proposed methods, especially the geometric-based reverberator and the acoustic rendering networks, are not well discussed. In dynamic acoustic scenes with complex geometry, modeling the diffraction phenomenon is important, especially for the early echoes. Since the delay lines of the acoustic rendering networks is closely tied with the surface patches of the geometry, the reflection kernel and corresponding feedback matrix can be extended to support a model of diffraction.

For the geometric-based late reverberator proposed in Chapter 3, the existing grouping scheme is based on the uniform energy principle. Further extensions can investigate other grouping schemes, in order to obtain a more flexible group design, and possibly to introduce spatial information for each group. In some extreme environments, the slightly metallic sound and modal frequencies are the desired sound effects. Thus there is a need to have a generalized paradigm of choosing the delay lengths in order to model the sound effects of various application scenarios.

To increase the modeling accuracy of the early echoes using the acoustic rendering equation in Chapter 5, the idea of pre-computing the subdivision and the construction of the beam tree of beam tracing method can be used to optimise the discretization of geometry. The limitation will be that such discretization is only optimal to a specific source location. By using the reciprocity property, this can be applied in scenarios where at least one of the two communication parties is static and the other is dynamic, such as a musical concert and a public speech. The optimization of discretization is also applicable to the acoustic rendering networks in Chapter 6.

In the virtual acoustic rendering system for the social networking application in Chapter 7, only the geometric-based reverberator is currently integrated to model the diffuse reflections and late reverberation. Specular reflections and early echoes can be included using the acoustic rendering networks of Chapter 6. Computational optimization need be done to the acoustic rendering networks, for instance using fewer patches and utilising the sparsity of the feedback matrix, in order to make it applicable for real-time rendering.

Bibliography

T. Ajdler and M. Vetterli. Acoustic based rendering by interpolation of the plenacoustic function. In *SPIE/IS&T; Visual Communications and Image Processing Conference*, volume 5150, pages 1337–1346, 2003a.

T. Ajdler and M. Vetterli. The plenacoustic function, sampling and reconstruction. In *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on*, volume 5, pages V–616. IEEE, 2003b.

T. Ajdler, L. Sbaiz, and M. Vetterli. The plenacoustic function and its sampling. *IEEE Transactions on Signal Processing*, 54(10):3790–3804, 2006.

V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano. The cipic HRTF database. In *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*, pages 99–102. IEEE, 2001.

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

L. Antani, A. Chandak, L. Savioja, and D. Manocha. Interactive sound propagation using compact acoustic transfer operators. *ACM Transactions on Graphics*, 31(1):7, 2012a.

L. Antani, A. Chandak, M. Taylor, and D. Manocha. Direct-to-indirect acoustic radiance transfer. *IEEE Transactions on Visualization and Computer Graphics*, 18(2):261–269, 2012b.

F. Antonacci, M. Foco, A. Sarti, and S. Tubaro. Fast tracing of acoustic beams and paths through visibility lookup. *Audio, Speech, and Language Processing, IEEE Transactions on*, 16(4):812–824, 2008.

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)*, pages 1–4, New Paltz, NY, October 2013.

H. Bai, G. Richard, and L. Daudet. Geometric-based reverberator using acoustic rendering network. In *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 1–4, New Paltz, NY, October 2015a.

H. Bai, G. Richard, and L. Daudet. Late reverberation synthesis: From radiance transfer to feedback delay networks. *Audio, Speech, and Language Processing, IEEE/ACM Transactions on*, 23 (12):2260–2271, 2015b.

Bang&Olufsen. *Music for Archimedes*. Audio CD, 1992.

D. R. Begault. *3-D sound for virtual reality and multimedia*. New York: Academic, 1994.

M. Bertram, E. Deines, J. Mohring, J. Jegorovs, and H. Hagen. Phonon tracing for auralization and visualization of sound. In *Visualization, 2005. VIS 05. IEEE*, pages 151–158. IEEE, 2005.

L. Bianchi, D. Markovic, F. Antonacci, A. Sarti, and S. Tubaro. Deconvolution of plenacoustic images. In *Proceedings of the IEEE Workshop on Application of Signal Processing to Audio and Acoustics (WASPAA)*, pages 1–4, New Paltz, NY, October 2013.

J. Blauert. *Spatial hearing: the psychophysics of human sound localization*. MIT press, 1997.

J. Blauert, H. Lehnert, J. Sahrhage, and H. Strauss. An interactive virtual-environment generator for psychoacoustic research. i: Architecture and implementation. *Acta Acustica united with Acustica*, 86(1):94–102, 2000.

J. Borish. Extension of the image model to arbitrary polyhedra. *The Journal of the Acoustical Society of America*, 75(6):1827–1836, 1984.

J. Borish and J. B. Angell. An efficient algorithm for measuring the impulse response using pseudorandom noise. *Journal of the Audio Engineering Society*, 31(7/8):478–488, 1983.

D. Botteldooren. Finite-difference time-domain simulation of low-frequency room acoustic problems. *The Journal of the Acoustical Society of America*, 98(6):3302–3308, 1995.

C. Cadoz, A. Luciani, and J. L. Florens. Cordis-anima: a modeling and simulation system for sound and image synthesis: the general formalism. *Computer music journal*, pages 19–29, 1993.

G. L. Calhoun, G. Valencia, and T. A. Furness. Three-dimensional auditory cue simulation for crew station design/evaluation. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 31, pages 1398–1402. SAGE Publications, 1987.

S. Carlile. Virtual auditory space: Generation and applications, (Landes, Austin, TX). 1996.

M. Carroll and C. Chien. Decay of reverberant sound in a spherical enclosure. *The Journal of the Acoustical Society of America*, 62(6):1442–1446, 1977.

A. Chandak, C. Lauterbach, M. Taylor, Z. Ren, and D. Manocha. Ad-frustum: Adaptive frustum tracing for interactive sound propagation. *Visualization and Computer Graphics, IEEE Transactions on*, 14(6):1707–1722, 2008.

M. Chemistruck, K. Marcolini, and W. Pirkle. Generating matrix coefficients for feedback delay networks using genetic algorithm. In *Proceedings of the 133rd Audio Engineering Society Convention*, San Francisco, CA, USA, 2012.

D. H. Cooper and J. L. Bauck. Prospects for transaural recording. *Journal of the Audio Engineering Society*, 37(1/2):3–19, 1989.

B. Cowan and B. Kapralos. Spatial sound for video games and virtual environments utilizing real-time gpu-based convolution. In *Proceedings of the 2008 Conference on Future Play: Research, Play, Share*, pages 166–172. ACM, 2008.

L. Dahl and J.-M. Jot. A reverberator based on absorbent all-pass filters. In *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00)*, Verona, Italy, December 2000.

B.-I. L. Dalenbäck. Room acoustic prediction based on a unified treatment of diffuse and specular reflection. *The journal of the Acoustical Society of America*, 100(2):899–909, 1996.

E. De Sena, H. Hacihabiboglu, and Z. Cvetkovic. Scattering delay network: an interactive reverberator for computer games. In *Audio Engineering Society Conference: 41st International Conference: Audio for Games*. Audio Engineering Society, 2011.

I. Dokmanić, Y. M. Lu, and M. Vetterli. Can one hear the shape of a room: The 2-d polygonal case. In *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, pages 321–324. IEEE, 2011.

I. Dokmanić, R. Parhizkar, A. Walther, Y. M. Lu, and M. Vetterli. Acoustic echoes reveal room shape. *Proceedings of the National Academy of Sciences*, 110(30):12186–12191, July 2013.

P. Dutre, P. Bekaert, and K. Bala. *Advanced global illumination*. CRC Press, 2006.

F. A. Everest, K. C. Pohlmann, and T. Books. *The master handbook of acoustics*, volume 4. McGraw-Hill New York, 2001.

H. Fletcher and W. A. Munson. Loudness, its definition, measurement and calculation*. *Bell System Technical Journal*, 12(4):377–430, 1933.

J. Frenette. *Reducing artificial reverberation algorithm requirements using time-variant feedback delay networks*. PhD thesis, University of Miami, 2000.

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*, pages 21–32, Orlando, FL, USA, July 1998.

C. M. Goral, K. E. Torrance, D. P. Greenberg, and B. Battaile. Modeling the interaction of light between diffuse surfaces. In *ACM SIGGRAPH Computer Graphics*, volume 18, pages 213–222. ACM, 1984.

H. Hacihabiboğlu, E. De Sena, and Z. Cvetković. Frequency-domain scattering delay networks for simulating room acoustics in virtual environments. In *Signal-Image Technology and Internet-Based Systems (SITIS), 2011 Seventh International Conference on*, pages 180–187. IEEE, 2011.

P. S. Heckbert and P. Hanrahan. Beam tracing polygonal objects. In *Proceedings of the 11th annual conference on Computer graphics and interactive techniques*, volume 18, pages 119–127, New York, NY, USA, July 1984.

R. Heinz. Binaural room simulation based on an image source model with addition of statistical methods to include the diffuse sound scattering of walls and to predict the reverberant tail. *Applied Acoustics*, 38(2):145–159, 1993.

M. Hodgson and E.-M. Nosal. Experimental evaluation of radiosity for room sound-field prediction. *The Journal of the Acoustical Society of America*, 120(2):808–819, 2006.

B. Holm-Rasmussena, H.-M. Lehtonenb, and V. Välimäki. A new reverberator based on variable sparsity convolution. *work*, 5(6):7–8, 2013.

J. Huopaniemi, V. Välimäki, M. Karjalainen, and R. Väänänen. Efficient and parametric reverberator for room acoustics modeling. In *Proceedings of the International Computer Music Conference*, pages 200–203, Thessaloniki, Greece, September 1997.

J. Hurchalla. A time distributed fft for efficient low latency convolution. In *Audio Engineering Society Convention 129*. Audio Engineering Society, 2010.

ISO226. Acoustics – normal equal-loudness-level contours. 2003.

ISO9613-3. Acoustics – attenuation of sound during propagation outdoors – part 1: Calculation of the absorption of sound by the atmosphere, international norm. 1996.

ITU-1534. Method for the subjective listening tests of intermediate audio quality. Geneva, Switzerland, 2001.

M. Jeub, M. Schäfer, and P. Vary. A binaural room impulse response database for the evaluation of dereverberation algorithms. In *Digital Signal Processing, 2009 16th International Conference on*, pages 1–5. IEEE, 2009.

J.-M. Jot. An analysis/synthesis approach to real-time artificial reverberation. In *Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on*, volume 2, pages 221–224. IEEE, 1992a.

J.-M. Jot. *Etude et réalisation d'un spatialisateur de sons par modèles physiques et perceptifs*. PhD thesis, l'Ecole Nationale Supérieure des Télécommunications, Télécom Paris, 1992b.

J.-M. Jot. Efficient models for reverberation and distance rendering in computer music and virtual audio reality. In *Proc. 1997 International Computer Music Conference. http://citeseer. ist. psu. edu/jot97efficient. html*, 1997.

J.-M. Jot and A. Chaigne. Digital delay networks for designing artificial reverberators. In *Proceedings of the 90th Audio Engineering Society Convention*, Paris, France, February 1991.

J. T. Kajiya. The rendering equation. In *ACM Siggraph Computer Graphics*, volume 20, pages 143–150. ACM, 1986.

B. Kapralos, M. Jenkin, and E. Milios. Sonel mapping: A stochastic acoustical modeling system. In *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, volume 5, pages V–V. IEEE, 2006.

S. Kiminki. *Sound propagation theory for linear ray acoustic modelling*. PhD thesis, MSc. Thesis. Helsinki University Of Technology, Department of Computer Science and Engineering, 2005.

M. Kleiner, B.-I. Dalenbäck, and P. Svensson. Auralization-an overview. *Journal of the Audio Engineering Society*, 41(11):861–875, 1993.

G. I. Koutsouris, J. Brunskog, C.-H. Jeong, and F. Jacobsen. Combination of acoustical radiosity and the image source method. *The Journal of the Acoustical Society of America*, 133(6):3963–3974, 2013a.

G. I. Koutsouris, J. Brunskog, C.-H. Jeong, and F. Jacobsen. Combination of acoustical radiosity and the image source method. *The Journal of the Acoustical Society of America*, 133(6):3963–3974, 2013b.

K. Kowalczyk and M. Van Walstijn. Room acoustics simulation using 3-d compact explicit fdtd schemes. *Audio, Speech, and Language Processing, IEEE Transactions on*, 19(1):34–46, 2011.

W. Krebber, H.-W. Gierlich, and K. Genuit. Auditory virtual environments: basics and applications for interactive simulations. *Signal processing*, 80(11):2307–2322, 2000.

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*, 8(1):118–125, 1968.

M. Kubovy and D. Van Valkenburg. Auditory and visual objects. *Cognition*, 80(1):97–126, 2001.

A. Kulowski. Error investigation for the ray tracing technique. *Applied Acoustics*, 15(4):263–274, 1982.

A. Kulowski. Algorithmic representation of the ray tracing technique. *Applied Acoustics*, 18(6): 449–469, 1985.

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

S. Laine, S. Siltanen, T. Lokki, and L. Savioja. Accelerated beam tracing algorithm. *Applied Acoustics*, 70(1):172–181, 2009.

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*, 13(6):1672–1679, 2007.

H. Lee and B.-H. Lee. An efficient algorithm for the image model technique. *Applied Acoustics*, 24(2):87–115, 1988.

S.-R. Lee, J.-S. No, E.-H. Shin, and H. Chung. On eigenvalues of row-inverted sylvester hadamard matrices. *Results in Mathematics*, 54(1-2):117–126, 2009.

E. Lehmann, A. M. Johansson, et al. Diffuse reverberation model for efficient image-source simulation of room impulse responses. *Audio, Speech, and Language Processing, IEEE Transactions on*, 18(6):1429–1439, 2010.

E. A. Lehmann and A. M. Johansson. Prediction of energy decay in room impulse responses simulated with an image-source model. *The Journal of the Acoustical Society of America*, 124 (1):269–277, 2008.

H. Lehnert and J. Blauert. Principles of binaural room simulation. *Applied Acoustics*, 36(3):259–291, 1992.

S. Lesoinne and J.-J. Embrechts. Size-adaptive spherical receptor acceleration method for acoustical ray tracing. *Journal of the Acoustical Society of America*, 123(5):3769, 2008.

T. Lewers. A combined beam tracing and radiatn exchange computer model of room acoustics. *Applied Acoustics*, 38(2):161–178, 1993.

R. Loyet. *Dynamic sound rendering of complex environments*. PhD thesis, Université Claude Bernard Lyon 1, 2012.

D. Markovic, F. Antonacci, A. Sarti, and S. Tubaro. 3d beam tracing based on visibility lookup for interactive acoustic modeling. 2016.

J. Martin, D. Van Maercke, and J.-P. Vian. Binaural simulation of concert halls: A new approach for the binaural reverberation process. *The Journal of the Acoustical Society of America*, 94(6): 3255–3264, 1993.

D. A. Mauro, R. Mekuria, and M. Sanna. Binaural spatialization for 3d immersive audio communication in a virtual world. In *Proceedings of the 8th Audio Mostly Conference*, page 8. ACM, 2013.

S. G. McGovern. Fast image method for impulse response calculations of box-shaped rooms. *Applied Acoustics*, 70(1):182–189, 2009.

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*, Tokyo, Japan, October 2010.

F. Menzer and C. Faller. Binaural reverberation using a modified jot reverberator with frequency-dependent interaural coherence matching. In *Audio Engineering Society Convention 126*. Audio Engineering Society, 2009.

F. Menzer and C. Faller. Unitary matrix design for diffuse jot reverberators. In *Audio Engineering Society Convention 128*. Audio Engineering Society, 2010.

R. Mignot, G. Chardon, and L. Daudet. Compressively sampling the plenacoustic function. In *SPIE Optical Engineering+ Applications*, pages 813808–813808. International Society for Optics and Photonics, 2011a.

R. Mignot, L. Daudet, and F. Ollivier. Compressed sensing for acoustic response reconstruction: Interpolation of the early part. In *Applications of Signal Processing to Audio and Acoustics (WASPAA), 2011 IEEE Workshop on*, pages 225–228. IEEE, 2011b.

J. D. Miller and E. M. Wenzel. Recent developments in SLAB: A software based system for interactive spatial sound synthesis. In *Eighth International Conference on Auditory Display*, pages 403–408, Kyoto, Japan, 2002.

H. Møller. Reproduction of artificial-head recordings through loudspeakers. *Journal of the Audio Engineering Society*, 37(1/2):30–33, 1989.

B. C. Moore. *An introduction to the psychology of hearing*. Brill, 2012.

J. A. Moorer. About this reverberation business. *Computer Music Journal*, pages 13–28, 1979.

S. Müller and P. Massarani. Transfer-function measurement with sweeps. *Journal of the Audio Engineering Society*, 49(6):443–471, 2001.

D. Murphy and R. Stewart. A hybrid artificial reverberation algorithm. In *Audio Engineering Society Convention 122*. Audio Engineering Society, 2007.

G. M. Naylor. Odeon—another hybrid room acoustical model. *Applied Acoustics*, 38(2):131–143, 1993.

F. E. Nicodemus, J. C. Richmond, J. J. Hsia, I. W. Ginsberg, and T. Limperis. *Geometrical considerations and nomenclature for reflectance*, volume 160. US Department of Commerce, National Bureau of Standards Washington, DC, USA, 1977.

E.-M. Nosal, M. Hodgson, and I. Ashdown. Improved algorithms and methods for room sound-field prediction by acoustical radiosity in arbitrary polyhedral rooms. *The Journal of the Acoustical Society of America*, 116(2):970–980, 2004.

Open-AIR. Open acoustic impulse response library. *[online]* http://www.openairlib.net/.

M. Pasin, A. Frisiello, J. Wall, S. Poulakos, and A. Smolic. A methodological approach to user evaluation and assessment of a virtual environment hangout.

A. Pietrzyk. Computer modeling of the sound field in small rooms. In *Audio Engineering Society Conference: 15th International Conference: Audio, Acoustics & Small Spaces*. Audio Engineering Society, 1998.

A. Primavera, S. Cecchi, J. Li, and F. Piazza. Objective and subjective investigation on a novel method for digital reverberator parameters estimation. *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, 22(2):441–452, 2014.

Z. Rafii and B. Pardo. Learning to control a reverberator using subjective perceptual descriptors. In *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR)*, pages 285–290, Kobe, Japan, October 2009.

N. Raghuvanshi, J. Snyder, R. Mehra, M. Lin, and N. Govindaraju. Precomputed wave simulation for real-time sound propagation of dynamic sources in complex scenes. In *ACM Transactions on Graphics (TOG)*, volume 29, page 68. ACM, 2010.

J. H. Rindel. The use of computer modeling in room acoustics. *Journal of Vibroengineering*, 3(4): 41–72, 2000.

D. Rocchesso. Maximally diffusive yet efficient feedback delay networks for artificial reverberation. *Signal Processing Letters, IEEE*, 4(9):252–255, 1997.

D. Rocchesso and J. O. Smith. Circulant and elliptic feedback delay networks for artificial reverberation. *IEEE Transactions on Speech and Audio Processing*, 5(1):51–63, 1997a.

D. Rocchesso and J. O. Smith. Circulant and elliptic feedback delay networks for artificial reverberation. *Speech and Audio Processing, IEEE Transactions on*, 5(1):51–63, 1997b.

S. Sakamoto, A. Ushiyama, and H. Nagatomo. Numerical analysis of sound propagation in rooms using the finite difference time domain method. *The Journal of the Acoustical Society of America*, 120(5):3008–3008, 2006.

J. Sánchez and M. Lumbreras. Virtual environment interaction through 3d audio by blind children. *CyberPsychology & Behavior*, 2(2):101–111, 1999.

J. Sandvad. Dynamic aspects of auditory virtual environments. In *Audio Engineering Society Convention 100*. Audio Engineering Society, 1996.

L. Savioja. Real-time 3d finite-difference time-domain simulation of low-and mid-frequency room acoustics. In *13th Int. Conf on Digital Audio Effects*, volume 1, page 75, 2010.

L. Savioja and U. P. Svensson. Overview of geometrical room acoustic modeling techniques. *The Journal of the Acoustical Society of America*, 138(2):708–730, 2015.

L. Savioja, J. Backman, A. Järvinen, and T. Takala. Waveguide mesh method for low-frequency simulation of room acoustics. 1995.

L. Savioja, J. Huopaniemi, T. Lokki, and R. Väänänen. Creating interactive virtual acoustic environments. *Journal of the Audio Engineering Society*, 47(9):675–705, 1999.

L. Savioja, T. J. Rinne, and T. Takala. Simulation of room acoustics with a 3-{D} finite difference mesh. 2006.

L. Savioja, D. Manocha, and M. Lin. Use of gpus in room acoustic modeling and auralization. In *Proc. Int. Symposium on Room Acoustics*, 2010.

J. W. Scarpaci and J. A. White. A system for real-time virtual auditory space. 2005.

M. R. Schroeder. Natural sounding artificial reverberation. *Journal of the Audio Engineering Society*, 10(3):219–223, 1962.

M. R. Schroeder. New method of measuring reverberation time. *The Journal of the Acoustical Society of America*, 37(3):409–412, 1965.

S. Siltanen and T. Lokki. Diffraction modeling in acoustic radiance transfer method. *Journal of the Acoustical Society of America*, 123(5):3759, 2008.

S. Siltanen, T. Lokki, S. Kiminki, and L. Savioja. The room acoustic rendering equation. *The Journal of the Acoustical Society of America*, 122(3):1624–1635, 2007.

S. Siltanen, T. Lokki, and L. Savioja. Frequency domain acoustic radiance transfer for real-time auralization. *Acta Acustica united with Acustica*, 95(1):106–117, 2009.

S. Siltanen, T. Lokki, and L. Savioja. Rays or waves? understanding the strengths and weaknesses of computational room acoustics modeling techniques. In *Proc. Int. Symposium on Room Acoustics*, 2010a.

S. Siltanen, T. Lokki, and L. Savioja. Room acoustics modeling with acoustic radiance transfer. *Proc. ISRA Melbourne*, 2010b.

S. Siltanen, A. Southern, and L. Savioja. Modeling reflections from rough surfaces in complex spaces. *The Journal of the Acoustical Society of America*, 132(3):1889–1889, 2012.

A. Silzle, H. Strauss, and P. Novo. Ika-sim: A system to generate auditory virtual environments. In *Audio Engineering Society Convention 116*. Audio Engineering Society, 2004.

J. O. Smith. A new approach to digital reverberation using closed waveguide networks. 1985.

J. O. Smith. *Physical audio signal processing: For virtual musical instruments and audio effects*. W3K Publishing, 2010.

A. Southern, S. Siltanen, and L. Savioja. Spatial room impulse responses with a hybrid modeling method. In *Audio Engineering Society Convention 130*. Audio Engineering Society, 2011.

A. Southern, S. Siltanen, D. T. Murphy, and L. Savioja. Room impulse response synthesis and validation using a hybrid acoustic model. *Audio, Speech, and Language Processing, IEEE Transactions on*, 21(9):1940–1952, 2013.

J. Stautner and M. Puckette. Designing multi-channel reverberators. *Computer Music Journal*, 6 (1):52–65, 1982.

J. C. Strikwerda. *Finite difference schemes and partial differential equations*. Siam, 2004.

J. W. Strutt. On our perception of sound direction. *Philosophical Magazine*, 13:214–232, 1907.

P. Svensson and U. R. Kristiansen. Computational modelling and simulation of acoutic spaces. In *Audio Engineering Society Conference: 22nd International Conference: Virtual, Synthetic, and Entertainment Audio*. Audio Engineering Society, 2002.

N. Tsingos, T. Funkhouser, A. Ngan, and I. Carlbom. Modeling acoustics in virtual environments using the uniform theory of diffraction. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 545–552. ACM, 2001.

V. Välimäki, 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*, 20(5):1421–1448, 2012.

V. Välimäki, J. Parker, L. Savioja, J. O. Smith, and J. Abel. More than 50 years of artificial reverberation. In *Audio Engineering Society Conference: 60th International Conference: DREAMS (Dereverberation and Reverberation of Audio, Music, and Speech)*. Audio Engineering Society, 2016.

M. Van Walstijn and K. Kowalczyk. On the numerical solution of the 2d wave equation with compact fdtd schemes. In *Proc. Int. Conf. Digital Audio Effects (DAFx'08)*, pages 205–212. Citeseer, 2008.

E. Vickers, J.-L. L. Wu, P. G. Krishnan, and R. N. K. Sadanandam. Frequency domain artificial reverberation using spectral magnitude decay. In *Proc. AES 121th Convention*, 2006.

M. Vorländer. Simulation of the transient and steady-state sound propagation in rooms using a new combined ray-tracing/image-source algorithm. *The Journal of the Acoustical Society of America*, 86(1):172–178, 1989.

M. Vorländer. *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*. Springer Verlag, 2008.

E. M. Wenzel. *Three-dimensional virtual acoustic displays*. National Aeronautics and Space Administration, Ames Research Center, 1991.

E. M. Wenzel, J. D. Miller, and J. S. Abel. Sound lab: A real-time, software-based system for the study of spatial hearing. In *Audio Engineering Society Convention 108*. Audio Engineering Society, 2000.

B. Xie. *Head-related transfer function and virtual auditory display*. J Ross, 2013.

W. A. Yost. *Fundamentals of hearing: An introduction*. Academic Press, 1994.

C. Zhang and B. Xie. Platform for dynamic virtual auditory environment real-time rendering system. *Chinese Science Bulletin*, 58(3):316–327, 2013.

Appendix A

List of Abbreviations

3D Three dimensional. v, vi, 47, 51

ARE Acoustic Rendering Equation. 15, 29, 31, 53, 115, 116, 122, 133

ARN Acoustic Rendering Networks. 135, 142, 148, 151, 152

BEM Boundary Element Methods. xvi, 47, 48, 58

BRDF Bidirectional Reflectance Distribution Function. 116, 118, 119, 121, 124, 138

CPU Computation Processing Unit. 152

CS Compressed Sensing. 41

DFT Discrete Fourier Transform. 64

DWM Digital Waveguide Mesh. 56

EDC Energy Decay Curve. 40

ER Early Reflections. 39, 115

FDN-RTM Feedback Delay Networks (based on) Radiance Transfer Method. 70, 71, 73, 82–88

FDNs Feedback Delay Networks. v, xvi, 28, 29, 55–58, 66, 67, 70, 73, 76

FDTD Finite-Difference Time-Domain. xvi, 47, 48, 57, 58

FEM Finite Element Methods. xvi, 47, 48, 58

FIR Finite Impulse Response. 57, 74

GPU Graphics Processing Unit. 54

HRTF Head-Related Transfer Function. 43

IIR Infinite Impulse Response. 74

ILD Interaural Level Difference. 43

ISM Image Source Method. xvi, 28, 49, 50, 56–58

ITD Interaural Time Difference. 43

LR Late Reverberation. 39, 40, 115

LTI Linear and Time-Invariant. 26, 41

MLS Maximum Length Sequence. 39

MUSHRA MUltiple Stimulus with Hidden Reference and Anchor. 88

RIR Room Impulse Response. 27, 29, 39, 40, 50, 51, 60, 88

RMS Root Mean Square. 34

RT Reverberation Time. 44

RTM Radiance Transfer Method. v, xvi, 28, 52, 57, 58, 60, 66, 67, 69, 70, 85–88, 115, 116

SNR Signal to Noise Ratio. 41

SPL Sound Pressure Level. 43