Analysis and Synthesis of Directional Reverberation

Benoit Alary
Analysis and Synthesis of Directional Reverberation

Benoit Alary

A doctoral dissertation completed for the degree of Doctor of Science (Technology) to be defended, with the permission of the Aalto University School of Electrical Engineering, at a public examination held online on September 20, 2021, at 16:00.

Aalto University
School of Electrical Engineering
Department of Signal Processing and Acoustics
Audio Signal Processing
Supervising professor
Prof. Vesa Välimäki, Aalto University, Finland

Thesis advisors
Dr. Archontis Politis, Tampere University, Finland
Prof. Sebastian J. Schlecht, Aalto University, Finland
Prof. Vesa Välimäki, Aalto University, Finland

Preliminary examiners
Prof. Hüseyin Hacihabiboğlu, Middle East Technical University, Turkey
Prof. Sylvain Marchand, La Rochelle University, France

Opponent
Dr. Jean-Marc Jot, iZotope Inc., United States of America
Abstract

The reproduction of acoustics is one of the key challenges in spatial sound reproduction. In order to reach high levels of realism and immersion, reproduced signals must be processed through a set of filters accounting for real-life phenomena. From the scattering and absorption of sound on walls to its diffraction around objects and our head, the propagation of sound waves in a room creates a complex sound field around a listener. While recreating all of the underlying parts of sound propagation is not yet within our reach for real-time applications, spatial sound techniques aim to minimize the computational cost by focusing the efforts on the most perceptually critical components.

In this dissertation, the reproduction of reverberant sound fields is investigated through the development of analysis methods, reverberation algorithms, and decorrelation techniques. A particular emphasis is given to the perception and reproduction of directional characteristics present in late reverberation.

Most reverberation algorithms consider that the sound energy is evenly distributed across space in all directions, after an initial period of time, since the diffusion of energy leads to more homogeneous and isotropic sound fields. However, previous work has demonstrated that insufficient diffusion in a room leads to anisotropic, and directional, late reverberation.

In this dissertation, a complete framework is proposed for the objective and subjective analysis of directional characteristics as well as a novel delay-network reverberation method capable of producing direction-dependent decay properties. The reverberator is also expanded to offer efficient frequency- and direction-dependent processing. The proposed algorithm contains all the required elements for the auralization of reverberant sound fields, which may be modulated in real-time to support six-degrees-of-freedom sound reproduction.

The decorrelation of audio signals, which occurs naturally during the diffusion of energy in a sound field, is another important aspect of sound reproduction. This dissertation considers the use of velvet-noise sequences, a special type of sparse noise signals, as decorrelation filters and offers a method to optimize their characteristics. Velvet noise is also proposed for the reproduction of an existing impulse response using a small set of time- and frequency-dependent information, along with a reverberator using velvet noise to improve the echo density of a delay network reverberator.

Overall, the results contained in this dissertation offer new insights into the perceptual ramifications of reverberant sound fields containing directional characteristics and their reproduction. The methods presented bring applications in the context of immersive sound reproduction, such as in virtual and augmented reality.

Keywords Acoustics, reverberation, decorrelation, digital signal processing, signal analysis
The research work detailed in this dissertation has been carried out in the Acoustics Lab, part of the Department of Signal Processing and Acoustics at Aalto University. This research was part of the ICHO project (Immersive Concert at Home, September 2016-2020) funded by the Academy of Finland (Aalto University project no.13296390) and the work was part of the activities of the “Nordic Sound and Music Computing Network—NordicSMC”, NordForsk project number 86892. Part of the work of this dissertation was conducted at the Institut de Recherche et Coordination Acoustique/Musique (IRCAM) (UMR STMS IRCAM-CNRS-Sorbonne Université) in Paris during a research visit in the Autumn of 2018. The research visit was funded by the Foundation for Aalto University Science and Technology.

The main research question that stemmed from the funding project was how to best reproduce the acoustics of a concert hall on immersive media platforms such as virtual reality. One of the main research objective was to develop new efficient reproduction methods using velvet noise. Towards these goals, the perceptual characteristics of reverberant sound fields were studied along with decorrelation techniques and artificial reverberation methods.

I would like to thank my supervisor, Prof. Vesa Välimäki, for sharing his vast expertise in the field and for all of his guidance throughout this Ph.D., as well as his unwavering trust in allowing me to pursue my research objectives. I would also like to thank my thesis advisors, Dr. Archontis Politis and Prof. Sebastian J. Schlecht, for their invaluable contributions to this work as well as for their support and guidance throughout my research. I also thank the pre-examiners, Prof. Huseyin Hacıhabiboğlu and Prof. Sylvain Marchand, for their comments that helped improve this dissertation, and Dr. Jean-Marc Jot for agreeing to be my opponent.

For hosting me at their facilities at IRCAM in Paris and for a fruitful collaboration, I thank Markus Noisternig, Pierre Massé, Thibault Carpentier, and Olivier Warusfel. I also thank Luis R. J. Costa, whose mastery of the English language was invaluable when publishing the work contained in this dissertation.
I thank my many colleagues at the Acoustics Lab for all the profound conversations and their general awesomeness throughout my time at Aalto. Doing a Ph.D. is a complex journey, let alone doing it in a land that intermittently dips in complete darkness. Having proximity to such a stellar group of colleagues who can encourage, challenge, inspire, and support each other is immensely valuable. For these reasons, I thank Prof. Tapio Lokki, Prof. Ville Pulkki, Prof. Lauri Savioja, Abraham, Aki, Aksu, Alec, Aleks, Alessandro, Andrea, Antti, Catarina, Christoph, Craig, Dimitri, Eloi, Étienne, Eero-Pekka, Fabiàn, Georg, Henna, Henri, Ilkka, Jan, Janani, Janne, Javier, Jon, Jose, Juhani, Juho, Julie, Jukka, Jussi, Karolina, Lauri, Lauros, Leo, Leonardo, Michael, Nils, Otto, Pablo, Pedro, Petteri, Raimondo, Rapolas, Ricardo, Sebastian, Sakari, Stefan, Symeon, Taeho, Thomas, Vasileios, and Winfried...

I also thank my close friends in the research community who’s work has been inspiring, Brian, Julian, Michele, Stefan, and Thanasis. Finally, I would like to thank my family who has always been so supportive of my endeavours and especially my parents, Sylvie and Pierre, for being such immovable pillars of support.

Helsinki, August 18, 2021,

Benoit Alary
Contents

Preface 1

Contents 3

List of Publications 5

Author's Contribution 7

List of Abbreviations 9

List of Symbols 11

1. Introduction 13

2. Reverberant Sound Field 17
   2.1 Sound Propagation ............................. 17
   2.2 Statistical Properties .......................... 21
   2.3 Diffuse Field .................................. 27
   2.4 Properties of Reverberant Sound Fields .......... 28
      2.4.1 Objective Measures ....................... 28
      2.4.2 Perception ............................... 29
   2.5 Reproduction of Sound Fields ................... 30
      2.5.1 Ambisonics ............................... 31
      2.5.2 Binaural Reproduction .................... 33

3. Decorrelation 35
   3.1 Overview ...................................... 35
   3.2 Objective Evaluation ........................... 36
      3.2.1 Cross-Correlation ......................... 36
      3.2.2 Coherence ............................... 37
      3.2.3 Covariance Matrix ......................... 37
      3.2.4 Incoherence Profile ....................... 38
   3.3 Decorrelation Methods .......................... 39
### Contents

3.4 Velvet-Noise Decorrelation ........................................ 40

4. **Artificial Reverberation** ........................................ 43
   4.1 Historical Overview ........................................... 43
   4.2 Captured Impulse Response ................................... 44
   4.3 Virtual Acoustics .............................................. 46
     4.3.1 Geometric Methods ....................................... 46
     4.3.2 Wave-based Methods .................................... 47
   4.4 Artificial Noise .............................................. 48
   4.5 Delay Networks ............................................... 50
     4.5.1 Velvet-Noise Feedback-Delay Network ............... 57

5. **Hybrid Reverberation Methods** ................................ 59
   5.1 Geometrically-Informed Reverberators ..................... 59
   5.2 Binaural Reverberation ...................................... 60
   5.3 Digital Waveguide ........................................... 63
   5.4 Ambisonics Delay Networks .................................. 64
   5.5 Directional Reverberation ................................... 65
     5.5.1 Analysis ................................................ 66
     5.5.2 Reproduction ............................................ 68
     5.5.3 Directional Feedback Delay Network ............... 69
     5.5.4 Frequency-Dependent Directional Feedback De-
         lay Network .................................................. 70

6. **Summary of Main Results** ...................................... 75

7. **Conclusion** ..................................................... 79

References ............................................................ 81

Errata ................................................................. 97

Publications .......................................................... 99
List of Publications

This thesis consists of an overview and of the following publications which are referred to in the text by their Roman numerals.


III Vesa Välimäki, Bo Holm-Rasmussen, Benoit Alary, and Heidi Maria Lehtonen. Late reverberation synthesis using filtered velvet noise. *Applied Sciences*, vol. 7, no. 5, May 2017.


VI Benoit Alary, Archontis Politis, Sebastian J. Schlecht, and Vesa Välimäki. Directional feedback delay network. *Journal of the Audio Engineering Society*, vol. 67, no. 10, pp. 752–762, October 2019.
Author’s Contribution

Publication I: “Velvet-noise decorrelator”

The present author developed and implemented the method, produced all of the figures, and wrote most of the manuscript. The second author had the original idea for the method and wrote part of the introduction.

Publication II: “Optimized velvet-noise decorrelator”

The author provided the original implementation for velvet-noise decorrelation. The first author had the idea for the improved method, implemented it, and wrote most of the article. The present author helped validate and fine-tune the method. He also planned, implemented, and conducted the two perceptual studies detailed in the article and helped produce Figure 6. The author wrote part of the introduction and most of Section 5. The author co-presented the paper in the conference.

Publication III: “Late reverberation synthesis using filtered velvet noise”

The author was responsible for improving the initial implementation of the algorithm. The author planned, implemented, and conducted the perceptual study. The author participated in the writing of Section 7 and produced Figures 5, 10, 11, and 16. He also produced the data used in Tables 2, 3, 4, and 5.
Publication IV: “Velvet-noise feedback delay network”

The author planned the research project with the co-authors. The present author also assisted and provided research directions to the first author throughout the development, implementation, and evaluation of the method. The author wrote the abstract, co-wrote the introduction and parts of the manuscript. He produced Figure 8 and helped produce Figures 6 and 7.

Publication V: “Perceptual analysis of directional late reverberation”

The author planned the research project with the co-authors. The present author implemented the method and produced all of the figures except for Figure 1. He also wrote most of the manuscript, except for the background Sections 2.A and 2.C-D. The author implemented and conducted the perceptual study.

Publication VI: “Directional feedback delay network”

The author produced the original idea for the method and developed the algorithm. He wrote most of the manuscript and produced all of the figures. The second author developed the recursive spatial transform and wrote Sections 1.3, 1.4, as well as Sections 3.2-3.4. The co-authors helped formalize the method and the terminology used in the manuscript.

Publication VII: “Frequency-dependent directional feedback delay network”

The author implemented the method, produced all of the figures, and wrote the manuscript. The co-author contributed to the original idea, provided critical feedback, and proofreading.
### List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACN</td>
<td>Ambisonics channel number</td>
</tr>
<tr>
<td>AP</td>
<td>all-pass filter</td>
</tr>
<tr>
<td>DFDN</td>
<td>directional feedback delay network</td>
</tr>
<tr>
<td>DIR</td>
<td>directional impulse response</td>
</tr>
<tr>
<td>DirAC</td>
<td>directional audio coding</td>
</tr>
<tr>
<td>DWG</td>
<td>digital waveguide</td>
</tr>
<tr>
<td>EDC</td>
<td>energy-decay curve</td>
</tr>
<tr>
<td>EDD</td>
<td>energy-decay deviation</td>
</tr>
<tr>
<td>EDR</td>
<td>energy-decay relief</td>
</tr>
<tr>
<td>FDN</td>
<td>feedback-delay network</td>
</tr>
<tr>
<td>FDTD</td>
<td>finite-difference time-domain</td>
</tr>
<tr>
<td>FFT</td>
<td>fast Fourier transform</td>
</tr>
<tr>
<td>FIR</td>
<td>finite impulse response</td>
</tr>
<tr>
<td>FLOPS</td>
<td>floating-point operations</td>
</tr>
<tr>
<td>FOA</td>
<td>first-order Ambisonics</td>
</tr>
<tr>
<td>HRTF</td>
<td>head-related transfer function</td>
</tr>
<tr>
<td>IACC</td>
<td>interaural cross-correlation</td>
</tr>
<tr>
<td>IEDD</td>
<td>interaural energy-decay deviation</td>
</tr>
<tr>
<td>IFFT</td>
<td>inverse fast Fourier transform</td>
</tr>
<tr>
<td>IIR</td>
<td>infinite impulse response</td>
</tr>
<tr>
<td>ILD</td>
<td>interaural level difference</td>
</tr>
<tr>
<td>ISM</td>
<td>image-source method</td>
</tr>
<tr>
<td>IR</td>
<td>impulse response</td>
</tr>
<tr>
<td>ITD</td>
<td>interaural time difference</td>
</tr>
<tr>
<td>SDM</td>
<td>spatial decomposition method</td>
</tr>
<tr>
<td>SH</td>
<td>spherical harmonics</td>
</tr>
</tbody>
</table>
List of Abbreviations

SIR  spatial impulse response
SIRR spatial-impulse-response rendering
SMA spherical microphone array
SNR  signal-to-noise ratio
VBAP vector-base amplitude panning
VNS  velvet-noise sequence
List of Symbols

A  recirculating matrix

\(c\)  speed of sound

\(D\)  distance

\(f\)  frequency

\(f_s\)  sample rate

\(f_{\text{Schroeder}}\)  Schroeder frequency

\(g\)  attenuation coefficient

\(G\)  directional gain matrix

\(G\)  absorbent filter

\(h\)  impulse response

\(h_{\text{noise}}\)  noise signal

\(H\)  coloration filter

\(j\)  imaginary unit

\(k\)  wavenumber

\(K\)  number of plane waves

\(L_d\)  length of a discontinuity

\(L\)  number of spherical harmonics

\((L_x, L_y, L_z)\)  room dimensions

\(n\)  sample number

\((n_x, n_y, n_z)\)  mode number

\(p\)  sound pressure

\(P\)  number of directional signals

\(Q\)  number of plane waves

\(r\)  radius

\(R\)  reflected factor

\(s\)  Ambisonics signal
List of Symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{s} )</td>
<td>plane-wave signal</td>
</tr>
<tr>
<td>( S )</td>
<td>surface</td>
</tr>
<tr>
<td>( S_\alpha )</td>
<td>equivalent absorptive area</td>
</tr>
<tr>
<td>( t )</td>
<td>time</td>
</tr>
<tr>
<td>( T )</td>
<td>directional weighting transform</td>
</tr>
<tr>
<td>( T_{60} )</td>
<td>reverberation time</td>
</tr>
<tr>
<td>( v )</td>
<td>impulse vector</td>
</tr>
<tr>
<td>( V )</td>
<td>volume</td>
</tr>
<tr>
<td>( w )</td>
<td>beamforming weight</td>
</tr>
<tr>
<td>( x )</td>
<td>input signal</td>
</tr>
<tr>
<td>( x )</td>
<td>multichannel input signal</td>
</tr>
<tr>
<td>( y )</td>
<td>output signal</td>
</tr>
<tr>
<td>( y )</td>
<td>multichannel output signal</td>
</tr>
<tr>
<td>( Y(\phi, \theta) )</td>
<td>spherical harmonics coefficient</td>
</tr>
<tr>
<td>( y(\phi, \theta) )</td>
<td>spherical harmonics vector</td>
</tr>
<tr>
<td>( Y )</td>
<td>spherical harmonics matrix</td>
</tr>
<tr>
<td>( Z )</td>
<td>wall impedance</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>absorption coefficient</td>
</tr>
<tr>
<td>( \delta )</td>
<td>Dirac delta function</td>
</tr>
<tr>
<td>( \Delta H )</td>
<td>differential filter</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>cross-correlation</td>
</tr>
<tr>
<td>( \Gamma_i )</td>
<td>acoustic admittance</td>
</tr>
<tr>
<td>( \Gamma )</td>
<td>covariance matrix</td>
</tr>
<tr>
<td>( \nabla )</td>
<td>Laplace operator</td>
</tr>
<tr>
<td>( \phi_{max} )</td>
<td>maximum angle</td>
</tr>
<tr>
<td>( (\phi, \theta) )</td>
<td>spherical coordinate</td>
</tr>
<tr>
<td>( \phi_i )</td>
<td>phase lag</td>
</tr>
<tr>
<td>( \rho )</td>
<td>density parameter</td>
</tr>
<tr>
<td>( \rho_0 )</td>
<td>characteristic impedance of air</td>
</tr>
<tr>
<td>( \omega )</td>
<td>frequency band</td>
</tr>
</tbody>
</table>
1. Introduction

The reverberation of sound plays a crucial role in our perception of a space: from the first few reflections, giving us a perceptual hint of the proximity of walls, to the late reverberation, giving us some intuition on the overall shape and materials in a room [1]. As such, the reverberation properties of a room are determined by its shape, the absorbing material it contains, and the scattering properties of its surfaces [2]. Although the acoustics of a concert hall are carefully crafted to ensure a pleasing and enveloping listening experience, many other types of rooms exist with rich and varied acoustic properties creating a unique reverberation in each room [3].

Artificial reverberation is an important aspect of sound reproduction, which strives to efficiently reproduce the perceptual qualities of sound waves propagating in a room [4]. Several artificial reverberation methods exist, and the choice of one method over another is usually determined by its application. For instance, music production requires algorithms with good tuning capabilities and an aesthetically pleasing reverberation, whereas architectural acoustics benefits from very accurate acoustics simulations, which often carry a heavy computational cost.

The emergence of immersive applications, such as virtual and augmented reality, creates a specific set of requirements [5, 6]. Due to the limited computing resources on these platforms, the sound reproduction methods used need to be efficient and highly adaptive to the movement of a listener and a sound source. In augmented reality, where real sounds overlap with virtual ones, the reverberation needs to be compatible with other spatial sound reproduction methods to provide accurate and perceptually transparent reverberation [6].

This dissertation focuses on the directional properties of late reverberation in the context of immersive sound reproduction. More specifically, an objective and subjective analysis framework is introduced to study the directional characteristics observed in late reverberant sound fields and evaluate their importance in the context of sound reproduction (Publication V). Artificial reverberation algorithms, capable of producing directional decay characteristics, are also presented (Publication VI, VII). In this work,
directional reverberation refers to reverberant sound fields characteristics, such as the decay rate, that are not uniform in all directions. Novel use of sparse noise sequences are also proposed in the context of artificial reverberation and the decorrelation of audio signals.

When developing artificial reverberation algorithms, different objectives may be realized [4]. Here, we describe the main purposes of reverberation methods along three main axes (Figure 1.1). Some algorithms focus more on the aesthetic qualities, in which case the reverberation is controlled through a set of parameters to adjust the reverberation. Some methods aim to accurately simulate the acoustics of a room [7]. Such methods may reproduce the fundamental behavior of waves propagating or approximate this process by using key physical properties of a room, such as its size and the materials it contains.

Perception is another goal which may define the characteristics of a reverberator aiming to produce realistic sounding reverberation based on perceptual characteristics, such as the decay rate of different frequencies [8]. These methods offer some flexibility and may be informed by physical properties to estimate their perceptual characteristics [9], but they may also be tuned aesthetically. The artificial reverberation methods considered in this work are mainly from this last category.

The methods presented in this dissertation have two main objectives. The first is to propose highly efficient algorithms, suitable for use in interactive applications with limited computing resources. The second objective is to analyze and reproduce key perceptual aspects of a reverberant sound field, namely, how the energy is distributed spatially during the reverberation. These two objectives allow the realisation of reverberation algorithms suitable for applications in virtual and augmented reality, which requires an efficient six-degrees-of-freedom spatial sound reproduction [10].

Throughout this dissertation, the figures illustrating key measures of a reverberant sound field were produced using a spatial impulse response (SIR) recorded using a fourth-order Ambisonics microphone in the facilities of the Acoustics Lab at Aalto University. The captured room is a reverber-
ant chamber with variable acoustics called Arni, but at the moment of the capture, in May 2019, it was being renovated. For this reason, it had exposed concrete walls and no furniture in the room. This provided a unique opportunity to capture the acoustics of a shoe-box room with no absorption or scattering beyond its rigid painted walls. The room has the following dimensions: 6.17 m wide, 8.69 m long, and 3.6 m high (Figure 1.2). The same data is used in Section 5.5.2 to reproduce the directional properties of the decay using an existing SIR.

The introductory part of this dissertation is structured around four main topics: the physical and statistical properties of sound propagation, the decorrelation of audio signals, artificial reverberation, and directional reverberation. Chapter 2 introduces some of the necessary background theory on sound propagation to discuss the nature of reverberant and diffuse sound fields. Statistical measures are reviewed as well as the analysis of captured impulse responses and spatial audio reproduction methods.

In Chapter 3, we discuss the decorrelation of audio signals in the context of sound reproduction. A brief overview of measures and existing decorrelation methods is given along with contributions made to this topic using short sequences of sparse noise called velvet noise (Publication I, II).

In Chapter 4, we give an overview of the history of artificial reverberation methods used to reproduce acoustics. Virtual acoustics and noise-based reverberator are covered, and an extensive review of delay-network methods is given. The chapter also includes two novel uses of velvet noise in the context of artificial reverberation (Publication III, IV).

Chapter 5 is focused on hybrid reverberation algorithms and their use for spatial sound reproduction. This chapter also highlights key contributions made in this dissertation, which includes an analysis framework for the evaluation of directional characteristics in a reverberant sound field (Publication V) and two reverberation algorithms capable of reproducing these decay characteristics (Publication VI, VII). Finally, the work of this dissertation is summarized and future research directions are explored in the conclusion.
2. Reverberant Sound Field

When a sound source emits energy in an enclosed space, the energy radiates outwards and propagates in the surrounding environment. At any given time, if we sample the sound pressure at different points in a room, we will find signals at various phases carrying different amounts of energy, and together these pressure points form a sound field [11]. As such, a sound field may be described as a portion of space with acoustical disturbances. In spatial sound reproduction, a sound field commonly refers to a set of points captured by a spherical microphone array (SMA).

Propagating waves eventually interact with the various surfaces contained in a room, and these surfaces absorb and scatter the energy when the waves are incident on it. This interaction quickly increases the density of waves contained in the sound field while decreasing the overall amount of energy, thus forming a reverberant sound field. The propagation of sound in an enclosed space can also be separated into two distinct time periods, the early part, where reflections are individually salient, and the late part, where the sound field contains a high density of reflections that are better described using statistical methods.

In this chapter, the fundamental principles of sound propagation are reviewed along with the essential statistical properties used in the study of room acoustics. The concept of diffuse fields, the implications they carry, and the measures to define them are also examined. Finally, methods used for the reproduction of sound fields are discussed.

2.1 Sound Propagation

The propagation of sound energy through mechanical waves, which includes longitudinal and transverse waves, follows the general laws of thermodynamics to transmit energy from particle to particle in a medium. In free space, an omnidirectional sound source emits a wave expanding spherically and causes the displacement of air particles.

However, this displacement is susceptible to a varied range of interfer-
ence such as the diffraction of waves around obstacles, the absorption of energy when reflecting off materials, the scattering of waves in multiple directions when reflecting off irregular surfaces.

In practice, many waves propagate simultaneously, and the cumulative effect produced by the displacements they cause makes it very challenging to assess acoustic phenomena as a whole. For this reason, to study the acoustics of rooms, it is useful to break down this complex system into smaller parts that can be observed in controlled conditions.

The first step in describing the behavior of waves in a three-dimensional space, is to use the basic equations to describe the propagation. A fundamental equation, known as the wave equation, describes the propagation of sound in free field, formulated using the following partial differential equation [11]:

\[
\nabla^2 p = \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2},
\]

(2.1)

where \( c \) is the speed of sound, \( \nabla^2 \) corresponds the second derivative of the spatial components called the Laplace operator, and \( p \) is the sound pressure.

In the frequency domain, waves are described through a linear partial differential equation, independent of time, called the Helmholtz equation [11]

\[
\nabla^2 \hat{p} = -k^2 \hat{p},
\]

(2.2)

for a given wavenumber \( k \).

In an open space, an ideal omnidirectional sound source emits vibrations that expand spherically, spreading energy outwards. The wave equation for the resulting spherical wave is given by [12]

\[
\frac{\partial^2 p}{\partial r^2} + \frac{2}{r} \frac{\partial p}{\partial r} = \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2},
\]

(2.3)

where \( r \) is the radius of the spherical wave and \( t \) is time, which may also be written as [2]

\[
\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2} = \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2},
\]

(2.4)

where \( x \), \( y \), and \( z \) are the Cartesian coordinates.

When the radius is large enough with respect to the wavelength, the observed curvature of the wavefront is almost flat. For this reason, the propagation of sounds is best described through a set of plane waves traveling in three-dimensional space.

On their own, these equations only represent waves traveling in a free field. In room acoustics, boundaries must be introduced into these wave equations to account for sound absorption [13].
In an enclosed space, a wave will eventually reach a wall, which will absorb some of the energy and reflect the rest. The magnitude of energy and phase of the reflected wave is determined by the impedance of the wall $Z$, a complex number, and the characteristic impedance of air $\rho_0$. For a flat surface of infinite dimensions, the reflection factor corresponds to the fraction of reflected energy and is defined as [2]

$$R(\theta) = \frac{Z \cos \theta - \rho_0 c}{Z \cos \theta + \rho_0 c}, \quad (2.5)$$

where $\theta$ is the angle between the normal vector of the wall and the incident wave.

Here, the reflection factor $R(\theta)$, also a complex number, has a real value constrained between $[-1, 1]$, based on the incident angle ($0 < \theta < \frac{\pi}{2}$) and the wall impedance. If the wave vector is perpendicular to the wall ($\theta = 0$), Equation 2.5 becomes

$$R(0) = \frac{Z - \rho_0 c}{Z + \rho_0 c}, \quad (2.6)$$

and at a theoretical 90° angle, parallel to the wall, the equation becomes

$$R(\frac{\pi}{2}) = -1, \quad (2.7)$$

hence, when $\rho_0 c < Z$, the energy being reflected $|R(\theta)|$ increases with the angle. Using the reflection factor, the absorption at the wall boundary is

$$\alpha_b(\theta) = 1 - |R(\theta)|^2, \quad (2.8)$$

which illustrates that the energy absorbed during the reflection of single plane wave is dependent on the incident angle of the wave.
However, if the surface of the wall is not regular, the reflection angle is dependent on the wavelength of a sound wave. If we consider a discontinuity of length $L_d$ and an incident wave with wavelength $\lambda > 2L_d$, the discontinuity will not substantially influence the wave. However, a wave having a smaller wavelength $\lambda < 2L_d$ will be reflected at the angle of the discontinuity, and a wave with $\lambda \approx 2L_d$ will be scattered in multiple directions [15].

In practice, a very irregular surface will cause a wave with different wavelengths to reflect in many directions. As such, the reflection of sound energy in a room is both angle- and frequency-dependent. The scattering of waves promotes the diffusion of sound energy in a room throughout the propagation.

Near a flat surface, an incoming plane wave interacts with its reflected wave in the specific region where the two overlap, illustrated in Figure 2.1. In a room with many flat surfaces, the overlapping of waves from all directions causes an increase of potential energy in the reverberant sound field in areas further away from the center [14].

In a rectangular room, when the distance between two walls is a multiple of a wavelength being propagated, the reflected energy combines either constructively or destructively to introduce stable energy patterns, called standing waves or room modes. This also occurs when higher-order reflection paths, combining different walls, form more modes. As such, the sound pressure in three-dimensional space can also be described as a linear
combination of eigenmodes.

Individual eigenmodes are described through an eigenfunction for a given mode number \((n_x, n_y, n_z)\). In a rectangular room with rigid boundaries, the eigenfunctions are calculated using [2]

\[
p_{n_x n_y n_z}(x, y, z) = \cos \left( \frac{n_x \pi x}{L_x} \right) \cos \left( \frac{n_y \pi y}{L_y} \right) \cos \left( \frac{n_z \pi z}{L_z} \right),
\]

where \((L_x, L_y, L_z)\) are the dimensions of the room, and \((x, y, z)\) are coordinates along these dimensions. The wavenumber \(k\) of a given mode and its associated frequency \(f\) can be calculated from [2]

\[
k_{n_x n_y n_z} = \frac{\pi}{L_x} \sqrt{\left(\frac{n_x}{L_x}\right)^2 + \left(\frac{n_y}{L_y}\right)^2 + \left(\frac{n_z}{L_z}\right)^2}, \tag{2.10}
\]

and

\[
f_{n_x n_y n_z} = \frac{c}{2 \pi} k_{n_x n_y n_z}, \tag{2.11}
\]

respectively. Therefore, higher modes resonate at higher frequencies. In Figure 2.2, low-order two-dimensional eigenfunctions are illustrated using the dimensions of the shoe-box room described in Chapter 1. The areas with larger \(|p|\) values represent a larger potential energy. Blue and red represent, respectively, a negative and positive amplitude.

Figure 2.3 illustrate higher-order modes. To activate a mode in a room, a wave oscillating at the mode frequency needs to be introduced. A reverberant sound field comprises of a superposition of all the activated modes.

### 2.2 Statistical Properties

As discussed in the previous section, the reflection of an individual plane wave is determined by its frequency, its incidence angle, as well as the impedance and shape of the reflective surface. In a reverberant sound field, many reflected waves combine together to form a dense distribution of plane waves.

Having many circulating plane waves from many directions means that the energy becomes more uniformly distributed as time progresses. Hence, it is more practical to leverage the stochastic nature of reverberant sound fields rather than attempting to follow the cumulative impact of individual reflected plane waves. This section contains an overview of the statistical properties of a reverberant sound field used to describe the acoustic properties of a room.

In the early stage of reverberation, individual echoes are sparse in time. For this reason, early reflections are particularly critical in the perception of acoustics [1]. However, in the later stage of sound propagation, the near infinite number of reflected plane waves circulating implies a high density
Figure 2.3. Normalized higher-order two-dimensional eigenmodes ($n_z = 0$).
of echoes. In a room, the number of echoes per second for a given time \( t \) is estimated as \[16\]

\[
\rho_e(t) = \frac{4\pi c^3}{V} t^2,
\]  

(2.12)

where \( V \) is the volume of the room. In artificial reverberation, the echo density is an important consideration and must rise rapidly to ensure a natural sounding reverberation \[17\].

In the frequency domain, the reverberant sound field is perceived as a superposition of room modes. The density of superposing modes increases with the frequency. The modal density for a frequency \( f \) can be estimated using \[16, 2\]

\[
\rho_m(f) = \frac{4\pi V}{c^3} f^2,
\]  

(2.13)

for a frequency \( f \) given in hertz. Schroeder suggested that above a certain frequency \( f_{\text{Schroeder}} \), \( \rho_m(f) \) is sufficiently high that the individual mode becomes perceptually indistinguishable \[18, 19, 20\]. The cross-over frequency where this transition occurs is estimated from the volume and decay time through \[18\]

\[
f_{\text{Schroeder}} \approx 2000 \sqrt{\frac{T_{60}}{V}},
\]  

(2.14)

where \( T_{60} \), a measure of reverberation, is given in seconds and \( V \) is the volume in cubic meters. The \( T_{60} \), called reverberation time, represents the time required for the sound energy to decay by 60 dB. \( T_{60} \) may be estimated based on the dimensions of the room and the absorbing properties on its surfaces. Sabine formulated the reverberation time as \[21\]

\[
T_{60,\text{Sabine}} = 0.163 \frac{V}{S\bar{\alpha}},
\]  

(2.15)

where \( V \) is the volume of the space in m\(^3\), \( S \) is the total surface area, and \( \bar{\alpha} \) is the average absorption per m\(^2\). Here, the absorption considered is also known as the random-incidence absorption coefficient \[22\] and is based on statistical measurements performed in a reverberant room. The term, 0.163, from this equation is derived from \((24 \ln(10)/c)\) \[2\].

This decay time estimation formula was later revised after it was discovered that rooms with highly absorptive materials were underestimated. The revised formula given by Eyring is \[23\]

\[
T_{60,\text{Eyring}} = 0.163 \frac{V}{S \ln (1 - \bar{\alpha})},
\]  

(2.16)

which introduces extra terms when solving the natural logarithm using a
Taylor-series expansion defined by
\[ \ln (1 - \alpha) = \alpha + \frac{\alpha^2}{2} + \frac{\alpha^3}{3} + \frac{\alpha^4}{4} + \cdots, \] (2.17)
whereas \( T_{60}^{\text{Sabine}} \) only considers \( \alpha \) in the denominator.

In modern versions of \( T_{60}^{\text{Sabine}} \) and \( T_{60}^{\text{Eyring}} \), an extra term is also added to the denominator (+4V\( \alpha_a \)), where \( \alpha_a \) represents air absorption. Air absorption becomes an important factor to consider in bigger rooms [2].

Both \( T_{60}^{\text{Sabine}} \) and \( T_{60}^{\text{Eyring}} \) make the assumption that the absorption is evenly distributed in the space, which can lead to an estimation error in certain rooms [24, 25, 26, 27]. Other methods account for a less uniform distribution of absorption. For instance, the Fitzroy formulation groups parallel pairs of walls to give [28]

\[ T_{60}^{\text{Fitzroy}} = \frac{0.163 V}{S} \left[ \frac{S_x}{-S \ln(1 - \alpha_x)} \right] + \left[ \frac{S_y}{-S \ln(1 - \alpha_y)} \right] + \left[ \frac{S_z}{-S \ln(1 - \alpha_z)} \right], \] (2.18)

where \( S_x \) represents the total area of the side walls, \( S_y \) the area of the floor and ceiling, and \( S_z \) that of the front and back walls, while \( \alpha_x, \alpha_y, \) and \( \alpha_z \) are their respective average absorption coefficients.

When considering the perception of late reverberation in a room, an important parameter to consider is the critical radius \( r_c \). This radius represents the distance at which the direct sound energy is equivalent to the reverberant energy. This measure has an implication, for example, on speech intelligibility for example. The critical radius can be calculated using [29]

\[ r_c = \sqrt{\frac{S_\alpha}{V}}, \] (2.19)

where \( S_\alpha \) is called the equivalent absorptive area, a value representing the total absorption in a room calculated using [29]

\[ S_\alpha = \sum_i S_i \alpha_i, \] (2.20)

where \( S_i \) is the surface area of the room and \( \alpha_i \) is the corresponding absorption coefficient.

For the statistical analysis of a room, it is also possible to capture an impulse response (IR) by playing a short transient sound or a sine sweep through a loudspeaker placed in the room [30, 31]. A broad-spectrum stimulus will ensure more room modes will resonate, yielding a more accurate frequency response of the reverberant sound field. In the time domain, and for a single microphone-loudspeaker pair, we define the IR as

\[ h(t) = \sum_{k=1}^{\infty} a_k \delta(t - \tau_k) + h_{\text{noise}}(t), \] (2.21)
where $a_k$ is the amplitude of an individual reflection, $\tau_k$ is the corresponding delay, and $\delta$ is a Dirac delta function. $h_{\text{noise}}(t)$ represents a noise sequence introduced when capturing an IR. Figure 2.4 illustrates $h(t)$ for the example captured room, and in Figure 2.4(b) individual reflections are shown.

To analyze the behavior in the frequency domain, a fast Fourier transform (FFT) is performed on the captured IR, resulting in the frequency response in Figure 2.5. In Figure 2.5(b), individual low-frequency modes are shown. The individual modes identified in Figure 2.5(b) correspond to the modes predicted using the dimensions of the shoe-box room in Equation 2.9 and illustrated in Figure 2.3.

As illustrated in Figure 2.4, the density of impulses tends to be much higher during the late reverberation stage. In an IR, the late part tends to look very similar to a random noise sequence. As such, the echo density of a captured IR may be estimated from how much it deviates from the distribution of Gaussian noise, representing ideally dense and stochastic late reverberation [32, 33].

To analyze the properties of the decay, the energy-decay curve (EDC) [34], an integral of energy from a moment in time $t$ until $\infty$, is defined as

$$
EDC(t) = \int_t^\infty h^2(\tau) \, d\tau, 
$$

which is expanded as a time-frequency surface using the energy-decay
relief (EDR) \[35\]

\[
EDR(t, \omega) = \int_{t}^{\infty} h^2(\tau, \omega) \, d\tau,
\]

where individual curves are calculated for different frequency bands \(\omega\).

In Figure 2.6, the EDR of the captured shoe-box room is analyzed. From this figure, we observe that the decay properties are frequency-dependent and also that the presence of noise hides a portion of the exponential decays. Hence, when recording an IR, the noise captured from the room and the recording equipment hides some of the decay occurring below the noise threshold. The decay rates contained in these curves are suitable to estimate frequency-dependent reverberation times \([T_{60}(\omega)]\), but in noisy IR, the reverberation time may be estimated from shorter segments of exponential decay \([T_{10}, T_{20}, T_{30}, \ldots] [12]\).

Since the EDC and EDR formulas integrate the noise energy contained in an IR, further considerations are necessary to interpret the data beyond the noise floor. In \[36\], different strategies to minimize the impact of noise on EDC analysis are explored. The decay rates in the beginning of the IR for different frequencies \[37\] and directions \[38\] may also be used to replace the noisy part of SIIR with decaying noise, which is beneficial in sound reproduction using captured IRs.
2.3 Diffuse Field

A diffuse sound field, or diffuse field for short, is a theoretical sound field in which the energy is ideally distributed. This specific state may arise if the reflection of sound energy in a room creates a near infinite amount of uncorrelated plane waves and reaches a statistically uniform distribution. The precise definition of a diffuse field varies throughout the literature, but it is commonly defined through three key preconditions: homogeneity, isotropy, and incoherence [14, 2, 39].

Homogeneity refers to a state where the mean energy density is the same at all locations in a space. In this state, the energy is perfectly uniform at every point in the sound field.

Isotropy implies that the energy coming from all directions for a given point is equivalent. From a statistical standpoint, this means that the probability for a wave to be incident from a specific direction is the same in all directions. From the homogeneity principle, a diffuse field is also isotropic at all locations. A sound field which is not isotropic is called anisotropic.

Finally, incoherence means that individual wavefronts have only a small amount of correlation between them. This implies that the phase of different frequency components is distributed randomly between reflected waves. This precondition is important in order to avoid constructive and destructive interference that arises from highly correlated sound waves, which in turn, would break the homogeneity and isotropy.

When all three preconditions are met, we obtain a statistically diffuse state. A diffuse field is useful, for instance, for measurements of absorption materials in a reverberant room. However, building a room with sufficient diffusion to create such a diffuse field requires special care [40, 41, 42].
evaluate the characteristics of reverberant rooms, recent work employs SMAs to analyze and validate the isotropy in a room [43, 44, 45, 46, 47].

In practice, these ideal conditions are never fully realized. Physical attributes such as parallel walls, uneven distribution of absorbing materials, and the presence of non-scattering surfaces can all impact negatively on the diffusion of energy in a room [48, 14, 40, 2]. Nonetheless, in some applications, the diffuse-field conditions are not required to be fully met. In such applications, any reverberant sound field is assumed to be sufficiently diffuse, which allows sufficient simplifications for practical tasks such as speech enhancement and de-reverberation [49].

In [14, 50], Waterhouse demonstrates how a rigid wall creates interference patterns in the sound field, which leads to up to 6 dB differences in the potential energy between the center of the room and its corners [14]. For this reason, it is recommended to measure the absorption coefficient of materials at the center of a reverberant chamber. Using simulated acoustics, the distribution of incident directions of individual reflections may be analyzed with good precision to demonstrate the anisotropy in simple rooms, such as a shoe box room [51, 52]. In [53], the direction-of-arrival of late reflections in a concert hall were analyzed using captured and simulated impulse responses, and the study determined that the direction of late reverberation is influenced by both the shape of a room and the distribution of absorption material in a room.

Recently, new measures were proposed to quantify the amount of anisotropy in a captured spherical sound field by measuring the difference between different spherical harmonics components [44, 45, 46]. In [54], a perceptual study demonstrated that the differences in the energy decay profiles can be perceived above a certain threshold.

### 2.4 Properties of Reverberant Sound Fields

A diffuse field carries important implications and as such, it is useful to determine whether a sound field is diffuse or not. In this section, we briefly review two objective measures and key subjective studies to characterize the properties of a reverberant sound field in the context of sound reproduction.

#### 2.4.1 Objective Measures

Diffuseness is a common term used for various measures estimating whether the conditions for a diffuse field are satisfied [55]. However, only one characteristic of a diffuse field is usually evaluated, and for this reason, the choice of a diffuseness measure is determined by some assumptions on the sound field. For instance, some definitions use the acoustic energy
Reverberant Sound Field

[56, 57], while others define diffuseness using acoustic intensity [2, 58]. These measures determine that a sound field is diffuse by assessing its isotropy.

In another approach, the covariance of a sound field is analyzed to measure the coherence between spherical harmonics (SHs) [55]. In the context of this dissertation, a key benefit of using spatial coherence to rate diffuseness is that it makes no assumptions on the isotropy of a reverberant sound field [38].

Generally, it is considered that a reverberant sound field transitions from a highly directional energy distribution during the early stage of reverberation to a more isotropic and incoherent distribution during the later stage. The moment of this transition, called mixing time $t_{\text{mix}}$, is of particular interest when studying sounds fields, since different assumptions are made before and after this moment. This is similar to $f_{\text{Schroeder}}$ in Equation 2.14, which defines the boundary of salient room modes.

One definition of $t_{\text{mix}}$ looks at the echo density and the moment where $\rho_e(t)$ (Equation 2.12) reaches a perceptual threshold where individual reflections are indistinguishable [59]. One limitation of this approach is that while individual reflections may not be perceptually salient, an anisotropic sound field may still maintain more overall energy in specific directions throughout the decay, resulting in a different perceptual cue.

In the context of spatial sound reproduction, another approach is to define $t_{\text{mix}}$ as the moment where a specific threshold of diffuseness is met [60]. This allows the use of a covariance analysis method which is suitable for the study of sound fields that may be incoherent but anisotropic (Section 3.2.4).

2.4.2 Perception

Whereas analysis method such as diffuseness and mixing time may be important as objective measures of a reverberant sound field, subjective evaluation is essential to determine the appropriate level of details required in the design of reproduction methods.

The importance of early reflections on the perception have long been demonstrated [61]. Specifically, they play a key role in the spatial impression of a concert hall, which may be characterized by the ratio of lateral and non-lateral energy in the first 80 ms of sound propagation [62]. This ratio has been found to be a good descriptor of the spatial impression of a concert hall, thus demonstrating the need for good lateral reflections in concert halls [63, 1]. The perceptual impact of the spatial distribution of early reflections was also studied from the perspective of musicians on stage [64].

The importance of the early reflections is also echoed in sound reproduction, with the development of recording techniques preserving their
characteristics in stereo reproduction [65], and more recently their reproduction was assessed in the context of virtual reality and a listening experiment found that an accurate reproduction of the first six reflections was optimal [66].

However, late reverberation also plays an integral role in the perception of acoustics. In [67], the impact of late reverberation on the sense of envelopment was studied, and a perceptual test showed that envelopment decreases a subject’s sensitivity to changes in the apparent source width of the sound field. Many other studies have also demonstrated the positive correlation between envelopment and late reverberation [3, 68], which is also influenced by the distance between a listener and a sound source [69].

In a different perceptual study, the reproduction of late reverberation in binaural recordings was investigated using a lecture hall with a short reverberation time [70]. In the experiment, the late reverberation tail was substituted with another one captured from different source and microphone locations in the same room. The listeners were not able to correctly detect the substitution if it occurred after the first 40 ms in the binaural IR.

Another listening experiment examined the connection between objective and subjective measures of mixing time in a binaural reproduction system [71]. In this experiment, the early part of the reverberation, before the mixing time, was updated to follow the head movement of a participant while the late reverberation remained static. The results of the perceptual study showed that, for the reproduced halls, the sound field was perceptually isotropic after the estimated $t_{\text{mix}}$.

However, other studies have indicated that, in conditions such as double-slope reverberation [72, 73] or if the differences in decay energy for different directions reach a specific threshold [54], a reverberant sound field becomes perceptually anisotropic. This suggests that, since isotropy is determined by the diffusion properties of a room [53], the perception of anisotropy also varies from one room to another. This is explored in Publication V, detailed in Section 5.5.1.

### 2.5 Reproduction of Sound Fields

A reverberant sound field contains many plane waves incident from all directions. In this section, we review Ambisonics methods, which decompose a spatial sound field into plane waves incident from many directions, and binaural reproduction, which considers the physiological properties of a listener’s head to reproduce a sound field through headphones.
2.5.1 Ambisonics

The reproduction of sound fields is crucial for multichannel sound reproduction. Ambisonics, a popular term used to describe a set of methods, originally developed in the 1970s to encode a sound field using an SMA [74, 75, 76, 77, 78]. Ambisonics methods were later generalized to higher order SMAs through the use of SH functions, suitable for the capture and reproduction of sound fields using a higher spatial resolution [79, 80, 81, 82, 83].

In higher order Ambisonics, a sound field is expressed by performing a plane wave expansion to obtain coefficients of a set of orthonormal SH basis functions covering the surface of a sphere. To obtain SH signals, a sound field is first sampled from $K$ individual plane wave signals $s_i$, incident to $(\phi_i, \theta_i)$ in spherical coordinates, and is encoded using

$$s(n) = \sum_{i=1}^{K} s_i(n) y(\phi_i, \theta_i),$$

where $n$ is the index of an audio sample and $y(\phi, \theta)$ corresponds to a vector containing the SHs for a given direction defined as a vector of coefficients

$$y(\phi, \theta) = \begin{pmatrix} Y_1(\phi, \theta) \\ Y_2(\phi, \theta) \\ \vdots \\ Y_Q(\phi, \theta) \end{pmatrix},$$

for a set of $Q = (L + 1)^2$ coefficients and a maximum order $L$. The SH coefficients in vector $y(\phi, \theta)$ are placed in a pre-determined order, following one of the existing conventions to ensure the portability of Ambisonics signals during sound reproduction [84].

In the Ambisonics channel number (ACN) convention, the index number corresponds to $i = l^2 + l + m + 1$ where $l$ and $m$ are the SH order and degree, respectively, with $0 \leq l \leq L$ and $-l \leq m \leq l$ [84]. Individual SH coefficients are commonly defined as [85]

$$Y_{lm}(\phi, \theta) = \sqrt{\frac{2 - \delta_{m0}}{4\pi}} \frac{(2l + 1)(l - |m|)!}{(l + |m|)!} \left( y_m(\phi) P_{l|m|}(\cos \theta) \right),$$

where $\delta$ is the Kronecker delta, $P_{l|m|}$ are the associated Legendre polynomials, and $y_m$ is

$$y_m(\phi) = \begin{cases} 
\sin |m|\phi & m < 0, \\
1 & m = 0, \\
\cos m\phi & m > 0.
\end{cases}$$
Once encoded, an Ambisonics signal is converted back to individual plane waves through

\[ \tilde{s}(n) = Y^T s(n), \]  

where \( \tilde{s}(n) \) contains a set of plane wave signals

\[ \tilde{s}(n) = \begin{pmatrix} \tilde{s}_1(n) \\ \tilde{s}_2(n) \\ \vdots \\ \tilde{s}_K(n) \end{pmatrix}, \]  

and \( Y \) is a matrix defined as

\[ Y = \begin{pmatrix} y(\phi_1, \theta_1) \\ y(\phi_2, \theta_2) \\ \cdots \\ y(\phi_K, \theta_K) \end{pmatrix}, \]  

which performs a plane wave decomposition in the direction of each SH vector \( y(\phi_i, \theta_i) \).

For sound reproduction, an Ambisonics signal is decoded to a set of channels by using a beamformer pointing in the direction of each loudspeaker. A beamformer is defined as

\[ y(n, \theta, \phi) = \sum_{i=1}^{Q} s_i(n) w_i(\theta, \phi), \]  

where \( w_i \) contains the individual SH weights for a given direction. A simple hypercardoid beamformer is obtained when \( w_i(\theta, \phi) = Y_i(\theta, \phi) \). The shape of a beamformer can be modified to modify the shape of the main lobe as well as the side and back lobes [82, 86]. To analyze and reproduce spherical reverberant sound fields, a SIR is captured using an Ambisonics microphone and decoded to a set of directional impulse responses (DIR) for reproduction.

In certain applications, it is necessary to manipulate an Ambisonics signal [87, 88, 89, 90]. Specifically, we may want to alter the energy of an Ambisonics signal in different directions [91, 92]. If we consider \( G = \text{diag}(g(\phi_1, \theta_1), g(\phi_2, \theta_2), \ldots, g(\phi_K, \theta_K)) \), a diagonal matrix containing a set of directional gain values, we formulate the transformation matrix \( T \) as

\[ T = Y' G Y^T, \]  

where \( Y' \) is the output SH matrix and \( T \) is a matrix which is used to apply energy weighting to an Ambisonics signal. Depending on the smoothness of the distributed gains in \( G \), spatial aliasing may occur, which is minimized by increasing the order of the encoding matrix \( Y' \) [91, 92].
2.5.2 Binaural Reproduction

Due to the reflection and diffraction of sound waves around the human head and the ears’ pinnae, the perceived spectrum of sounds reaching our ears is altered based on their incident angles. Using binaural sound-reproduction techniques [93], a spherical sound field is encoded into a stereo signal by using a set of head-related impulse responses (HRIR) or head-related transfer functions (HRTF) in the frequency domain, which will reproduce the interaural time difference (ITD) [94, 95, 96], interaural level difference (ILD) [94, 95], and the spectral cues of sound reaching our ears at different angles [93].

In the context of a reverberant sound field, a key characteristic to consider in binaural sound reproduction is the perceived interaural cross-correlation (IACC). A formal definition of cross-correlation is given in Section 3. The IACC corresponds to the frequency-dependent cross-correlation between the two ears, defined as [97, 98]

\[
\gamma_{\text{IACC}} = \frac{\int_{t_0}^{t_1} x_L(t) x_R(t + \tau) \, dt}{\sqrt{\int_{t_0}^{t_1} x_L(t)^2 \, dt \int_{t_0}^{t_1} x_R(t)^2 \, dt}}, \tag{2.33}
\]

where \(x_L\) and \(x_R\) are the two channels of a binaural signal and \(\tau\) represent a time lag. In a diffuse field, the IACC approximates a sinc function along the frequency axis [99]. However, in practice, the shape of a listener’s head will greatly impact the IACC in a diffuse field [93].
3. Decorrelation

Since one of the primary characteristics of reverberant sound field is the low-coherence between individual plane waves, the decorrelation of audio signals relates to the natural process of sound diffusion occurring during sound propagation [2]. For this reason, a multichannel reverberator should act as a decorrelator (Section 5.1). However, in some applications, it is useful to control the decorrelation independently from the reverberation, which requires the use of decorrelation methods that do not introduce reverberation.

In this context, cross-correlation refers to a measure of the similarities between two signals. Coherence is a frequency-dependent measure of these similarities, and covariance is used in the context of multichannel sound reproduction. The decorrelation of audio signals discussed in this section refers to signal processing methods that are used to modify a signal's waveform slightly, ideally just its phase response, to enhance the perceptual spatial attributes during multichannel reproduction.

3.1 Overview

As seen in Equation 2.5, the phase of a reflected wave is a function of the impedance of a wall material, and uneven surfaces will scatter a wave in different directions depending on its wavelength. As such, decorrelation occurs naturally due to the impact of reflection, scattering, and absorption of sound altering the phase of propagating waves.

In signal processing terms, an effective decorrelator should randomize the group delay of a signal while minimizing the spectral distortions [100]. Furthermore, the modification of the phase should be restricted in order not to exceed the frequency-dependent perceptual threshold of phase shifting [101] to avoid echoes and the smearing of transients.

Decorrelation is used in multichannel sound reproduction to reduce the comb-filtering effect encountered when multiple channels produce the same signal and recombine slightly out of phase in a sound field. For
stereo headphones reproduction, highly correlated signals are perceived as coming from inside the listener’s head. For this reason, up-mixing, a common application of decorrelation, is used to convert between different channel-based configurations during sound reproduction [93].

In a sound field, some signals should be highly correlated, such as the direct sound and early reflections, whereas late reverberation must be well decorrelated. For this reason, when decorrelation is performed on a continuous signal encoded in a multichannel format, such as Ambisonics, an analysis step is required to separate the diffuse part of the signal from the non-diffuse part, and the decorrelation is applied only to the diffuse part. This approach is used to improve the reproduction of Ambisonics [58, 102, 103].

Another use for decorrelation is to extend the perceived size of a sound source when mono signals are reproduced on a loudspeaker array. If the same signal is used on multiple loudspeakers, the incident direction of the source is perceived from a small area between the loudspeakers [100, 104]. On the other hand, this attribute is beneficial in panning techniques, such as the vector-base amplitude panning [105].

3.2 Objective Evaluation

To evaluate the correlation of audio signals, several measures are used depending on if we are interested in the correlation of signals over time, frequency, or space.

3.2.1 Cross-Correlation

The amount of correlation between two signals $x$ and $y$ over time is measured through the cross-correlation, calculated as

$$\gamma_{xy}(l) = \sum_{n=1}^{N} x(n + l)y(n), \quad (3.1)$$

where $l$ is the lag. As such, two out-of-phase signals may still be highly correlated if, for instance, one is a delayed copy of the other. When evaluating decorrelation, the zero-lag cross-correlation $\gamma_{xy}(0)$ is of particular interest to measure the correlation at a specific moment. The normalized zero-lag cross-correlation is given by

$$\gamma_{xy} = \frac{N}{\sqrt{\sum_{n=1}^{N} x(n)^2 \sum_{n=1}^{N} y(n)^2}} \sum_{n=1}^{N} x(n)y(n), \quad (3.2)$$

which is bounded between $[0, 1]$. A lower $\gamma_{xy}$ value indicates dissimilarities
between the two signals.

### 3.2.2 Coherence

Coherence is a frequency-dependent measure of cross-correlation. In practice, \( x(n, \omega_c) \) and \( y(n, \omega_c) \), the discretized signals given \( \omega_c \), the center frequency for a frequency band in a filterbank. Zero-lag coherence is defined as

\[
\gamma_{xy}(\omega_c) = \frac{\sum_{n=1}^{N} x(n, \omega_c)y(n, \omega_c)}{\sqrt{\sum_{n=1}^{N} x(n, \omega_c)^2 \sum_{n=1}^{N} y(n, \omega_c)^2}}.
\] (3.3)

### 3.2.3 Covariance Matrix

Covariance is another related statistical measure. In the context of multichannel sound reproduction, covariance is commonly used to measure the correlation between pairs of channels. For this purpose, a covariance matrix built from the sample covariance of every pair of channels, is defined as

\[
\Gamma_x = \frac{\sum_{n=1}^{N} x(n)x(n)^T}{N},
\] (3.4)

where vector \( x \) is a signal containing multiple channels and \( N \) is the number samples.

As discussed in Section 2.5.1, in Ambisonics, the beamformer used to extract directional signals from Ambisonics signals contains side and back lobes (Equation 2.31). These lobes and the width of the main lobe introduce correlation between the output signals. As such, the covariance matrix is a useful tool to evaluate various output configurations in the context of Ambisonics sound reproduction.

In Figure 3.1, we evaluate the covariance for different output configurations. Here, the input signal is the SIR of the example shoe-box room (Section 1), encoded in first-order Ambisonics. For each configuration evaluated, \( K \) output channels are uniformly distributed around a sphere. The diagonal line in the normalized matrix is the autocovariance between a channel and itself and will always have a value of 1. In Figure 3.2, the same analysis is performed using the SIR encoded in fourth-order Ambisonics as input. These figures illustrate the the limitation of using low-order Ambisonics signals to output on a high-order multichannel loudspeaker system. In Publication V (Section 5.5.1), the coherence matrix is used to validate the reproduction method used in a perceptual study.
3.2.4 Incoherence Profile

The coherence, or its opposite, incoherence, may be used as a measure of the diffuseness of a reverberant sound field (Section 2.4). As discussed previously, during the propagation of sound, the increase of mutually uncorrelated reflected plane waves in a sound field should lower the coherence. To analyze the evolution of coherence in a spatial sound field, we measure the incoherence profile from a multichannel signal.

For this purpose, we first reformulate the zero-lag cross-correlation using the covariance matrix as

$$\gamma_{ab} = \frac{\Gamma_{ab}}{\sqrt{\Gamma_{aa}\Gamma_{bb}}},$$  \hspace{1cm} (3.5)

where $a, b$ are two different channels. The coherence between channels $\gamma_{ab}$ is then averaged for the different pairs of channels using to obtains a coherence profile $\gamma_{\pi}$.

Finally, we set the incoherence profile to $1 - \gamma_{\pi}$ on a short-time moving window to obtain a time-dependent measure that may be used to estimate the evolution of diffuseness in a reverberant sound field.

In Publication V (Section 5.5.1), a measure of incoherence is used to estimate the mixing time of captured SIRs. A more formal measure of spatial coherence is given in [55].
3.3 Decorrelation Methods

Introduced in the context of artificial reverberation [16], the all-pass filter (AP) has good properties for the decorrelation of signals [100]. The AP consists of combining a negative feed-forward path to a positive feedback path of a delay-line using the same absolute attenuation gain $|g|$ for both paths (Figure 3.3). The combination of the two paths creates a zero and a pole, mirrored on opposite sides of the unit circle, which yields a flat magnitude response.

The transfer function of an AP filter is [16]

$$H(\omega) = e^{-j\omega \tau} \frac{1 - ge^{j\omega \tau}}{1 - ge^{-j\omega \tau}}, \quad (3.6)$$

where $\omega$ is the angular frequency, $g$ is the attenuation, $\tau$ is the delay time, and $j$ is the imaginary unit. From this equation, we obtain the magnitude response for the filter $|H(\omega)| = 1$, implying that the input and output signal energy is the same at all frequencies. What is caused by the filter is the frequency-dependent phase lag

$$\phi_l(\omega) = \omega \tau + 2 \arctan \frac{g \sin \omega \tau}{1 - g \cos \omega \tau}, \quad (3.7)$$

which results in an uneven distribution of phase shifts.

Since the AP filter modifies the phase of a signal without affecting its magnitude spectrum, it is a useful tool for decorrelation. However, to ensure sufficient decorrelation, multiple AP filters are cascaded, and their delay components should be kept short to avoid smearing of transients [100]. In larger sets of cascaded APs, the smearing of transients, causing a chirp-like effect, may be leveraged as a sound synthesis method [106], or individual AP characteristics may be optimized to obtain better decorrelating properties [107, 108].

Sequences of random white noise also have a flat spectrum and random phase, which have been shown to provide good decorrelation when used as a filter [100]. In this approach, an input signal is convolved with a short sequence of decaying white noise. Here again, the length of the noise sequence should be kept relatively short to avoid undesired artefacts on transients and the introduction of a reverberation effect. The challenge of decorrelating signals containing transients has led to the development...
of decorrelation methods where the transient elements of signals are separated before applying decorrelation [109].

The Fourier transform of a signal is a complex variable, which can be represented in the polar form using the magnitude and phase at each frequency. As such, this is a convenient domain to perform phase modification without changing the magnitude [110, 111]. This decorrelation method may be performed at minimal computational cost in applications where signals are already processed in the time-frequency domain [110]. Recently, physically-informed decorrelation filters were proposed to account for the specific phase relationship between loudspeakers in a room [112].

3.4 Velvet-Noise Decorrelation

In Publication I (p.99), the use of velvet noise, a special type of random sparse noise introduced in [113], is proposed as a decorrelation method (Fig. 3.4). With properties similar to white noise, velvet-noise sequences are suitable to randomize the phase of a signal without significantly impacting its spectrum. Since velvet noise only contains a few non-zero elements, it has the benefit of reducing the computational cost of the convolution [113].

To form velvet noise, a sequence is divided into windows of fixed length, where each window contains exactly one impulse at a random location. This guarantees a smoother distribution of impulses compared to other types of sparse noise and prevents larger gaps from occurring. Here, the density parameter controls the window length, and the specific location and sign of each non-zero impulses is randomized separately.

A velvet-noise sequence (VNS) is created using a density parameter $\rho$, which determines the length of a windowing function that constrains the locations of individual impulses. Since every window contains exactly
one impulse, the maximum distance between two impulses is $2\rho - 1$. The location of the $m^{th}$ impulse is formulated as [114]

$$v(m) = \left\lfloor \frac{m f_s}{\rho} + r \right\rfloor,$$  (3.8)

where $\lfloor \rfloor$ represents the rounding operation, $f_s$ is the sample rate, and $r$ is a random value constrained to be within a window through [114]

$$r \in \left[ 0, \left(\frac{f_s}{\rho} - 1\right) \right],$$  (3.9)

and a separate random process determines the sign of each impulse. Thus, the sequence contains values that are either $-1$, 0, and 1. In Publication I, several variants of decaying velvet-noise sequences are proposed in order to better preserve transient signals.

While velvet noise has a relatively flat spectrum [114], the short decaying sequences introduced for decorrelation in Publication I are not perfectly flat. For this reason, in Publication II (p.109), the use of an optimisation function is proposed to slightly modify the location of individual impulses and minimize the spectral coloration of a sequence.

Using this optimization method, a large set of sequences is generated from which we find the best pairs for use as decorrelation filters. In Figure 3.5, we see the decaying VNS of the original method, in black, and the sequence resulting from the optimization process, in blue.

Once a set of sequences is obtained, they are further analyzed to rate their overall flatness and to find the pairs with the lowest coherence. As such, a compromise between a flat response and low coherence is required. In Figure 3.6, we see the impact of the optimization process on the coherence.
In Publication II (p.109), two perceptual studies were conducted to evaluate the perceived flatness of the sequences and the performance of the method in the context of stereo up-mixing. Overall, the method has been found to provide decorrelation comparable to white noise while using between 76% and 87% fewer floating-points operations (FLOPS) compared to a segmented FFT-based convolution used for white-noise decorrelation.
4. Artificial Reverberation

The propagation of sound energy in a room produces a large number of echoes with increasingly decaying amplitudes over a short period of time (Equation 2.12). While the early reflections are considered perceptually salient [1], the high density of reflections during the late stage of propagation is perceived as its own complex auditory cue called reverberation.

Methods capable of producing such a reverberation effect are used in a diverse range of applications with various requirements [115, 116, 117, 118, 119]. In music production, for aesthetic reasons, it is desirable to produce ideally dense and rich reverberant sound fields with tuning capabilities that may go beyond what can be experienced in real life [118]. Another application for reverberation is the precise simulation of acoustics during the design phase of new architectural structures [120].

In this chapter, we focus primarily on methods capable of reproducing realistic reverberant sound fields at low computational cost in the context of spatial and immersive sound reproduction, which includes virtual and augmented reality. We begin with a brief historical summary of early reverberation methods. This is followed by an overview of reproduction techniques, which includes capturing room IRs as well as simulating IRs using virtual acoustics methods or artificial noise.

An in-depth review of delay-based reverberators is given in Section 4.5, followed by a detailed survey on how these methods have evolved into complex modern approaches capable of producing multichannel sound fields for real-time interactive applications. Contributions made in this dissertation to the topic of noise-based and delay-based artificial reverberation are given in their respective sections.

4.1 Historical Overview

Artificial reverberation is an audio effect with a rich history that continued to evolve throughout the last century, constantly adapting to the growing capabilities and needs of emerging technologies [4, 121]. Artificial reverber-
Artificial Reverberation can be used to mimic anything from the precise frequency-dependent energy decay of a room to imaginary spaces with ethereal acoustics, thus providing a vast set of tools to creatively control the aesthetic sense of space in an audio production.

Today, more immersive applications such as augmented reality requires a new level of realism from artificial reverberators to allow reproduced sounds to blend with existing spaces. For efficiency, many artificial reverberation techniques focus primarily on the reproduction of the more perceptually salient aspects of room reverberation.

The first artificial reverberation method relied on physical spaces, called reverberant rooms [122, 4]. A reverberant signal was captured by using a loudspeaker and a microphone placed inside the room to playback the dry signal and record the reverberant sound from the room. To obtain an aesthetically pleasing reverberation effect, these rooms had to be carefully designed to promote the diffusion of energy and avoid the fluttering echoes caused by parallel walls. The ratio between of the wet reverberant signal and the original dry signal was adjusted during post-production.

While using a reverberant room can yield a natural-sounding effect, the need for a physical space for this approach has some obvious limitations. For instance, the absorbing materials inside the room need to be altered in order to modify the decay properties of the space.

For this reason, other media such as metallic plates and springs were later used as they provide interesting reverberation effects in a more compact form. Although the plates were still relatively bulky, the property of metals offers unique reverberation properties. This explains why both plates and springs are still used today and why they have been recreated digitally for more practical use [123, 124, 125].

4.2 Captured Impulse Response

As covered in Section 2.1, reverberation can be reduced to an FIR filter containing the response of a specific room (Equation 2.21). So a modern alternative to using reverberant chambers is to capture the IR of an existing room. To record an IR, a simple technique is to generate a very sharp transient sound in the room, such as a balloon pop (Section 2.2).

A captured IR is later used as an FIR filter to convolve with any dry sound [126]. In the time domain, convolution is defined as

\[ y(t) = \sum_{n=1}^{\infty} x(n)h(t - n), \]  

(4.1)

where \( x \) is a dry signal and \( h \) is the impulse response. Obviously, time-domain convolution leads to a prohibitive computational cost as \( h \) becomes longer. A more efficient approach is to perform the convolution in the
Artificial Reverberation

frequency domain, where convolution is defined as

\[ y = \text{IFFT}(\text{FFT}(x)\text{FFT}(h)), \]  

which requires both the input and IR to be first converted to the frequency domain through an FFT, followed by an inverse fast Fourier transform (IFFT). The FFT/IFFT operations introduce latency, but existing methods allow partitioning the convolution into shorter time windows, making the issue of latency negligible in practice [126, 127].

To capture an IR, another approach is to use a frequency sweep as a stimulus in the room. This ensures all modes in the room (Equation 2.10) are activated and captured, preferably with the same amount of energy [30, 31]. The signal recorded using the sweep must first de-convolved to obtain the IR.

The presence of ambient noise during the recording process and the noise introduced by the recording equipment, are carried into the captured IR and perceived as an infinite reverberation tail during reproduction. To limit the impact of noise, it is recommended to capture IRs using a logarithmic sine sweep signal, and to increase the length of the sweep to improve the signal-to-noise ratio (SNR) [31].

Since the late reverberation tail is similar to white noise, it is also possible to synthesize it artificially using Gaussian noise [128]. De-noising techniques require the detection of a noise threshold to determine the optimal part to be replaced and a spectral envelope informed by frequency-dependent decay rates, which may also be applied to SIRs [37, 38].

When an IR is captured, the specific location of the loudspeaker and microphone determines the source-listener location during reproduction. In an interactive application, where both the sound sources and the listener are moving, many source-listener locations need to be captured. However, in a diffuse field, this can be simplified since the late reverberation is uniform in all locations. When the late reverberant sound field is homogeneous, only the early reverberation is location-dependent. For this reason, it is common to combine a static late IR with dynamic early reflections obtained by interpolating the early reverberation from a set of IRs [129] or generated using geometric acoustics [130]. However, for heterogeneous sound fields, dynamic reproduction requires many recording locations and interpolation between the IRs at those points [131, 132].

In multichannel sound reproduction, a microphone array, such as an SMA, is used to record a SIR. During reproduction, a mono sound source is convolved with the captured SIR, and the loudspeaker location during the capture determines the incident direction of the reproduced sound source. For better performance, segmenting a SIR into shorter FIR filters has been proposed, allowing the use of HRTFs with lower spatial resolution for late segments in a binaural reproduction [133], and lower-order encoding in Ambisonics reproduction [134].
Artificial Reverberation

One shortcoming of Ambisonics reproduction occurs when lower order Ambisonics is reproduced on a high-order loudspeaker array, which produces highly correlated output signals. However, as covered in Section 2.3, a key characteristic of late reverberant sound fields is that they produce low coherent signals. To overcome the coherence limitation of Ambisonics, several parametric enhancement methods have been developed.

With the spatial impulse-response-rendering method (SIRR)[135], a first-order Ambisonics SIR is analyzed using short-time windows and a filter bank to find the dominant direction for each time window and frequency band. This yields a set of time- and frequency-dependent sparse IRs that can be used in reproduction. A similar approach, the spatial decomposition method (SDM) [136], uses broadband analysis and extends the analysis to arbitrary microphone arrays. Using parametric analysis methods, the direction of arrival of early reflections can be estimated [137, 138].

In some applications, instead of using a room IR, a continuous signal is recorded for reproduction, implying that both the sound source and the reverberation in the room are captured. Here, to enhance the spatial coherence during the reproduction of recorded Ambisonics signals, an analysis process extracts the reverberant part of the signal and a decorrelation filter (Section 3) is used to minimize its coherence [58]. This approach was recently extended to higher order Ambisonics [139] and more advanced parametric methods [140, 141].

4.3 Virtual Acoustics

Capturing IRs requires access to a physical space, good recording conditions, and potentially many recording points. For these reasons, it may be desirable instead to simulate numerically the propagation of sound in a room. The field of acoustic simulation, called virtual acoustics, is largely divided into two families of methods, geometric and wave-based methods [5].

4.3.1 Geometric Methods

In geometric acoustics [5, 142], the behavior of sound waves are approximated to a set of ideal reflection paths from a sound source to a listener. This offers a realistic rendering of the propagating of high frequencies when abstracting more complex elements such as diffraction [143, 13, 144, 5].

The image source method (ISM) considers each wall as a rigid surface, and the algorithm simulates reflections by using virtual sources that are mirror image positions (Fig. 4.1), from which the time-delay and incident angle of individual reflections are calculated [13]. With the ISM, the maximum reflection order determines the complexity of the approximation
and thus the computational cost.

Ray tracing is another geometry-based method that sends rays throughout a virtual geometry and finds possible reflection paths between a source and a listener [143, 15]. The computational cost is minimized by limiting the number of rays and reflection surfaces used to sample the space.

While neither ISM nor ray tracing inherently simulate the full physical characteristics of sound propagation, they efficiently produce specular reflections simultaneously approximating the stochastic nature of the late reverberation [145]. Geometric methods can be combined with more complex analytic methods to simulate specific aspects of the propagation such as scattering [15] and edge diffraction [146, 147]. Ray tracing can also be generalized into the more efficient beam-tracing method, which takes advantage of the spatial coherence of a geometric scene to expand the rays into a larger convex polyhedral shape that splits into smaller beams when multiple intersecting objects are detected [148, 149, 150, 151]. The radiance transfer method [152], inspired from light-propagation modeling techniques, estimates the properties of a reverberant sound field that are used to inform other reverberation methods [153, 154, 155].

4.3.2 Wave-based Methods

To simulate the physical properties of sound propagation as acoustical waves, wave-based methods perform numerical simulations which approximate the wave-behavior of sound transmitting in a medium [5]. Compared to geometric methods, wave-based simulations are inherently well suited for the simulation of diffraction, which plays an important role in the propagation of low frequencies when a scene contains objects and openings.
One of these methods, called finite-difference time-domain (FDTD) [156], solves the wave equation (Equation 2.1) on a fixed grid of points using finite difference operators. The simulation results in sound pressure values for each point that generates a sound field that evolves over time. However, the discretization of wave propagation on a fixed grid has limitations due to the inherent numerical errors of difference operators, such as staircase effects at boundaries [157] and numerical dispersion in high frequencies [158].

Another approach, the finite-volume time domain, performs a similar task using a set of flux terms to define the particle velocities at the boundaries of a three-dimensional cell volumes [159]. Similarly, finite-element methods commonly operate in the frequency domain to solve the Helmholtz equation (Equation 2.2) [5].

Due to their potential for a high level of accuracy, wave-based methods are suitable for applications, such as architectural acoustics, where accuracy is a priority over efficiency [7]. However, for real-time applications, such as virtual and augmented reality, the computational cost of wave-based methods becomes prohibitive. Nonetheless, it is also possible to perform the simulation offline to generate a dataset that can be used in real-time to inform a parametric reverberation algorithm [66].

4.4 Artificial Noise

During the propagation of sound waves, the density of echoes increases exponentially over time (Equation 2.12), and due to this high density, an IR eventually acquires a distribution of impulses over time similar to stochastic noise (Figure 2.4) [160, 161, 162, 17, 163, 144]. Based on this attribute, artificial reverberation algorithms were developed leveraging the use of random-noise sequences. This section explores this family of methods in more detail.

Random signals, such as Gaussian noise, have a flat frequency spectrum and random phase. To produce a reverberation effect, a frequency-dependent envelope is used to control the decay and create an artificial IR. This approach was used in [128] to replace the segment of a recorded IR containing background noise. In this model, an EDR analysis (Eq. 2.23) is performed to extract a frequency-dependent envelope from an earlier segment of the IR, that is then used to filter Gaussian noise to replace the IR after the noise floor. Just as for regular captured IRs, a convolution operation is used to convolve the input signal with the noise and to create a reverberated signal.

Other types of random noise offer similar characteristics with a lower computational cost. In particular, a category of noise known as sparse noise, containing a large proportion of zero values, have been shown to
Artificial Reverberation

Figure 4.2. Flow diagram of the segmented velvet-noise reverberator proposed in [165]. The coloration filters $H_{i,j}$ are designed to match the magnitude response from a segment of an existing IR. The final output $y$ also goes through a series of cascaded all-pass filters (omitted in this figure).

retain a perceptually flat spectrum on the condition that a sufficient density of non-zero values is maintained to avoid audible artefacts [164]. Since the noise sequence is sparse, the zero elements are ignored to efficiently compute the convolution in the time domain with no latency.

In [164], a recursive sparse pseudo-random FIR filter and a low-pass filter is proposed to replace late reverberation, whereas the early reflections are handled by a parallel signal path using an FIR filter. The pseudo-random signal only contains the values $-1$, $0$, and $1$, and the output is filtered with an exponentially decaying envelope before recirculating again. A density parameter controls the probability of having a non-zero element in a given sample. Finding the optimal density is an important aspect of this type of reverberator since it has a direct impact on the computational cost. An average density of at least 2000 impulses per second has been found necessary for this type of recursive sparse-noise reverberator [164].

In [113], a similar approach was used with a new type of sparse noise called velvet noise. As described in Section 3.4, a velvet-noise sequence is divided into windows of fixed length, each containing one impulse at a random location (Figure 3.4). The segmentation guarantees a more uniform distribution of impulses over time and prevents larger gaps from occurring, a potential issue in fully random sparse noise. This formulation of sparse noise was found to be perceptually smoother than Gaussian noise in a listening experiment, which inspired the name of the method [113]. A more in-depth perceptual study was later conducted to investigate the minimum density of non-zero elements required to obtain perceptually smooth velvet noise, which was found to be around 2000 impulses per second [114].

More complex reverberators using velvet noise were developed later. Several methods are proposed in [166] to make use of multiple noise sequences and prevent the audible periodicity that occurs when using a single recursive noise sequence. In [165], the reverberation is split into
Figure 4.3. Flow diagram of the segmented velvet noise reverberator presented in Publication III. The cascaded filters $\Delta H_i(z)$ are designed to match the difference in coloration between segments. The final output $y$ also goes through a series of cascaded all-pass filters (omitted in this figure).

$N$ segments over time. Each segment is delayed and filtered by a VNS and a coloration filter. The coloration filters $H_i$ are designed to reproduce the spectral envelope from a corresponding segment in an existing IR (Figure 4.2). Due to the filtering, the computational cost remains close to that of a segmented FFT-based convolution. However, the required memory is much smaller. VNSs can also be interleaved to create natural sounding reverberation at an even lower computational cost [167]. Since it was first introduced, velvet noise has also inspired a variety of new applications beyond reverberation, such as decorrelation (Section 3.4), time-stretching [168, 169], and 1-bit sound [170]. Improving on the previous model presented in [165] (Figure 4.2), a new formulation of the segmented velvet-noise reverberation was presented in Publication III (p.119). The reverberator is modified to use a cascaded filter structure, that reduces the required number of impulses and allows further computational saving. The filters are designed as differential filters $\Delta H_{i,j}(z)$ corresponding to the spectral differences between two segments of the segmented impulse response (Figure 4.3). The segmentation is also improved by using low-order linear-prediction analysis to identify the ideal locations of the segment boundaries. A listening test is also performed to compare the recorded IR to an artificial one generated using the method. The results showed that the recorded IR was easily identifiable, especially when the stimuli have sharp transients, but the differences were rated as small on average.

4.5 Delay Networks

Back in the middle of the 20th century, electro-mechanical devices, such as magnetic disks and tapes, were used as delay units in sound reproduction. In such devices, a signal is converted through electro-mechanical means
Artificial Reverberation

and is stored on the delay unit for a determined amount of time before being sent to a loudspeaker. By mixing the delayed signal back to the input, the signal is repeated endlessly, which creates a simple reverberation effect [171].

A common application of these early artificial reverberators was to enhance the perceived stereo image from venues equipped with a loudspeaker system. One of the first delay-based reverberation methods used a single delay unit, acting as a ring buffer, and the output from different points on the delay unit, or taps, were taken to feed different channels (Figure 4.4). The last delay-tap was used with a potentiometer to control the signal recirculating to the input of the delay unit [115].

This technique lowered the coherence between different output channels in the room, an important characteristic of a reverberant sound field (Section 2.3). However, one issue of this design was that the recirculating signal also produced a comb-filtering effect that impacted the magnitude spectrum of the reproduced signal.

As the production cost of these delay units became more accessible, more complex delay-network reverberators were created. An important milestone came with the publication of the first reverberation algorithm offering colorless reverberation [172, 16], where Schroeder introduced the AP filter (see Section 3.3 and Figure 3.3) that adds a feed-forward path to a feed-back comb filter to produce a spectrally-flat delayed signal. However, from Equation 3.7, we see that an AP filter produces an uneven distribution of phase delays across frequencies. Thus, several AP filters in cascade with different delay lengths are recommended to ensure a good modal distribution suitable for reverberation [16].

The combination of AP filters in series provides a good method to create increasingly dense copies of a signal without directly impacting its spectrum, thus creating colorless reverberation. Having a colorless reverberation prototype is a fundamental building block to build an artificial

![Flow diagram of an early multichannel reverberator using a recirculating multi-tap delay [115]. In this figure, the multi-tap delay is illustrated using separated delay lines for illustrative purposes.](image_url)
reverberator where the frequency-dependent attenuation is controlled separately. In his original design, Schroeder added feed-forward comb filters in parallel to introduce frequency-dependent attenuation. The comb filters can be placed either at the beginning or the end of this system due to its linear and time-invariant property. To obtain multi-channel outputs, a mixing matrix is placed at the end of the signal flow to combine different signal paths, which creates output signals with low correlation [17] (Figure 4.5). This reverberation model was the starting point of increasingly complex delay-network reverberation methods that are still being used and studied today.

An important characteristic in artificial reverberation is mixing time. As discussed in Section 2.4, it highlights the moment where the reverberant sound field transitions into a more diffuse state of reverberation. For artificial reverberation, it refers to the moment of transitions between distinct reflections and more temporally dense echoes [144, 128, 118, 59].

Many reverberation methods separate the early reflections from the late reverberation to allow each part to be processed separately. In [120], Schroeder added a multi-tap delay at the beginning of the system to ensure that the perceptually critical early reflections were reproduced.

Through the various combinations of gains, delays, and filters, delay networks offer a wide range of tuning capabilities while being efficient. This makes them convenient to use in creative design. However, the vast realm of possible arrangements posed a design challenge for the early delay networks, which had to be tuned perceptually [116, 173]. In [174], Moorer further studied these structures and proposed replacing the gains with frequency-dependent absorbent filters, inspired by the filter design
Artificial Reverberation

Figure 4.6. Flow diagram of Gerzon’s delay network [180].

proposed in [175] to account for the attenuation of high frequencies due to air absorption. Moorer also provided a specific set of practical values for the delays and gains yielding pleasing perceptual qualities, which made delay networks more accessible.

The commercialization of good-sounding reverberation units means that the practical design of many delay networks are not publicly available for researchers to explore, but a few exceptions exist [176, 177]. In [178], the design of a popular commercial reverberator is presented and it demonstrates how complex they can get.

Through a series of articles [179, 180, 181], Gerzon introduced the idea of organizing sets of delay lines in a recirculating network to create a reverberator capable of producing high density of echoes from a small set of interconnected delay lines. In this design, the delay lengths were chosen to be mutually prime, which was more challenging at the time because digital delay units were sold with 5-ms size increments.

The recirculating matrix $A$ (Figure 4.6) controls the feedback gain for each input-output combination of delays. The matrix is also constrained to be orthogonal to ensure lossless recirculation. A practical value of 0.7 is given for the recirculating gain $g$ to obtain a pleasant sounding reverberation.

Stautner proposed a delay-network structure for quadraphonic sound reproduction [182]. In this design, each delay path is connected to an output channel spatially distributed around the listener.

The signal from the direct sound is inserted into the system through the input $x_i$, which corresponds to its incident direction (Figure 4.7). Early reflections are simulated using the ISM [13] to calculate the pre-delay, attenuation, and incident direction for each reflection. These early reflections
Artificial Reverberation

Figure 4.7. Flow diagram for Stautner’s quadraphonic delay network [182]. The input \( x_i \) contains both panned direct sound and a set of early reflections generated using geometrical acoustics positioned at the nearest input channel. The filters \( G_i \) combine the low-pass, band-pass, and high-pass filters from the original formulation.

are also distributed in the system through the \( x_i(n) \) inputs.

In this method, each channel feeds the input of its two neighbouring delay path resulting in the orthogonal recirculating matrix

\[
A = \begin{bmatrix}
0 & 1 & 1 & 0 \\
-1 & 0 & 0 & -1 \\
1 & 0 & 0 & -1 \\
0 & 1 & -1 & 0
\end{bmatrix}, \quad (4.3)
\]

where some of the recirculating paths invert the phase of the signal. The attenuation gain used in the recirculation is set to

\[
g < \frac{1}{\sqrt{2}}. \quad (4.4)
\]

Absorbent filters \( F_i(z) \) are used to attenuate each delay line. In the article, this correspond to a set of low-pass, band-pass, and high-pass filters tuned with the absorbent coefficient of wall materials, and a low-pass filter that accounts for air absorption. The article also suggests continuously varying the lengths of the delay lines slightly (\( \leq 1 \text{ ms} \)) to vary the peaks in the frequency response and minimize flutter echoes.

In [8], Jot made a great leap forward in the formalism of feedback-delay networks (FDNs), allowing them to reach a level of popularity still maintained today. The FDN adds input and output gains \( b_i \) and \( c_i \) as well as a path for the direct sound with attenuation \( d \) (see Figure 4.8). The article provides a formal definition for the design of absorbent filters, and an in-depth analysis of the frequency response of the system is also given.
The exponential decay of energy is produced from a set of attenuation parameters defining frequency-dependent decay rates as

\[ g_{dB}(\omega) = -60 \frac{1}{T_{60}(\omega) f_s}, \]  

(4.5)

where \( T_{60}(\omega) \) is the specified frequency-dependent reverberation time. The per-sample linear attenuation gain is obtained from

\[ g_{lin}(\omega) = 10^{\frac{g_{dB}(\omega)}{20}}, \]  

(4.6)

which is used to calculate the attenuation required at the output of each delay line using

\[ g_i(\omega) = (g_{lin}(\omega))^{m_i}, \]  

(4.7)

where \( m_i \) is the length of the delay line.

In turn, these gains are used to design absorbent filters for their corresponding frequencies. For instance, the transfer function of a first-order low-shelf filter is specified with [183]

\[ G(z) = \frac{g \tan(\frac{\omega}{2}) + \sqrt{g} + [g \tan(\frac{\omega}{2}) - \sqrt{g}]z^{-1}}{\tan(\frac{\omega}{2}) + \sqrt{g} + [\tan(\frac{\omega}{2}) - \sqrt{g}]z^{-1}}, \]  

(4.8)

for an attenuation coefficient \( g \) and a center frequency \( \omega \). Finally, in the \( z \)-domain, the system is described as

\[ y(z) = c^T G(z)s(z) + dx(z), \]  

(4.9)

\[ s(z) = Z(z)[AG(z)s(z) + bx(z)], \]  

(4.10)
Artificial Reverberation

where \( s(z) \), \( b \), and \( c \) are

\[
s(z) = \begin{pmatrix} s_1(z) \\ s_2(z) \\ \vdots \\ s_N(z) \end{pmatrix}, \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{pmatrix}, \quad c = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \end{pmatrix},
\]

(4.11)

and \( Z(z) \), and \( G(z) \) are

\[
Z(z) = \begin{pmatrix} z^{-m_1} & 0 & \ldots & 0 \\ 0 & z^{-m_2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & z^{-m_N} \end{pmatrix},
\]

(4.12)

\[
G(z) = \begin{pmatrix} G_1(z) & 0 & \ldots & 0 \\ 0 & G_2(z) & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & G_N(z) \end{pmatrix}.
\]

(4.13)

This formulation from [8] has a single input and a single output, but this design is easily expanded to multiple inputs and outputs by considering the formulation in [182].

The lossless properties of the recirculating matrix was further developed in [184] with the introduction of circulant matrices, and [185] presented a general definition for the requirements of lossless matrices. Since these networks are designed as complex feedback filters, they do not have AP properties. More complex network designs with AP properties were proposed in [180, 186] and further formalized in [187].

To achieve natural sounding reverberation it is important that the system quickly reaches a sufficiently high echo density, which can be done by using shorter delays or adding more delay lines. To avoid multiple impulses landing on the same sample, thus losing density, it is important to ensure that delay lengths are mutually prime [16].

However, this is not sufficient to prevent different recirculation paths from landing on the same sample. For example, an impulse going through the first and second delay line will land on the same sample as one going through the second delay followed by the first. To minimize this occurrence, [188] proposed to quantifying this overlap to help select an optimal set of lengths. Recently, [189] proposed adding small delays in the recirculating matrix to ensure that every recirculation path has a unique combined length. However, these small delays have an impact on the modal distribution, which was improved in [190].

The modal density is another key characteristics of reverberation [191]. A popular method to increase the modal density in delay networks is
to vary slightly the length of delay lines over time thus producing more modes [182, 192, 116, 178], which requires the use of fractional delay lines [193]. In [194], a time-varying recirculating matrix is used to modulate the recirculation coefficients that changes the amplitude of modes over time. To reproduce the frequency-response of real rooms, the absorbent filters require precise design, usually through the use of a set of high-order equalization filters [195, 183, 196, 197, 198].

4.5.1 Velvet-Noise Feedback-Delay Network

As delay networks often suffer from a slow echo density build-up, the use of more delay lines is necessary to ensure that a sufficient echo density is reached in a short period of time. On the other hand, the echo density is a key strength of noise-based reverberators that are very dense from the very beginning.

In Publication IV (p.139), short sequences of velvet noise are added to the input and output of an FDN to increase its echo density. Using this method, it was demonstrated that the number of required delay lines could be reduced without impacting the perceptual characteristics of the reverberator. In the proposed method, the input and output gains from a traditional FDN (Figure 4.8) are replaced with velvet-noise sequences \( B_i(z) \) and \( C_i(z) \) (Figure 4.9).

With this design, a 32-delay-line FDN was found to be comparable to a 16-delay-line reverberator in the proposed model. The time-domain convolution with the velvet-noise sequences required extra computation, but due to the reduced number of delay lines in the system, up to a 52% reduction of floating-point operations (FLOPS) is be obtained.
5. Hybrid Reverberation Methods

The methods detailed in the previous chapter offer perceptually motivated and tuneable reverberation. Their design was substantially simplified when Jot introduced the formulation for absorbent filters, providing a comprehensive set of parameters for the system using $T_{60}(\omega)$ as the target [8]. However, some applications of artificial reverberation require a more physically-informed reproduction of a reverberant sound field. This means that certain physical attributes of a room, such as its dimensions, are used to specify some of the properties of the reverberator.

This chapter describes hybrid reverberation algorithms combining delay networks with geometrical acoustics as well as methods capable of producing more complex multichannel outputs, such as Ambisonics and binaural signals. Reverberation algorithms are grouped into sub-categories, although many of the methods described here overlap in various aspects used in the categorization.

The chapter concludes with a section on directional reverberation and contains a detailed description of the contributions made to this topic in Publication V, VI, and VII. In these publications, an analysis framework is proposed to assess the perceptual importance of directional characteristics in late reverberation, along with reverberation algorithms leveraging the use of delay networks to reproduce these characteristics.

5.1 Geometrically-Informed Reverberators

In [120], Schroeder suggested the use of geometrical acoustics to specify the multi-tap output of a delay line. Each tap accounts for an early reflection in a simulated room. The delayed signals serve as input for a late reverberation block (Figure 4.5). Because early reflections are a perceptually critical aspect of reverberation [1, 64], this design was proposed to offer simulated reverberation in the context of architectural design.

The reproduction of moving sources in reverberant rooms was studied by Chowning in [199]. Based on the location of the sound source relative to
the listener, frequency modulation is applied to produce the Doppler shift of a moving source. The source signal is then panned between loudspeakers on a four-channel system using pairwise panning defined as [199]

\[
y_a = \sqrt{1 - \frac{1}{2} \left(1 + \tan \left(\theta - \frac{\theta_{\text{max}}}{2}\right)\right)},
\]

\[
y_b = \sqrt{\frac{1}{2} \left(1 + \tan \left(\theta - \frac{\theta_{\text{max}}}{2}\right)\right)},
\] (5.1)

where \(\theta\) is the source’s incidence angle in the horizontal plane, \(\theta_{\text{max}}\) is the maximum angle between two adjacent loudspeakers, and \(y_a, y_b\) are two adjacent loudspeakers encompassing \(\theta\). The amplitude of the direct sound is also modulated based on the distance \(D\) between the source and the listener \((1/D)\).

Each output channel has an independent reverberator. The input of each of the reverberators consists of a mix between the panned signal, which is distributed to two channels as determined by Equation 5.1, and attenuated using

\[
\left(1 - \frac{1}{D}\right) \left(\frac{b}{\sqrt{D}}\right),
\] (5.2)

where \(b\) is a parameter to adjust the overall level of the reverberant signal. Similarly, the dry signal is also mixed into all channels using

\[
\left(\frac{1}{D}\right) \left(\frac{b}{\sqrt{D}}\right),
\] (5.3)

thus controlling the mix between the dry and wet signal. Here, the specific design of the late reverberator is not documented, and based on the reference in the article a Schroeder reverberator is assumed.

### 5.2 Binaural Reverberation

In a series of articles, Kendall developed a binaural reverberator using the ISM to specify a delay network [200, 9, 201, 202]. In this design, a set of delay lines are set from the distance of the first and second-order reflections, calculated using the ISM. The incident direction of individual reflections also determines their directions in the output of the system. A binaural output is produced by panning the direct sound and first-order reflections using HRTFs (Figure 5.1).

Together, the first-and second-order delays form a network of recirculating delays that produces decorrelated late reverberation. In the network, the second-order delays are split into two recirculating delay lines, each
representing the propagation to and from a wall. Having two recirculating paths yield more delayed signals and thus a higher echo density. Frequency-dependent filters are placed before and after each delay path to account for wall and air absorption, respectively.

The early 1990s was a prolific time for binaural reverberators using geometrical and statistical methods [203, 204, 205, 206]. In [203, 150], Martin divided the reverberation into three stages: an early reflection stage producing a fixed amount of geometrically-informed reflections, a pseudo-statistical stage using linear prediction to estimate higher-order reflections, and finally, a late reverberation stage using Gaussian white-noise to produce two decorrelated decaying signals \( y_1 \), and \( y_2 \).

The two noise sequences \( y_1 \), and \( y_2 \) were mixed together to reintroduce some correlation and produce binaural signals using

\[
\begin{align*}
  y_L(n) &= \cos \left( \frac{1}{2} \arcsin(\gamma_{IACC}) \right) y_1(n) + \sin \left( \frac{1}{2} \arcsin(\gamma_{IACC}) \right) y_2(n), \\
  y_R(n) &= \sin \left( \frac{1}{2} \arcsin(\gamma_{IACC}) \right) y_1(n) + \cos \left( \frac{1}{2} \arcsin(\gamma_{IACC}) \right) y_2(n),
\end{align*}
\]

(5.4)

where \( \gamma_{IACC} \) is a measure of the IACC, such as the one defined in Equation 2.33. In this system, the IACC is estimated from the binaural signal produced in the second stage of reverberation.

In [204], Gardner calculated image sources up to the fifth order and pruned by merging taps occurring within 1 ms of each other and keeping only the 40 loudest taps. These reflections are panned across a multichannel loudspeaker system using intensity panning [207], and the signal is fed to a delay network, tuned to follow the appropriate decay rate.

In [208, 209], Jot et al. cascaded early reflections and late reverberation in three stages. Starting from early reflections, which are panned either left or right, each reflection also serves as the input to the individual delay lines of a delay network. The output of the first delay network is used as the input to a second late reverberation unit. The final output is encoded...
to either binaural, multichannel, or Ambisonics format. For a binaural output, the late reverberation stage is filtered in order to adjust the IACC between the left and right output channels. The reverberator was also built into a larger system, producing moving sources, source directivity, distance attenuation, and room equalization.

A reverberator reproducing the characteristics of a captured binaural IR is proposed by Väänänen et al. in [210]. In this method, a set of early reflections are produced at each ear by analyzing the individual peaks contained in the first 100 ms of the binaural IR. The direct sound and the early reflections are then fed to a delay network that produces the late reverberation. In this delay network, each recirculating delay path contains an AP and a low-pass filter, and the latter is specified by frequency-dependent decay rates measured in the captured IR. Here, instead of using a recirculating matrix, the output of all the delay paths are summed together and fed back to the input of each delay.

In [211, 212], Christensen et al. proposed a multichannel reverberator with distinct signal paths for early and late reverberation. The last stage of the reverberator has a directional encoding stage that supports different output formats such as Ambisonics, HRTF, and channel-based formats through panning [105]. A key distinction with other models is that each channel has a separate reverberation block. Hence, this reverberator is capable of producing anisotropic decaying sound fields. However, the design outlined in the paper only specifies the intended goal as obtaining fully decorrelated outputs by using independent delay networks.

Recently, the emergence of applications for binaural sound reproduction has motivated the development of several binaural reverberation methods. As with previous binaural reverberators [200, 208], a key characteristic of these reverberators is the use of individually-panned early reflections and decorrelated late reverberation.

Recently, Hacıhabiboğlu et al. introduced an algorithm to cluster and reduce a set of simulated early reflections resulting in computational savings [213]. Leveraging on the precedence effect, the simulated reflections reaching the listener within a small time frame, and thus perceptually indistinguishable, are collapsed into the same sample before reaching a late reverberation stage.

In [214] by Menzer et al., the decorrelated stereo output of an FDN is mixed using a frequency-dependent IACC control to match the spectral envelope of an existing binaural response. In [215], the method is expanded to use two cascaded delay networks. The first network is specified by the directions and lengths of first-order reflections, and the second network is used to produce the diffuse reverberation similar to the one in [214]. Finally, in [216], the method is modified to take multiple sound sources as the input.

An extended FDN, informed by a set of simulated early reflections, is pro-
posed by Wendt et al. in [217]. Early reflections are individually processed using HRTF filters for binaural auralization and the hybrid reverberator expands on previous systems [9, 204] by incorporating spectral filtering based on a room’s material.

In [218], Borß et al. introduced a two-stage reverberator which uses a filter bank and a set of white-noise sequences to reproduce the frequency-dependent energy decay and IACC of a recorded binaural IR. A multichannel reverberator was proposed by Oksanen et al. in [219], that uses velvet noise to produce a set of direction-dependent decay characteristics from an existing SIR. The use of decorrelated noise sequences in a parametric binaural reverberation is further explored in [220].

In [221], the transfer of energy between the early reflection and the late reverberation stage is formalized to control the amount of energy entering the late reverberation stage based on the location of the listener in a sound scene. In [154], a simplified one-stage binaural delay network, using a set of delay lines determined by reflection paths, is perceptually compared to the ISM and captured binaural responses.

5.3 Digital Waveguide

Another approach to specify delay networks using room properties is called digital waveguide (DWG) networks [192]. DWG networks, originally a modified form of DWG filters, was first proposed by Smith as a reverberation algorithm [192]. It consists of a network of interconnected bi-directional delay lines representing physical propagation paths. The delay lengths represent propagation time, and the connections between delay lines are controlled through scattering junctions defining the sound pressure using the following equation [222]:

\[
p_j(n) = \frac{2}{\sum_{i=1}^{N} \Gamma_i} \sum_{i=1}^{N} \Gamma_i p_i^+(n),
\]

(5.5)

where \(\Gamma_i\) and \(p_i^+(n)\) are the admittance and sound pressure, respectively, of individual incoming waveguides \(i\) at junction \(j\).

The bi-directional delay lines of a DWG network are suitable to simulate a sound source propagating outwards from a specific point in the delay network. In its simplest form, a one-dimension DWG network is capable of producing the behavior of a guitar string, which shares some similarities with the Karplus-Strong algorithm [223, 224]. This method was later generalized through the discretization of a two-dimensional grid space, called DWG mesh, suitable for the physical modeling of musical instruments [225]. In three-dimensional space, a DWG mesh may be used to simulate sound propagating in a room [226, 227].

Since FDNs are the general building blocks to define a network of delays,
a DWG network may be converted to an FDN design where the scattering junctions are equivalent to the filters and the matrix controlling the recirculation of delays [184, 228]. More complex DWG meshes are mathematically comparable to FDTD under certain conditions [229].

A DWG network is formalized in [222] by using the location of geometrically-specified first-order reflections from a shoe box. The location of the reflection on each wall is used to create paths from the sound source to the walls, from the walls to the listener, and between walls. The distance of each path informs the delay-lengths in a recirculating delay network. Another DWG network was later proposed, which simulates sound propagation in a forest by incorporating the properties of wave scattering around cylindrical objects in the physical model, which form scattering junctions simulating trees [230].

The reverberation model proposed by [222] was later improved by De Sena et al. in [231, 232], removing some propagation paths to reduce the computational cost and using a more intuitive approach to define the filters at the scattering junctions where wall absorption coefficients are used. A DWG network design was proposed in [233] using higher-order propagation paths defined within for complex three-dimensional scenes.

To reduce the computational cost of three-dimensional meshes in the context of room acoustics simulation, the use of two-dimensional meshes has been studied in [234]. Various hybrid approaches combining two-dimensional approximations, Gaussian noise, and ray tracing to replace segments of propagation were proposed by Murphy et al. in [235].

In [236], a method to design unilossless and spatially-aware recirculating matrices is proposed. Using this method, the recirculating matrix of an FDN may be designed to simulate two interconnected rooms with different decay rates and a controlled amount of recirculation between them. This matrix design approach also has applications to obtain angle-aware scattering junctions for a DWG network.

5.4 Ambisonics Delay Networks

For immersive applications, an Ambisonics reverberator offers the benefit of supporting a wider range of output configurations, including head-tracked binaural reproduction. In [237], an Ambisonics input signal with \( Q \) channels is first converted into discrete \( K \) plane waves (Figure 5.2). The early part of the reverberation is produced by using multiple cascaded delay stages to increase the density gradually. These early reflections are encoded individually to Ambisonics. Multiple decorrelated signals are obtained from a late-reverberation delay network and a gain parameter controls the amount of the reverberant signal in different directions.

A first-order Ambisonics (FOA) reverberation is obtained using a set of
Schroeder-type reverberators in [238] by distributing the output of each reverberator spatially. A scattering matrix controls the redistribution of reverberated signals amongst the directions to compensate for the directivity of the input signal. Three steering parameters control the output gain between the left-right, front-back, and top-down directional axes. In this formulation, contrary to what was proposed in [8], every direction has an independent reverberation unit and the recirculating gains in the reverberator are set across all directions and independently of the delay length. For this reason, a longer delay line in the reverberator has a slower decay, resulting in different reverberation times in different directions. After encoding the final output of the reverberator in FOA, a spread parameter controls the proportion of the reverberated signal taken from the omni-directional channel.

5.5 Directional Reverberation

The physical properties of a room determine its ability to diffuse sound energy as waves propagate through it. Poor diffusion in a room due, for instance, to flat rigid walls, leads to interference patterns increasing the potential sound pressure at certain locations further away from the center (Section 2). As such, achieving good diffusion is one of the challenging aspects of building reverberant rooms suitable for scientific measurements [46]. Whereas concert halls are built to favor diffusion, since it gives a better musical experience, other types of rooms are well varied.

In the context of immersive sound reproduction, such as augmented reality, the accuracy of reproduction is imperative to ensure that a reproduced sound blends with the real world. Hence, it is important to assess the characteristics of reverberant sound fields and understand the extent in which we perceive their characteristics. A direction-dependent reverberant sound field, or directional reverberation for short, is also sometimes referred to as a non-diffuse field [54], while others use anisotropic sound fields [46].
This section contains an overview of key contributions made to the objective and subjective analyses of sound fields as well as their reproduction using artificial reverberation.

5.5.1 Analysis

Publication V (p.149), proposes an analysis framework to measure the anisotropy in a captured reverberant sound field as well as a perceptual study methodology to evaluate our capacity to hear anisotropic characteristics. This method is also useful to analyze the output of a multichannel reverberator and to evaluate its capacity to preserve direction-dependent characteristics in the decay.

The proposed objective analysis method is based on energy-decay measures such as EDC and EDR (Equation 2.22 and Figure 2.23). It compares directional decay properties to a mean in all directions to measure how much each direction deviates from an isotropic sound field similarly to the method proposed in [239].

From a SIR encoded in Ambisonics, a set of directional IRs are extracted using a beamformer, as expressed in Equation 2.31. The directional IRs are used to produce a set of directional decay curves in dB from

$$\text{EDC}_{dB}(n, \phi, \theta) = 10 \log_{10} \left( \sum_{i=n}^{N} (y(i, \phi, \theta))^2 \right),$$  \hspace{1cm} (5.6)

where $N$ is the total number of samples in the signal. The $\text{EDC}_{dB}(n, \phi, 0)$ analysis of the captured shoe-box room is illustrated in Figure 5.3. We can see that the distribution of energy, between the analyzed directions, diverges increasingly with time, which illustrates the difference in the rate of decay based on the direction on the horizontal plane. The energy decays...
faster at $0^\circ$ and $180^\circ$ compared to $90^\circ$ and $270^\circ$.

From the properties of a diffuse field (Section 2.3), the energy of an isotropic sound field is uniformly distributed in all directions. As such, a diffuse field’s directional decay characteristics should remain close to an average decay curve $\text{EDR}_{\text{dB}}(n)$. To quantify the anisotropy of a reverberant sound field, we look at the deviation of each of the directional decay curves from the mean. The energy decay deviation (EDD) is calculated as

$$\text{EDD}(n, \phi, \theta) = \text{EDR}_{\text{dB}}(n, \phi, \theta) - \text{EDR}_{\text{dB}}(n),$$

(5.7)

which yields a time- and direction-dependent measure containing negative and positive values highlighting the anisotropic characteristics of a reverberant sound field, as illustrated in Figure 5.4. One benefit of using an analysis based on the EDC is that it does not require any windowing of the signal in the time domain, as opposed to other methods using the mean of energy over a short time window. For these reasons, this method is not characterized by a windowing function and its length. However, this method is impacted by the Ambisonics order and the beamformer used, which may smooth out the results over multiple directions.

In the context of human perception, taking into account a specific head-orientation and binaural hearing is useful. Using a binaural IR, which may be obtained from a SIR in a specific orientation, we calculate the interaural energy-decay deviation (IEDD) from

$$\text{IEDD}(t, \omega) = \text{EDR}_{\text{dB}}^L(t, \omega) - \text{EDR}_{\text{dB}}^R(t, \omega),$$

(5.8)

Figure 5.4. Energy decay deviation (EDD($\phi, \theta$)), on the horizontal plane, of the example SIR.
where $\omega$ is a frequency band and $\text{EDR}^L_{\text{dB}}$ is computed from the left channel of the binaural IR and the $\text{EDR}^R_{\text{dB}}$ from the right channel. In order to get a general idea of the energy deviation per frequency, the root-mean-square in time is calculated as

$$\text{IEDD}(\omega) = \sqrt{\frac{1}{T} \sum_{t=0}^{T-1} \text{IEDD}(t, \omega)^2}.$$  (5.9)

The EDD and IEDD measures are useful to analyze a sound field objectively. To assess the same characteristics of late reverberation perceptively, a few important aspects must be considered. For instance, one challenging aspect of the perceptual evaluation of late reverberation is the effect of early reflections on perception. For this reason, in Publication V, we only assess the perception of the sound field after the estimated mixing time (Section 2.4). In the perceptual study, a reverberant sound field is reproduced in an anechoic room equipped with a spherical loudspeaker array, and an ABX test [240] is used to assess the perception of small differences occurring when the late reverberant sound field is rotated. This assesses the perception of the directional characteristics of the reverberant sound field.

Another aspect to consider is the presence of noise in a captured SIR, which may be perceivable during reproduction. To ensure that noise is not a factor in the perceptual evaluation, a spatial denoising algorithm [38] was used to remove it.

From the rooms analyzed in the study, the results suggest that a sound field with an IEDD$(t, \omega)$ value as small as 1 dB is perceivable, implying that these characteristics are an important factor to consider for spatial sound reproduction when accuracy is critical.

By proposing objective analysis measures as well as a formal perceptual study methodology for reverberant sound fields, Publication V offers a complete framework to evaluate anisotropic reverberant sound field.

### 5.5.2 Reproduction

The direction-dependent characteristics detailed in the previous section are analyzed in captured SIRs, which can also be used for their reproduction. However, capturing every possible combination of sound source and listener locations is not always practical. The computational cost of convolution with long impulse responses may also be prohibitive in some applications. For these reasons, the use of artificial reverberation methods in the reproduction may be preferable.

Using existing methods (Section 5.1) it is possible to simulate the early reflections of an interactive scene using geometric acoustics and to use a multichannel artificial reverberator to produce decorrelated late reverber-
Figure 5.5. Directional decays (EDC_{dB}(\phi, \theta)) using direction-dependent output gains, while the decay rate remains constant for all directions.

...ation. However, this type of reverberator does not inherently produce the directional decay characteristics illustrated in the previous section.

As an illustrative example, Figure 5.5 shows the output of an Ambisonics reverberator using direction-dependent output gains to control the reverberation. However, unlike the captured SIR in Figure 5.3, Figure 5.5 has a constant decay rate in all directions.

5.5.3 Directional Feedback Delay Network

Publication VI (p.163), proposes to expand the design of delay network reverberators to produce direction-dependent decay characteristics from a set of decay parameters (Figure 5.6). This multichannel reverberation method is called the directional feedback delay network (DFDN).

In this formulation, the delay lines of an FDN (Figure 4.8) become delay-groups containing the Q channels of an Ambisonics signal (Figure 5.6). An input signal is weighted in each delay group using b_i, an input-weighing vector.

The directional decay characteristics of the reverberator are controlled through a set of direction-dependent decay parameters T_{60}(\phi, \theta), which are converted to per-sample directional attenuation gains using Equations 4.5 to 4.7. The resulting direction-dependent gain vector g_i for each delay-group is used to define a directional weighting transform T_i using a method similar to the one presented in Equation 2.32, although a more thorough formulation of the transform is given in Publication VI.

Due to the Ambisonics processing, the individual delay-lines contained in a delay group are constrained to share the same length. To produce direction-dependent decay characteristics, the recirculation matrix A pre-
vents the individual channels of a delay group from mixing with other channels. This is done to preserve the spatial distribution of energy and to ensure that the directional decay follows the ones specified by the \( T_{60}(\phi, \theta) \) values. The Hadamard product is used to multiply individual channels with an output gain vector \( c_i \) in Figure 5.6.

### 5.5.4 Frequency-Dependent Directional Feedback Delay Network

In Publication VII (p.177), the delay groups of the DFDN are simplified to a set of \( K \) spatially distributed channels, each represented as a thick line in Figure 5.7. These channels replace the Ambisonics signals in Publication VI. While this formulation requires more channels to retain an equivalent spatial resolution, the simplifications in the processing ultimately allow for a more efficient direction- and frequency-dependent control of the decay. Similarly to the methods presented in Section 5.1, a series of geometrically-based early reflections are used as the input \( x(n) \) to the reverberator.

The decay in the reverberator is controlled through a set of direction- and frequency-dependent decay parameters \( T_{60}(\omega, \phi, \theta) \) that are used to specify absorbent filters at the output of each of the delay groups. The individual delay lines of each of the delay groups have their own set of absorbent filters, and a multi-band equalizer is used for precise tuning.

One benefit of using channels encoded for a set of directions instead of Ambisonics is that the delay groups are no longer required to share the same delay lengths for each individual channel. Removing this constraint allows all channels of each delay group to contribute independently to the reverberation, which significantly increases the echo density at the output. The increase in echo density allows a reduction of the number of delay
Further processing of the directional reverberation is possible using the directional input and output gains $b_i$ and $c_i$, respectively. Finally, the output $y(n)$ of the reverberator is encoded to a specified output format, which is either channel-based, Ambisonics, or binaural.

$$y(z) = \sum_{i=1}^{N} c_i^T G_i(z) s_i(z) + d \odot x(z),$$  \hspace{1cm} (5.10)$$

$$s_i(z) = Z_i(z) \left[ b_i \odot x(z) + \sum_{j=1}^{N} A_{ij} G_j(z) s_j(z) \right],$$  \hspace{1cm} (5.11)$$

where $\odot$ denotes element-wise multiplication (Hadamard product) and vectors $s(z)$, $b$, and $c$ are

$$s_i(z) = \begin{pmatrix} s_{i,1}(z) \\ s_{i,2}(z) \\ \vdots \\ s_{i,K}(z) \end{pmatrix},$$  \hspace{1cm} (5.12)$$

$$b_i = \begin{pmatrix} b_{i,1} \\ b_{i,2} \\ \vdots \\ b_{i,K}(z) \end{pmatrix},$$  \hspace{1cm} (5.13)$$
Hybrid Reverberation Methods

Figure 5.8. Directional energy-decay curves (EDC_{dB}(\phi, 0)) on the horizontal plane from the output of a DFDN.

\[
c_i = \begin{pmatrix} c_{i,1} \\ c_{i,2} \\ \vdots \\ c_{i,K} \end{pmatrix}, \tag{5.14}
\]

and the matrices \(Z_i(z)\) and \(G_i(z)\) are

\[
Z_i(z) = \begin{pmatrix} z^{-m_{i,1}} & 0 & \cdots & 0 \\ 0 & z^{-m_{i,2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & z^{-m_{i,K}} \end{pmatrix}, \tag{5.15}
\]

\[
G_i(z) = \begin{pmatrix} G_{i,1}(z) & 0 & \cdots & 0 \\ 0 & G_{i,2}(z) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & G_{i,K}(z) \end{pmatrix}. \tag{5.16}
\]

Figure 5.8 visualizes the EDC_{dB}(\phi, \theta) taken from the output of a DFDN. This specific DFDN was specified by a set of directional decay times \(T_{60}(\phi, \theta)\) measured in the captured shoe-box room (Figure 5.3). In Figure 5.9, the EDD (Section 5.5.1) is analyzed at the output of the same DFDN configuration. Therefore, the DFDN is suitable for reproduction of decay characteristics contained in an existing SIR. Furthermore, since these decay characteristics are controlled through a set of gains and filters, an interactive scene may be reproduced by simply modulating these gains and filters, which allows for six-degrees-of-freedom in the reverberation [10].
In the article, the echo density of different configurations is compared, along with the reproduction accuracy of distinct uniformly distributed spatial grids. Ultimately, this formulation allows flexibility to increase or decrease the spatial and frequency accuracy to accommodate various computational-cost requirements.
6. Summary of Main Results

This section presents the main results from the featured publications that are related to the author’s work.

Publication I - “Velvet-Noise Decorrelator”

In Publication I, the use of velvet noise as a decorrelation filter is introduced. Velvet-noise sequences have a perceptually flat spectrum and a random phase, which makes them suitable to be used as broadband time-domain decorrelation filters. Since a velvet-noise sequence usually consists of the values $-1, 0, \text{and } 1$, the article investigates the use of decaying envelopes together with the sequences. A decaying envelope is necessary to preserve the reproduction of transients in the decorrelated signals. The proposed method is compared with white noise, which has well-known decorrelation properties, and the results show a similar decorrelating potential.

Publication II - “Optimized Velvet-Noise Decorrelator”

Publication II improves on the velvet-noise decorrelator introduced in Publication I and proposes an optimization process to minimize the coloration and find ideal pairs of filters for stereo decorrelation. The optimization process slightly alters the location of individual impulses in a velvet-noise sequence to minimize its spectral coloration. A set of 500 optimized sequences is produced, and the best low-correlated pairs are selected from this set. A weighing factor is used to determine how to prioritize between spectral flatness and low cross-correlation. The use of optimized velvet-noise sequences using lower echo density is also investigated. Two perceptual studies are conducted to evaluate the method. The first evaluates the spectral differences between the optimized and non-optimised sequences. The second perceptual study evaluates the decorrelating prop-
Summary of Main Results

Properties of the filters for up-mixing a mono audio signal to stereo. Both studies were conducted over headphones, and white-noise decorrelation was used as the reference method. The results showed that the optimized sequences using less impulses were appropriate for use in stereo decorrelation, which yields 88% fewer floating-point operations when compared to the FFT-based convolution required for comparable white-noise filters.

Publication III - “Late Reverberation Synthesis Using Filtered Velvet Noise”

Publication III extends the late-reverberation method introduced in [165] using filtered velvet noise to reproduce the characteristics of an existing impulse response. The method uses velvet noise to reproduce segments of an existing impulse response. Velvet-noise sequences are used to produce a perceptually flat spectrum that is then altered using a coloration filter. This article improves on the previous formulation by using cascaded differential coloration filters, which reduces the required impulse density of the velvet noise used for later segments, thus lowering the overall computational cost. A new segmentation method is also proposed to find the optimal segment boundaries. A perceptual study evaluates the similarities between a recorded IR and the output of the reverberator. The results showed that the differences were noticeable in most cases, but these differences were rated as small by the test subjects. The computational cost is comparable to FFT-based convolution methods, but using much less memory.

Publication IV - “Velvet-Noise Feedback-Delay Network”

In Publication IV, a delay-network reverberator is modified to increase its echo density using velvet noise. Delay network reverberators, such as the FDN, require many delay lines to ensure a sufficiently high density of echoes early in their response. In this article, the velvet-noise sequences proposed for decorrelation in Publication I and II are used to increase the echo density in an FDN. Objective measures demonstrate that filtering both the input and output of each of the delay lines with velvet-noise sequences produces an optimal increase in echo density and with low spectral deviations. One of the key benefits of this method is the reduction in the number of delay lines required to achieve a sufficient echo density. The propose method requires between 35% and 52% less floating-point operation in practical configurations.
**Publication V - “Perceptual Analysis of Directional Late Reverberation”**

In Publication V, an analysis framework is proposed to assess the perception of directional characteristics of a reverberant sound field. Based on existing methods used to analyze impulse responses, the publication introduces two objective measures to analyze the directional properties of SIRs. The proposed method measures the deviation of energy in the decay between a mean and a set of directions, or between the two channels of a binaural signal. This process is also extended to the frequency domain for further analysis. These objective measures are used to analyze SIRs from four different locations, and a perceptual study is conducted to assess whether these directional characteristics are subjectively significant. A testing methodology is proposed where a rotation is applied to the SIRs after their estimated mixing time. This provides a set of stimuli containing a rotated late reverberant sound field without altering the early part of the reverberation. In the proposed methodology, existing denoising techniques are also used to ensure that the background noise from the captured SIR does not influence the results of the perceptual study. Using the stimuli, an ABX perceptual study was conducted using eighteen test subjects, and the results indicate that a deviation of approximately 1 dB in the proposed binaural objective measure is sufficient to differentiate between the original and the rotated stimuli. These results demonstrate that the directional characteristics of individual SIR should be evaluated to ensure the appropriate level of details in applications where the accuracy of reproduction is important.

**Publication VI - “Directional Feedback Delay Network”**

Publication VI introduces a delay-network reverberation method capable of reproducing directional reverberation characteristics found to be perceptually important in Publication V. This method is based on previous formulations of an FDN. For multichannel sound reproduction, FDNs may provide sets of mutually decorrelated signals. However, the proposed method improves the multichannel capabilities of the FDN to allow the reproduction of directional decay characteristics, something that was not previously possible. Delay lines are modified as multichannel delay groups, where the signal is encoded in the spherical harmonics domain and the reverberator is characterized using a set of direction-dependent reverberation times. A spatial transform in the Ambisonics domain is placed after each delay group to alter the gains of different directions and to produce the specified reverberation times.
Publication VII - “Frequency-Dependent Directional Feedback Delay Network”

Publication VII improves the method introduced in Publication VI to efficiently produce frequency- and direction-dependent reverberation. In this formulation, the delay groups are encoded as individual plane waves distributed on a spatial grid. While this spatial representation of signals requires more channels, when compared to the previous formulation, this ultimately reduces the overall number of computations necessary to produce frequency- and direction-dependent reverberation. A set of geometry-based early reflections distribute the energy spatially at the input of the reverberator. The article demonstrates that by removing the Ambisonics processing from the recirculation loop, we are able to increase the perceived echo density since the delay groups are no longer constrained to share delay length. Another benefit is to simplify the spectral processing, allowing for more precise tuning. The error between the target decay properties and the output of the reverberator is also evaluated for various spatial grid resolutions.
7. Conclusion

In this overview, we have discussed the characteristics of reverberant sound fields, the decorrelation of audio signals, and artificial reverberation methods. The propagation of sound in a room was described, and some of the statistical properties used in room acoustics were detailed. The physical properties of reverberant and diffuse sound fields were reviewed along with the reproduction.

The decorrelation of audio signals and the use of coherence measures was summarized. The use of sparse sequences of random noise, called velvet noise, was proposed as a decorrelation method in Publication I, and this approach was further improved in Publication II by optimizing the location of individual impulses in a sequence. While the method provides very efficient broadband stereo decorrelation, more work remains to generalize the method to more output channels.

We examined the various methods used to reproduce reverberation artificially, and more specifically, methods using random noise to mimic late reverberation and the ones using delay networks for efficient reproduction. An improved reverberation algorithm using velvet noise along with coloration filters was introduced in Publication III, and its ability to reproduce segments of an existing IR was evaluated through a perceptual study which demonstrated that this method has good applications to reduce the storage required by large data sets of IRs. Velvet noise was also used in Publication IV to improve the echo density of an existing delay-network reverberation method, thus reducing the required number of delay lines in the system and the overall computational cost of the system.

Hybrid reverberation methods, which, for instance, may combine geometrical acoustics with delay networks, were also explored as also were binaural and Ambisonics reverberators. To better understand the requirement of spatial audio reverberation methods, we proposed a novel framework to analyze the directional characteristics of reverberant sound fields (Publication V). The proposed objective measurements method shows that the decay of energy in SIRs may not be uniform in all directions, and a perceptual study methodology was developed to demonstrate that once a
certain perceptual threshold has been met this anisotropic distribution of energy is perceivable. As such, the proposed framework is useful in assessing the necessary spatial resolution for the reproduction of reverberant sound fields.

A directional reverberation algorithm, based on the feedback-delay network and capable of producing direction-dependent decay characteristics, was proposed in Publication VI. This method was further improved in Publication VII to allow for a more efficient processing of frequency- and direction-dependent decay characteristics. Using this method, late reverberation is designed using a set of direction-dependent decay properties, which is suitable for the reproduction of the decay characteristics contained in existing SIRs. In this method, the late reverberation is combined with early reflections that may be determined through the use of a geometric virtual acoustics method. The decay characteristics of the reverberator may be modulated in real-time for applications requiring six-degrees-of-freedom, such as in virtual and augmented reality [10].

Although this dissertation offers new means to analyze and reproduce reverberant sound fields containing directional characteristics, more work remains to be done to fully understand the perceptual limits of directional reverberation, especially when other characteristics, such as early reflections or the direct-to-reverberant ratio, may have precedence on the perception. Nonetheless, we hope that by contributing to the understanding of directional reverberation, the work contained in this dissertation may inspire better reproduction methods in the future.
References


References


References


References


Errata

Publication I

Equation 11 should be

\[ T_{dl}(m) = \frac{T_T}{100} \cdot 10^{\frac{2\pi}{M}} - T_{dl}(m - 1), \]

where

\[ T_{dl}(0) = \frac{T_T}{100} \cdot 10^{\frac{2\pi}{M}}. \]

Publication II

In Equation 7, \( \delta \) is the discrete Dirac function. Also, the right term in Equation 18 should be 0.5T, where T is the sampling period.
In this dissertation, the reproduction of reverberant sound fields containing directional characteristics is investigated. A complete framework for the objective and subjective analysis of directional reverberation is introduced, along with reverberation methods capable of producing frequency- and direction-dependent decay properties. Novel uses of velvet noise are also proposed for the decorrelation of audio signals as well as artificial reverberation. The methods detailed in this dissertation offer the means for the auralization of reverberant sound fields in real-time, with applications in the context of immersive sound reproduction such as virtual and augmented reality.