HELSINKI UNIVERSITY OF TECHNOLOGY
Department of Computer Science and Engineering
Telecommunications Software and Multimedia Laboratory

Sami Kiminki

Sound Propagation Theory for Linear Ray Acoustic Modelling

Supervisor: Professor Lauri Savioja
Instructor: Tapio Lokki, D.Sc. (Tech.)

HELSINKI UNIVERSITY OF TECHNOLOGY

ABSTRACT OF THE MASTER'S THESIS

Author:	Sami Kiminki		
Name of the thesis:	Sound Propagation Theory for Linear Ray Acoustic Modelling		
Date:	March 7, 2005	**Number of pages:**	100+14
Department:	Computer Science and Engineering	**Professorship:**	T-111
Supervisor:	Professor Lauri Savioja		
Instructor:	Tapio Lokki, D.Sc. (Tech.)		

In this work, a linear ray acoustic modelling theory is constructed. The theory forms a base for linear ray acoustic modelling methods. As such, the theory can be used to derive and analyse ray methods. Three existing ray modelling methods (the image source method, the radiosity method, and the ray tracing method) are shown to be derivable from the theory. It is also suggested that the theory can be used to derive acoustic characteristics estimators such as the average reverberation time of a room. To the author's knowledge, this is the first attempt to create a theory for acoustic ray modelling.

The theory is divided into two parts: general and acoustic. The general theory consists of general definitions, time-dependent energy propagation equations, and detection equations. The general part yields time-independent ray modelling theory by eliminating time dependency, thus linking the acoustic and the graphic ray modelling. The acoustic part specifies the general definitions as acoustic definitions. The theory lacks sub-surface scattering reflection and edge diffraction. A well-defined extension path for the inclusion is considered, however.

The general definitions consist of mathematical and physical definitions. Energy propagation equations are constructed in detail, resulting in the reflection-iterative construction and the acoustic rendering equation. The first is a straightforward construction, and the second is a balance equation — extension of the Kajiya's rendering equation. The equations evaluate impulse energy responses and are shown to be equivalent using linear operator analysis. An example definition for auralization of energy responses is constructed.

Keywords: general modelling theory, image source method (ISM), radiosity, ray tracing, ray acoustic modelling

TEKNILLINEN KORKEAKOULU DIPLOMITYÖN TIIVISTELMÄ

Tekijä:	Sami Kiminki	
Työn nimi:	Äänen etenemisteoria lineaarisessa sädeakustiikassa	
Päivämäärä:	7.3.2005	**Sivuja:** 100+14
Osasto:	Tietotekniikan osasto	**Professuuri:** T-111
Työn valvoja:	Professori Lauri Savioja	
Työn ohjaaja:	TkT Tapio Lokki	

Työssä rakennetaan pohjateoria lineaariselle sädeakustiselle mallinnukselle. Teoriaa voidaan käyttää sädemenetelmien johtoon ja analyysiin. Kolme olemassaolevaa sädeakustista mallinnusmenetelmää osoitetaan olevan johdettavissa teoriasta (kuvalähde-, radiositeetti- ja säteenseurantamenetelmä). Lisäksi ehdotetaan, että teoriaa voitaisiin käyttää myös akustisten tunnuslukujen estimointiin, esimerkkinä jälkikaiunta-aika. Tämä on tekijän tietämyksen mukaan ensimmäinen yritys luoda kattava sädeakustisen mallinnuksen teoria.

Teoria jaetaan kahteen osaan, yleiseen ja akustiseen. Yleinen osa käsittää yleiset määritelmät, aikariippuvat energiankulkuyhtälöt sekä havainnointiyhtälöt. Teorian yleisestä osasta saadaan lisäksi teoria sädegrafiikalle, kun eliminoidaan aikariippuvuudet. Akustinen osa spesifioi yleiset määritelmät akustisiksi määritelmiksi. Teoriasta puuttuu pinnanalaissironta heijastuksissa sekä reunadiffraktio. Teorian laajennettavuus näiden puutteiden osalta on otettu huomioon.

Yleiset määritelmät koostuvat matemaattisista ja fysikaalisista määritelmistä. Energiankulkuyhtälöt konstruoidaan yksityiskohtaisesti. Tämä johtaa heijastusiteratiiviseen konstruktioon sekä akustiseen mallinnusyhtälöön. Ensimmäinen on suoraviivainen konstruktio. Jälkimmäinen on tasapainoyhtälö, joka on Kajiyan mallinnusyhtälön laajennus. Yhtälöt tuottavat energiaimpulssivasteita ja konstruktiot osoitetaan yhtäläisiksi lineaarioperaattorianalyysilla. Työssä rakennetaan esimerkinomainen määritys energiavasteiden auralisaatioon.

Avainsanat: yleinen mallinnusteoria, kuvalähdemenetelmä, radiositeetti, säteenseuranta, akustiikan sädemallinnus

Acknowledgments

First of all I thank my supervising professor Lauri Savioja and my instructor Tapio Lokki. They made it possible for me to conduct this work and provided invaluable insights into the vast field of acoustic modelling. I also thank my closest associate Heli Nironen for providing lots of background information and reference material, and especially for the tireless conversational efforts on various sound-related topics.

I thank professor Timo Eirola for very useful mathematical discussions, and for allowing me to present the work in his seminar in computational science and engineering. I also thank Jaakko Lehtinen for helping me in radiance calculus, and Janne Kontkanen for valuable information on computer graphics.

Additional thanks should go to Vesa Hirvisalo, Jan Jukka Kainulainen, and Juha Tukkinen for support and otherwise important influence on my work. Finally, I thank Maarit Tirri for numerous comments on the textual layout. Any remaining linguistic anomalies are due to my personal stubbornness and obsessions.

Contents

List of Figures

Notations

Latin alphabets

A		variable used in derivations and proofs
a	$\left[\mathrm{m}^3\right]$	point in surface geometry
area	$\left[\mathrm{m}^2\right]$	surface area operator (p. 77)
B		reflection operator (p. 62)
$\widehat{\overline{B}}$		all reflections operator (p. 66)
$\mathbf{c}$	$\left[\frac{\mathrm{m}}{\mathrm{s}}\right]$	speed of wave front (p. 36)
card		cardinality, size of set (p. 21)
D		delay pattern (p. 72)
$\widehat{D}$		linear system response operator (p. 14)
d		total detection (p. 48,64)
		representation function of linear system response operator (p. 14)
d_0		detection of direct source to observer radiation (p. 55)
d_r		detection of radiance, detection of total reflected radiation via any number of reflections (p. 57)
det		matrix determinant
dtf		directional transfer function for observation (p. 39,47)
$_{\mathrm{e}}$		exitant (prefix) (p. 37)
F		patch emission vector (p. 20,78)
f	$\left[\frac{1}{\mathrm{s}}\right]$	frequency (p. 42)
f		general function

f_r	$\left[\frac{1}{\mathrm{sr}}\right]$	bidirectional reflection distribution function (BRDF) (p. 41)
$f_{\mathrm{r,d}}$		BRDF of an ideal diffuse reflection (p. 45)
$f_{\mathrm{r,s}}$		BRDF of an ideal specular reflection (p. 47)
$\mathcal{G}$		environmental surface geometry (p. 27)
g		Kajiyan geometry term with propagation delay (p. 53)
$\hat{g}$		Kajiyan geometry term without propagation delay (p. 53)
H^α		medium absorption operator (p. 51)
I	$\left[\frac{\mathrm{W}}{\mathrm{m}^2}\right]$	intensity, irradiance, energy flow per surface area (p. 36)
I		identity matrix, identity operator
$_\mathrm{i}$		incident (prefix) (p. 37)
$\hat{\imath}(t)$		impulse response (p. 72)
$L(\Omega)$	$\left[\frac{\mathrm{W}}{\mathrm{m}^2}\right]$	radiance (p. 40,60)
$L_0(\Omega)$	$\left[\frac{\mathrm{W}}{\mathrm{m}^2}\right]$	primary radiance (p. 60)
$\ell(t)$		time-dependent radiance in non-absorptive medium (p. 52)
		total propagated radiance (p. 60,61)
ℓ_0		primary radiance (p. 54,57–58)
ℓ_k		primary radiance after k reflections (p. 57,60)
$\hat{\ell}$		time-dependent intensity in non-absorptive medium (p. 50)
M		mirror reflection operator (p. 45,87)
$\mathbf{n}$		surface normal (p. 27)
$\mathcal{O}()$		asymptotic complexity class
P		vector of energy flows of patches (p. 20,77)
P_k		patch, small but not infinitesimal surface area (p. 19)
p_e		pattern of emittance (p. 38)
R		reflection kernel (p. 53)
r		distance, radius of a sphere

S_r		lossless propagation operator for distance r (p. 50)
$\hat{S}_r$		propagation operator with medium absorption for distance r. Used in the construction of TIA (p. 49)
s		noise signal with average unit intensity (p. 72)
T		triangle (p. 89)
t		time
$\mathbf{v}$	$\left[\frac{\mathrm{m}}{\mathrm{s}}\right]^3$	velocity (vector quantity)
x	$[\mathrm{m}]^3$	point in space
x_o		location of the observer (p. 39,55)
x_p		ray path (p. 10)
x_s		location of the energy source (p. 37,55)

Lower-case Greek alphabets

β		reflectance factor, total reflectance (p. 42)
γ		sensitivity pattern (p. 72)
δ		Dirac delta functional (p. 15,45,91)
θ		elevation angle — angle between Ω and surface normal (p. 26)
ν		visibility function (p. 27)
ν^{-1}		visible geometry (p. 28)
ν_p		inverse projection (p. 28)
ρ		biconical reflectance factor (p. 42)
		probability density (p. 92)
σ		energy profile of sub-band filtered impulse (p. 66,72)
		signal (p. 14)
τ		general scalar or vector
ϕ		azimuth angle of Ω (p. 26)
φ		angle
ψ		wave representation of a fundamental particle (p. 13)
ω	$[\mathrm{sr}]$	solid angle, set of directions (p. 26)

Upper-case Greek alphabets

Φ	[W]	energy flow, energy flux (p. 19,35)
Ω	[sr]	infinitesimal solid angle, direction (p. 26)

Miscellaneous

$\int$	integration operator
∂	partial derivative
$\cdot$	scalar or dot product
$\times$	vector cross product
$*$	convolution operator
$\bigtriangledown$	gradient operator
$\lvert \cdot \rvert$	length, absolute value
$\lVert \cdot \rVert$	norm
2π	solid angle of the hemisphere
4π	solid angle of the sphere

Abbreviations

AEPT	Acoustic Energy Propagation Theory
ARE	Acoustic Rendering Equation
BRDF	Bidirectional Reflection Distribution Function
BSSRDF	Bidirectional Sub-surface Scattering Reflection Distribution Function
def.	definition
eq.	equation
fig.	figure
GEPT	General Energy Propagation Theory
HRTF	Head-related Transfer Function
ISM	Image Source Method
RE	Rendering Equation (by Kajiya)
p.	page
rem.	remark
sec.	Section, subsection
thm.	theorem
TIA	Temporal Intensity Algebra

Preface

I have always sought new, better theories.

When I began my studies, I had assumptions on the graduation work — Master's Thesis. It was supposed to be a challenge. It was supposed to be a proving ground for showing even the tiniest glimpse of comprehension on something, and to show that in the format of science.

Few years after that, around a year ago, it was finally the time for me to start thinking of that final great step until graduation. I began looking for a subject that would allow me to examine a real scientific problem with an academic degree of freedom. Albeit much of my way of thinking was reinvented since the freshman days, the old assumptions held still unchallenged. Through various twists — a story of its own — I found myself in the field of acoustic modelling.

That turned out to offer somewhat the best opportunity imaginable. Having no previous studies in the field, I was able to build my very own insight at the small cost of extra self-oriented studies. A perfect opportunity for theories.

So, I set up for acoustic modelling. Naturally, my studies included literature on general physics, mathematics, general acoustics, practical acoustics, and acoustic modelling methods — including BEMs, waveguides, the wave equation, and acoustic rays. Eager, I decided that I would construct a pressure disturbation propagation model for diffuse reflections using ray concepts. But it failed miserably.

The failure was due to very typical reasons for any such failure: insufficient background and problem analysis misguiding the intuition. A short essay to honour

the failure is included as Appendix B.

Back to the studies but this time looking more strongly to the related fields — computer ray graphics, optics, digital signal processing. I even glanced briefly at remote sensing, quantum physics, and Einstein's texts on general relativity to seek new insights.

When finally redefining the subject, I still felt the need to include diffuse reflections, and I still did not want to do that at the cost of losing generality. The ray thinking still felt to have some potential. But thinking in pressures needed replacement.

Again, I found it somewhat hard to get a firm grasp feeling that whenever I tried to construct anything it was hard to find a reasonable base. I reglanced at a certain class of ray methods for such base. Indeed, they seemed to have a lot in common as they should but, to my surprise, that common was not properly identified — at least not in the acoustics literature I knew about!

Instead of developing a new ray method, I started to work for the base, the common for the ray methods. That common — the unificative abstraction — could perhaps be used to create all new methods, or at least have some analytical value. It certainly would have better value for my personal understanding on acoustics than developing a new ray method.

***

Now, after a lot of work that unificative abstraction is finally constructed — the initial proposition for the acoustic energy propagation theory in terms of rays. Should the theory have any real scientific value, one can only hope. The scientific adventure was still real, that much is certain. But the theory in its initial propositional state is yet to be properly tested — and far from complete...

I still seek new, better theories.

Chapter 1

Introduction

A proper base is fundamental to any work. An unconnected, stand-alone work lacks easily any reasonable base for rationale, and cannot almost certainly be accepted as a scientific work. On the other hand, a work subclassing a more general theory can easily exploit the general results.

Ray acoustic modelling, as I understand it, lacks such proper base. The methods seem to always require specific considerations to support and justify the method. Within the methods, the considerations are often very similar or even equivalent, and thus, they should be provided by a base. By a brief example, we shall elaborate how the base hierarchies in theories could be constructed. After the example, the advantages should be obvious.

Our example takes place in computer program compiling — a field very different from acoustics. Let us assume that there is a proper theory for instruction selection for RISC processors. Very probably, one might easily adapt such theory for instruction selection for specific RISC processors, such as MIPS, Sun SPARC, and Alpha. Possibly, all that is needed is just the explicit specification of the instruction set.

The theory of instruction selection for RISC processors might be a subclass of a theory of more general instruction selection, which might be based on a theory

of tiling of intermediate trees, which might be based on a theory of general tiling of nets, which might be based on set theories, and the chain goes on. The very fundamental results of general tiling of nets would be inherited all the way up in the chain to the special processor-specific instruction selection.

Let us further assume a multi-platform compiler. In the compiler there might be a framework for general tiling of trees, which could be used by a framework of general instruction selection, which could be used by a framework of RISC instruction selection, which could be used by the processor-specific instruction selections. Adding support for a new RISC processor might require only adding proper instruction set semantics descriptions to the compiler — a very small task compared to the stand-alone instruction selection.

Apart from the compiler example, such examples are found very abundantly. In fact, counter-examples of successful theories that are not based on anything are hard to find. Even commonly used axiomatic theories, physical postulates, or other fundamental assumptions are always based on something. Often such something is the fact that the predictions based on the fundamental assumptions give results that are harmonious with experiments. In the development of the theory, the axioms might be based on strong intuition, and only afterwards the predictions might either prove the axioms acceptable, or false.[1]

Study of acoustics has also a strong base, and a lot of that base belongs to the classical physics. Open almost any book on acoustics, and you see Euclidean spaces, energies, vibrations, radiation, wave equations, *etc.*, which are indeed very physical concepts.

Scientific theories, as I see them, have two goals above all: to predict and to explain. These goals are not mutually exclusive, as one must understand the phenomena to make predictions, and understanding the basis of predictions helps

[1] One commonly known such theory is the General Relativity Theory by Albert Einstein. Many aspects that lead to the theory were indeed greatly based on well-justified intuition. Later, he found that the theory successfully predicted the trajectory deviation of Mercury — unexplained by the Newtonian physics — and the bending of light in the presence of a gravity, amongst others.

understanding the phenomena.

Acoustics study is not exceptional in respect of these goals. Possibly, the acousticians have first studied the behaviour of sound in different circumstances, and formed some initial theories. The initial theories were tested by experiments, and either they survived or were rejected. The long process eventually led us to understand sound as wave motion and further, we learned to describe wave motion as solutions to the wave equation.

The wave equation is a very solid tool, or theory. Immediate predictions of it are the constant bounded speed of sound in homogenous medium, and the reflection of the sound when incident to a wall, for example. These predictions are easy to test — clap your hands towards a distant wall, and you will hear an echo after some time.

Given time, predictions have become more sophisticated. Take for example the famous formula of Sabine [25] which evaluates the reverberation time for a room. Improved reverberation time prediction has been then suggested by *e.g.* Kuttruff [17], and further, the prediction has been finessed by *e.g.* ray tracing techniques. There exist also other important *characteristic* predictions.

A few decades ago, we have begun to move beyond characteristic predictions, to form a new study of computational acoustic modelling. Instead of prediction of characteristics, we now ask *how* a sound source sounds to an observer in some acoustic environment. The initial steps were taken at the noon of the computer era, which made it possible to perform complex computations required by acoustic modelling.

Since then, computational acoustic modelling has branched into various substudies, and one of them is ray acoustic modelling — sometimes also referred as geometric acoustic modelling. Ray acoustic modelling itself has a strong base, formed by acoustics, ray optics, and even remote sensing to some extent. Many results are borrowed from a similar but fundamentally somewhat simpler field of computer ray graphics which is also vastly based on ray optics.

Ray acoustic modelling has many special characteristics. Although fundamentally sufficient, acoustics itself is too general to form a practical base for ray acoustic modelling methods. The other strong base, computer ray graphics, generally lacks time dependency in the ray propagation and a continuous spectrum of radiation, which are both of great importance in acoustics. In these respects, computer ray graphics should be more of a subclass of ray acoustic modelling than the other way around.

This work is a proposal towards a base for ray acoustic modelling — the Acoustic Energy Propagation Theory. I have collected some existing results from closely related fields, and formed the initial theory. Some borrowed results were directly applicable (*e.g.* the energy flow concepts, geometrical considerations), and some required a bit modification (*e.g.* the rendering equation). Additionally, there are some suggestions for concepts more specific to the ray acoustic modelling (*e.g.* radiation in various frequencies, temporal intensity algebra). Finally, and perhaps surprisingly, some results which are intuitive using time-dependency can be applied back to time-independent forms (*e.g.* the reflection-iterative construction as a solving method to the rendering equation) where the intuitivity is less obvious.

Now I should answer my own question: what does this initial theory try to predict or explain? To be honest, the initial goal was neither, instead the unification of already existing results for ray acoustic modelling purposes. The already existing results explain and predict themselves. Inherited, the theory tries to explain a simplified version of the propagation of sound in terms of rays, and therefore, open some views for further considerations. Furthermore, by proposing a base for existing methods, I hope that it could provide an inspiration for new powerful methods and analysis tools, as has happened with Kajiya's rendering equation [13].

1.1 Organization of the Thesis

This thesis is organized as follows. In Chapter 2, the field of ray acoustic modelling is introduced along with the base for the modelling, finishing with an introduction to some well-known modelling methods.

The next two chapters consist of the theory, which is divided into the general part and the acoustic part. Chapter 3 is solely dedicated to the general part of the theory, and Chapter 4 is of adapting the general theory for acoustics. The separation is made in the hope that the general part of the theory could be useful in other ray-based radiation propagation modelling fields — especially in computer ray graphics. The chapters also include some discussion on using the theory, and the known limitations of the theory with some suggestions on overcoming the limitations.

In Chapter 4 also the unification of the modelling methods introduced in Chapter 2 is shown. In addition, some suggestions on using the theory as a basis to the new modelling and analysis methods are given.

Chapter 5 concludes this thesis. In that chapter I try to discuss, as objectively as possible, how the theory succeeds in what is expected of it. The conclusion ends with a summary of possibly interesting topics for the related future work.

Two appendices are included. Appendix A introduces briefly the most important mathematical concepts used in the work. Especially, the solid angle integration considerations might be useful for a not-so-deeply mathematically oriented reader.

Finally, Appendix B is a short essay on impulse responses of diffuse reflections, and the very problematic essence of them. I felt important that some consideration on the actual pressure-field realization of diffuse reflections were included in this work that so generously speaks of them.

I do not claim originality on any theorems, proofs, solutions, methods or insights in any specific areas presented in this work. They are all already present in various

fields of study, in one form or another. Nevertheless, no other theoretical construction, targeted this specifically to general ray acoustic modelling, has reached my attention.

From the reader, you, I assume some understanding on calculus and physics. As already hinted, Appendix A reveals the required mathematical skills essential for the complete understanding of this work. Nevertheless, with some vector integration skills the most important parts of this work should be easily reachable. The most important physical concept is the radiance, introduced in Section 3.2.4. Preliminary knowledge of that is not required but [7] and [5] contain a more detailed introduction.

This work is written such that familiarity with the existing ray acoustic modelling methods should not be required to understand the theory. However, the familiarity is still probably required to understand correctly all the motivations behind this work, and to correctly assess the outcome.

Chapter 2

Background

In this section, the background for linear ray acoustic modelling is briefly covered. We shall begin with an introduction to the underlying fundamental physics of the linear sound propagation and the fundamental mathematical equation — the *wave equation*. After the brief introduction, we shall give a glance at the Huygens' Principle, which can be considered as the justification of the ray concept, the foundation of this work.

After the physical considerations, we introduce *acoustics* and the linearity assumptions used in this work. Naturally, a brief discussion will be given upon the inferred consequences of the assumptions.

Finally, we introduce acoustic modelling and the goal of that field. The field is large and thus, we shall restrict the introduction to the general concepts and to a subclass called *ray acoustic modelling*. Of that subclass, a brief introduction to three well-known modelling methods (the image source method, the radiosity method, and the ray tracing method) will be given with a short discussion on what they have in common.

Physical and mathematical foundations for ray acoustic modelling will not be given in this chapter. Instead, the physical foundations are considered as a part

of the theory which is constructed in Chapters 3 and 4. The most important mathematical concepts are introduced in Appendix A.

2.1 Sound and Acoustics

Quoting from the first sentence of the first chapter in [22]:

> *Acoustics* is the science of sound, including its production, transmission, and effects.

The concentration in this work will be on the linear sound and transmission in a linear medium. Very little of actual production, observation, or effects of sound will be discussed.

We start this section by introducing the concept of *sound* as a physical phenomenon, and further, how it can be described using mathematics. Some physical characteristics are also discussed, such as energy in the sound.

This section owes much to Allan D. Pierce's *Acoustics* [22]. Supplementary physical definitions are found in *e.g.* [34]. As so throughoutly used, the citations to these references in the next two subsections are sometimes omitted.

2.1.1 Physics of Sound

In an informal context, the term sound refers to all that is *audible*, all that produces hearing sensations. Often the sound propagates through gaseous substances such as air, but it may also travel through solid or liquid substances such as walls and water. It is also heard in spoken language that sound travels through electric lines when speaking of sound information propagation via phone lines, for example.

But what is sound physically? The natural sound, *i.e.*, the sound that is audible without transducers (the other examples above except the phone line one), is propagation of *disturbances* in a medium — regardless of whether the substance of the

medium is of gaseous, liquid, solid, or of a more exotic form. When the disturbances reach our hearing organs, they are transformed into hearing sensations by our brains. Next we elaborate the term disturbance a little.

Substances have stable states, at least in the macroscopic sense. A stable state is a state with a local energy minimum, a state where nearby states are susceptible to revert to. In gaseous substances, the stable state could be the uniform pressure. Put yourself in an airtight room, and verify by a manometer that the air pressure is constant. The pressure will be constant everywhere after some time, even if you try to change the pressure of only one corner of the room by releasing highly pressurized air from a pressure bottle.

The disturbance in the air pressure is a deviation in the constant pressure. The surrounding stable state pressure "tries" to compensate the difference, which "pushes" the disturbance further away. The result is that the disturbance propagates in the medium. For linear acoustics, the justification to this simplified description is provided by the wave equation, which will be introduced soon.

The disturbance propagates at the *speed of sound*. In gaseous mediums the propagation of a small enough disturbation is often assumed *linear* in respect to other disturbations. Linearity dictates that the disturbances ignore each other and thus, the disturbances may be examined separately. Non-linear mediums exist also[1]. In them, such separation is not generally plausible.

In linear mediums, the time-dependent pressure deviation at some point can be separated into different frequency components by the *Fourier* transformation. Assuming linear reflections, one may consequently separate the examination of propagation into sub-bands, and later, sum the partial results for the complete result. A sub-band consists of all frequency components inside a frequency range. The frequency components can be considered as *vibrations* at certain frequencies. This justifies a common saying that sound is vibration in medium.

[1] In fact, all natural mediums for sound are non-linear. Some mediums with certain conditions can be approximated as linear mediums with only very little approximation error. Sound propagation in air is example of such.

***

The mathematical analysis of sound dates back to the late 17th century, when Isaac Newton published *Principia*. Newton correctly interpreted sound as pressure pulses, but with only very limited mathematical consideration. During the 18th century, d'Alembert, Lagrange, and Euler carried significant contribution to the development of further analysis of wave propagation. This, amongst others, resulted in the 3-dimensional wave equation (by Euler): [22]

$$\bigtriangledown^2 u(x,t) = \frac{1}{\mathbf{c}^2}\frac{\partial^2 u(x,t)}{\partial t^2} \tag{2.1}$$

The typical derivation of the wave equation assumes a grid.[2] Each point in the grid has a potential state u, and the current speed of potential change u'. The potential difference between nearby points cause linear change to the speed of the potential change u''. Such system is often represented by a linear spring-mass system, as illustrated in figure 2.1. By letting the grid point distance close to zero, the wave equation is obtained. Finally, the constant tying the potential state difference and the speed of potential change is the wavespeed constant $\mathbf{c}$. The solutions to the wave equation represent the time-dependent disturbation behaviour given some initial and boundary conditions appropriate to the problem.[3,4]

The Huygens' Principle constructs the concept of moving wave fronts. The analysis is strictly in the wave field itself. The extended version of the principle can be represented by equation

$$\frac{\partial x_p}{\partial t} = \mathbf{v}(x_p,t) + \mathbf{n}(x_p,t)\mathbf{c}(x_p,t) = \mathbf{v}_{\text{ray}} \tag{2.2}$$

where $\mathbf{v}(\cdots)$ is the velocity of the medium, $\mathbf{n}(\cdots)$ is the unit normal of the wave

[2]The *typical* applies to the general wave equation. Specific wave equations, such as the electromagnetic wave equation [5], are derived rather from the field-specific fundamental equations.

[3]A detailed derivation is found in *e.g.* [11]. A derivation specific to sound is found in [22]. The latter contains an examination of the linearity assumptions often used in sound propagation, and their consequences.

[4]Renote that the wave equation is only applicable in *linear* acoustics.

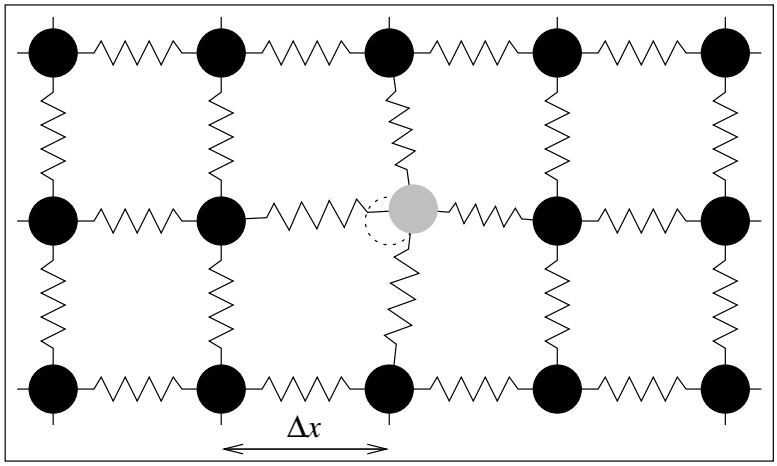

Figure 2.1: A 2-dimensional spring-mass system with displacement

front, and $\mathbf{c}(\cdots)$ is the speed of sound relative to the medium. When this equation is satisfied, $x_p(t)$ represents the *ray path* and $\mathbf{v}_{\text{ray}}$ the ray velocity.[5]

We shall assume stationary, homogenous medium throughout this work. This assumption reduces the equation above to:

$$\frac{\partial x_p}{\partial t} = \mathbf{c}\ \mathbf{n}(x_p, t) = \mathbf{v}_{\text{ray}} \tag{2.3}$$

Further, point sources emit spherical waves. Ignoring edge diffraction, the wave front normal $\mathbf{n}(\cdots)$ of the spherical wave is changed only in reflections. This results in that sound originated from a point source propagates *straightforwardly* in homogenous medium. This is true also more generally with non-point sources, as a source of another shape can be considered as a locally distributed point source.[6]

This *reshaping* may be done by introducing infinitesimal point sources. Such thinking is required in the analysis of reflected radiation. We consider walls reflecting radiation as *radiation sources* in Section 3.3. Essentially these radiation sources are considered to consist of many infinitesimal point sources which receive their energy from the incident radiation.

[5] See [22] for details and derivation.

[6] Examination of the wave equation gives equivalent results.

2.1.2 Energy of Sound

As implied in the previous section, sound carries energy, and the propagation of sound is comparable to the propagation of radiation. In linear acoustics, the energy is carried along the wave fronts, or rays. This is because of the separability of a linear wave field. In this section, we shall briefly justify the use of energies in the modelling instead of pressure disturbations.

The separability of a wave field into linear components (temporal, spatial, and frequential) is fundamental, and derivable from the wave equation[7], and the separability applies also to the energy. The separability of energy may be used to devise a particle concept. For a point source, we can separate the emission to small particles, where each particle is a fundamental piece of waveform. The particles can then be used to describe the initial conditions by a linear combination (or integration). Chosen properly, the fundamental particles do not change shape in propagation and reflections so that the propagated and reflected result cannot be described by a linear combination of the fundamental particles. In that case, the fundamental particles and environmental operations form a closed group.

From the base, we can derive an orthogonal base. For each fundamental particle, we can calculate the energy, and thus, we can calculate the energy of the whole field by summing the energies of the particles forming the field. The propagation can now be considered as particles propagating along ray paths, and reflections can be considered as surfaces absorbing particles and re-emitting one or more or

[7] If $u(x,t)$ is a solution to equation 2.1 for initial conditions $u(x,0)$ and $u'(x,0)$, and $v(x,t)$ is a solution for initial conditions $v(x,0)$ and $v'(x,0)$. Then $w(x,t) = u(x,t) + v(x,t)$ is a solution for initial conditions $w(x,0) = u(x,0) + v(x,0)$ and $w'(x,0) = u'(x,0) + v'(x,0)$. This can be found out by direct substitution to the wave equation, and proves the *spatial* separability.

The *temporal* separability is proven similarly for boundary conditions $w(x,t) = u(x,t) + v(x,t)$. In applications, the borders may be vibrating surfaces (*e.g.* loudspeakers).

The frequential separability is a corollary of the linearity of the field, and elementary property of the *Fourier* transform.

even an infinite number of particles per absorbed particle.[8]

The most important difference to particle physics is that in this particle system, we can freely scale the particles in size, if required. If particle ψ has the unit energy, then particle $\tau\psi$ has energy τ^2 (refer to the energy of wave motion in [34]). Thus, for distance attenuation, it is a matter of taste to consider either that rays "attenuate" or that a smaller number of rays hit an object further away. In addition, we can make a transformation from wave field to energy particles by reconsidering wave particles as energy particles, calculate everything in energies[9] and, after all computation, transform back into wave field representation.

By particle decomposition, it is justified to use energy carrier rays for modelling. The concept of wave field decomposition to particles has further value in discussion of sound propagation of various frequencies and phase effects.[10]

Note that the base formed by the fundamental particles is not specified. An exact specification is not required here. It is sufficient to know that it is possible in principle, and in this way the particle concept is comparable to mathematical distributions [26]. A concrete base could be constructed from an orthonormal wavelet base, for example.

To summarize, for each particle it is sufficient to know its 1) type, 2) position, 3) energy, and 4) state of motion for wave field recomposition. For reflection behaviour it is sufficient to know their absorption and re-emission behaviour on the surfaces. Rays and particles are interchangeable concepts in this context because of the particle size scaling.

[8] The concepts of bases, and representation of linear objects by base elements, can be familiarized in *e.g.* [15].

[9] Reflection calculations are typically much easier to do with energies.

[10] See discussion in Section 5.2.

2.2 Acoustic Modelling

Freely applying the definition of acoustics:

> *Acoustic modelling* is the science of modelling sound, including its production, transmission, and effects.

The very foundation of any typical acoustic modelling is the modelling of sound, whether one models the production, transmission, or effects. Sometimes the modelling of sound is very implicit to the method (the use of precalculated impulse responses, for example), sometimes the acoustic modelling is all about sound modelling (solving numerically the wave equation, for example). This work is placed somewhere between, as the concentration is on the sound propagation modelling.

A typical division of sound modelling based on physics is as follows: 1) models based on solving the actual wave field by *e.g.* solving the wave equation by discretation, or solving Kirchhoff's equations by boundary element methods, and 2) methods that solve some derived property of the wave field, such as energy. Examples of the latter are various ray tracing and radiosity methods.

Models not directly based on physics exist also. Examples of such models are *perceptual* models. They modify sound characteristics such as spaciousness, brightness, and clarity. The characteristics analysis is based on human perception, hence the name.

2.3 Impulse Responses

In linear signal propagation systems, it is typical and adequate to model only the propagation of a single impulse. The impulse propagation information can be used to derive the propagation of an arbitrary signal.

This is strictly because of the linearity. The detection of the propagated signal can be described by a linear operator $\widehat{D}$ such that, for any signal $\sigma(t)$ emitted by the

source, the detection is $\widehat{D}\sigma(t)$. Further, the linear operator $\widehat{D}$ can be represented by a function $d(t)$, where $\widehat{D}\sigma(t)\widehat{=}d(t)*\sigma(t)$, and $*$ is the *convolution operator* [15].[11]

The convolution operator[12] commutes ($a*b=b*a$). In addition, $\sigma(t)=\delta(t)*\sigma(t)$, where $\delta(t)$ is the Dirac delta functional. Hence, the detection of an arbitrary signal $\sigma(t)$ can be written as

$$\begin{aligned}\widehat{D}\sigma(t)\widehat{=}&d(t)*\sigma(t)\\ =&\delta(t)*d(t)*\sigma(t)\\ =&\sigma(t)*(d(t)*\delta(t))\\ \widehat{=}&\sigma(t)*\widehat{D}\delta(t)\end{aligned} \tag{2.4}$$

which justifies the modelling of only impulses.

The detected impulse $\widehat{D}\delta(t)$ is referred as the *impulse response*. Impulse response analysis is widely used in acoustics and other signal processing related fields, see for example [19, 29, 8].

Depending on the context, the impulse may refer to the pressure disturbation impulse or the energy impulse. By the impulse response we refer to the disturbation impulse response of the system, and by the energy response, we refer to the energy impulse response, respectively.

2.4 Ray Methods

Ray methods form a subset of acoustic modelling methods, and this subset is fundamentally based on modelling the propagation of wave fronts. As introduced briefly in Section 2.1.1, wave front propagation is equivalent to ray propagation,

[11] The symbol $\widehat{=}$ means here exactly $\exists\widetilde{I}:\ \widehat{D}\sigma(t)=\widetilde{I}(d(t)*\sigma(t))$ where $\widetilde{I}$ is a homomorphic mapping.

[12] $(a*b)(t)=\int a(\tau)b(t-\tau)\mathrm{d}\tau$

as rays should be understood as propagation directions of specific points in wave fronts. In linear acoustics with point sources, this is an unambiguous definition.

Typical of ray methods is that they model rather the energy of the sound than the pressure disturbations. This is due to reflection modelling: reflections are much easier to model by using energy concepts than pressure concepts. There exists, however, an important exception. When the reflections are assumed to be always specular (see lead-in to def. 3.2.11), the ray is simply "bounced" from the surface with some possible attenuation. The lack of reflection scattering makes it easy to perform the necessary calculus using pressures.

Methods that do not assume specular reflections consider almost always only energies. A fundamental problem in pressure disturbation modelling — and quite a difficult one — is the reflection scattering. It is hard to model and measure accurately. Even more it is hard when the reflection is not only from a small surface area but from a larger area.

For example, consider a distant surface. Let us assume that we know the impulse response for that surface. If the surface is distant enough, the impulse response of the reflection is approximately peak-formed, as the distance between various points in the surface is almost constant. What about replacing the surface with a surface with twice the area? The pressure disturbation field, at least, does not get twice the amplitude of the original, as that would imply a quadratic change in the energy. A more general answer easily leads to that one would be required to do quite heavy calculations for the determination of the impulse responses, *e.g.* to solve the Kirchhoff's equations by using a boundary element method.[13]

Energies, on the contrary, are much easier. Twice the size, twice the reflected energy — at least approximately. For many purposes, such approximate results are adequate.

In the following subsections, three conventional ray modelling methods are introduced. We start with the vanilla image source method (ISM), which is used to

[13] Similar considerations are presented in *e.g.* [14].

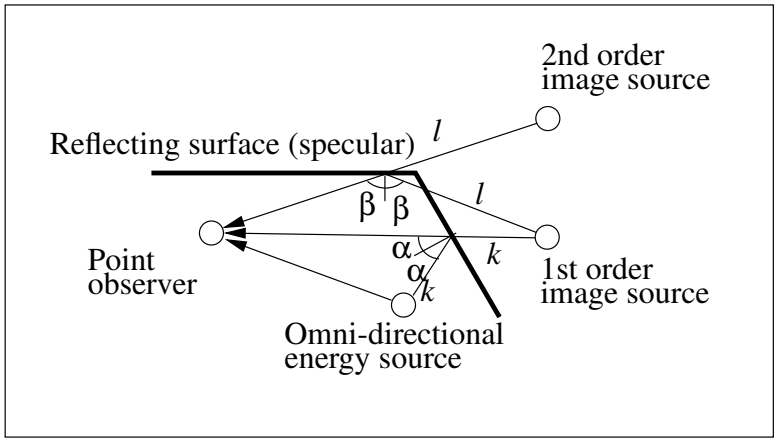

Figure 2.2: Direct, first, and second order image sources. Not all possible image sources are shown.

model the propagation in specular reflecting environments. For the other extreme — ideal diffuse reflections — the radiosity method is a typical choice. Finally, we introduce the ray tracing method which can be used with more general reflections.

2.4.1 Image Source Method

The image source method (ISM) [1, 4] is a method originally devised for modelling sound propagation via specular reflections on planar surfaces. More modern approaches also incorporate edge diffraction to some extent [28].

The image source method is a geometrical method. It uses simple[14] geometrics to determine the virtual positions of the sound source. A virtual position — image source — represents a ray path from the source to the observer. The virtual positions, applied by appropriate filters for the effects of reflections, are ready for auralization. The essence of the beauty lies in the calculation of the image sources, of which we introduce a simplified version.

Let us assume a setup as in figure 2.2. The direct sound is obvious. The first order

[14] Albeit the choice of word, the author nevertheless does not wish to undervalue the beauty!

possible paths are easy to come by, when keeping in mind that specular reflections obey the mirror reflection requirement. That is, the elevation angle of the incident and exitant rays relative to the surface normal must be equal, and the azimuth angles must be the opposite (180 degree difference). Thus, there is at most one point per a planar surface to examine. The image source is placed such that the vector from the observer to the reflection point is extended by the length of the vector from the reflection point to the source.

The beauty is in the calculation of higher order image sources. When determining nth order sources, one must only consider the image sources of order $n-1$. The same rule applies — consider only reflection points in the surfaces, where the specular reflection requirement exist. When the reflection point is found, create a new image source in the direction of the reflection point and the distance increased as with the 1st order case.

In real implementations the situation is somewhat more complicated, however, because not all image sources need to be visible, *i.e.* represent an unobstructed ray path. Non-visible image sources can even create visible image sources in higher order image source calculation. This increases the implementational complexity, see *e.g.* [27] for details.

The edge diffraction extensions consider also the edges of the surfaces as possible reflection points but these reflections do not need to satisfy the specular reflection requirement. Also, diffuse reflection image sources do not need to satisfy the mirror reflection requirement.

The relaxation of the mirror reflection requirement imposes a weakness to the image source method. Edge diffraction and diffuse reflection create *fuzzy* image sources — image sources that do not exist at one well-defined point. Instead, they are distributed to continuous volumes. Fuzzy image sources are also present with non-planar surfaces. A fuzzy image source, distribution, must be approximated with a sufficient amount of point image sources for results of any accuracy. Thus, diffuse reflections and edge diffractions are heavy in respect of computation time, compared to specular reflections. Without edge diffraction or diffuse reflections

the resulting sound resembles often the flutter echo acoustics of a bathroom, without even mentioning that all-specular reflections seldom exist in nature.

The number of image sources is of order $\mathcal{O}(k^n)$, where k is the average number of reflections per order, and n is the number of successive reflections. In practice, this limits the examination to at most three or four consecutive reflections.[15] Some optimization techniques, however, might be used to decrease the complexity. Such optimizations could try to aggressively merge nearby image sources, for example.

2.4.2 Radiosity Method

The family of radiosity methods, originating in thermodynamics, consists of powerful methods for radiation propagation computation in diffuse reflective environments. Below, a brief introduction to the radiosity method is given. For further details, see *e.g.* [5] for electromagnetic radiosity, and *e.g.* [21] for a modern version of the radiosity method for sound.

The geometry in the radiosity method is usually divided into small surface areas — *patches* — and the communication between the patches is examined. The communication analysis is typically such that if we know the exitant energy flow $\Phi_k(t)$ for some patch P_k, we examine the reflection of that energy flow in other patches to determine their exitant energy flows caused by the exitant energy flow $\Phi_k(t)$. Further, one can do such analysis for every patch to get a set of linear equations. Calculating the exitant energy flows in every patch in the system is then simply a matter of solving a linear system.[16]

The simple communication analysis is possible because the ideal diffuse reflection is *memoryless*, *i.e.* the exitant radiation pattern does not depend on the incident

[15] For average 100 reflections per order, and 3 successive reflections, there is an order of $100^3 = 1\,000\,000$ image sources of 3rd order.

[16] Albeit simple in theory, solving such system with currently available technology is far from trivial when the number of patches reaches the orders of thousands and millions.

pattern.[17] Because of the memoryless nature, the computational complexity is typically much less than that of the image source method. The communications can be represented by a matrix A. Thus, if we have a vector F of initial (primary, zero-order) exitant radiosities in patches, the first order of reflected radiosity can be calculated by operating F with A, noted by AF, and the nth order reflections with A^nF. Finally, summing the different orders, we obtain:

$$P = \sum_{n=0}^{\infty} A^n F \tag{2.5}$$

Alternatively, one can begin with Kajiya's rendering equation [13] and substitute the reflection distribution function with a memoryless function (constant for ideal diffuse reflections), to get a vector equation of form:

$$P = AP + F \tag{2.6}$$

Then it is a matter of finding an energy flow vector that satisfies the equation. For construction details, see for example [7].

For later, we remark the equivalency between the two radiosity methods. A simple reorganization of equation 2.6 yields

$$P = (I - A)^{-1} F \tag{2.7}$$

and by assuming $\|A\| < 1$ (which is typically the case in any realistic problem, refer to sec. 3.3.5), one can solve this by the Neumann series [15, 11]:

$$P = \left(\sum_{n=0}^{\infty} A^n \right) F \tag{2.8}$$

which is obviously equivalent to equation 2.5.

To emphasize, in acoustics we use time-dependent energy flows. Consequently, the vectors P and F are not in $\mathbb{R}^n$ but in $(\mathbb{R} \to \mathbb{R})^n$ (vector of functions), and

[17] This implies, that the reflection distribution function for such surface is *separable* to incident sensitivity pattern and exitant radiation pattern, *i.e.* $f_{\mathrm{r}}(\Omega_i, \Omega_e) = f_{\mathrm{r,i}}(\Omega_i) f_{\mathrm{r,e}}(\Omega_e)$. See derivation of definition 3.2.10 for details.

similarly, the members of A contain shifting operations for propagation delay. Therefore, the items in the vectors, at least, require additional discretation, and the actual computational complexity of equation 2.5 is not simply of linear class in respect to the reflection order.

The radiosity method is well-applicable only for memoryless reflections when used in general geometries. It is possible to extend the radiosity model for general reflections but then it would not be enough to simply compute the patch-patch communications but patch-patch-patch communications instead. This is because the reflection angles are required. Then, the size of the communication matrix, however, poses easily unrealistic memory and computational requirements for currently available hardware.

2.4.3 Ray Tracing Methods

The prologue of ray tracing in acoustic modelling was set up in the dawn of the computer era. In 1958 Allred and Newhouse used the Monte Carlo integration method[18] to model architectural acoustics [2, 3], and after about a decade, Krokstad *et al.* published an article in terms of ray tracing [16].

Essentially, the Monte Carlo integration method and the ray tracing method are equivalent. The Monte Carlo integration method is a numerical method to calculate integrals. Shortly, the integration can be written as

$$\int_a^b f(x)\mathrm{d}x \approx \frac{b-a}{\operatorname{card} X} \sum_{x\in X} f(x) \tag{2.9}$$

where X is a random finite set with uniform distribution in range $[a,b]$, or more generally

$$\int_A f(x)\mathrm{d}x \approx \frac{1}{\operatorname{card} X} \sum_{x\in X} \frac{f(x)}{\rho(x)} \tag{2.10}$$

where X is a random finite subset of A, and the distribution probability density is defined by ρ.

[18] See definition A.2.4 in Appendix A

In ray tracing, the set A contains typically the ray paths, or possible directions of a ray, and the probability measure adjusted properly to take account the reflection patterns. Ray tracing can be done from the viewer to the sources, or vice versa. The vanilla ray tracing cannot cope with point sources in from-viewer model, or with point viewers in from-source tracing. Extensions exist to overcome such limitations [9].

***

The equivalency between the ray tracing and the Monte Carlo integration can be easily understood by a little fictional setup involving light rays, with apparent analogy to sound rays. Let us assume that there is a narrow-beam detector pointing towards an ideally diffuse reflective surface, and that there are light-emitting surfaces which cast radiation to the reflective surface. A portion of the reflected energy is detected by the detector.

Now, the detector essentially measures the incident energy flow from a small area in the reflective surface. In Section 3.2.4 we will learn that the detected brightness of a diffuse surface is independent of the orientation of the surface. We will also learn that the brightness of the surface is dependent only on the incident energy flow, or irradiance, to the surface. The irradiance, detected by the detector, can be written as (sec. 3.2.4 and eq. 3.19)

$$\mathrm{I_i} = \int_{2\pi} L_\mathrm{i}(\Omega_\mathrm{i}) \cos\theta_\mathrm{i} \mathrm{d}\Omega_\mathrm{i} \tag{2.11}$$

where L_i is the incident radiance[19] to the surface caused by the emitting surfaces, Ω_i is the unit vector direction of the incident radiance, 2π is the hemisphere surface of directions, and θ_i is the angle between the surface normal and the incident radiance direction. This can now be approximated by the Monte Carlo method using equation 2.10 with the uniform distribution:

$$\mathrm{I_i} \approx \frac{1}{\mathrm{card}\, X} \sum_{\Omega_i \in X} \frac{L_i(\Omega_i)\cos\theta}{\frac{1}{2\pi}}, \qquad X \subset 2\pi \tag{2.12}$$

[19] Radiance is a form of radiation distribution.

where X is a random finite set in the hemisphere. The sum elements can now be understood as individual rays of different strengths arriving in different directions to the detected surface area.

The equation 2.12 has also a physical meaning. Comparing the incident and reflected intensities yields reflectances which in turn yields the detection probability of a single photon arriving from a radiating wall to the reflective patch. Thus, there exists physical equivalence in addition to the mathematical.

Chapter 3

Simplified General Energy Propagation Theory

In this chapter, we shall construct a theory which unifies the ray methods described in Section 2.4 and subsections. The theory will be subsequently referred as the general energy propagation theory (GEPT) and, as implied by the name, it is of general energy propagation. By straightforward specialization, it is usable as *e.g.* an acoustic energy or a light energy propagation theory.

We begin the construction by the environmental (sec. 3.1) and physical (sec. 3.2) definitions fundamental to the theory. We then construct two single-band energy propagation models (sec. 3.3). As the name suggests, the single-band models are usable in modelling the energy propagation of radiation at a single frequency, or radiation in a homogenously behaving frequency band. Logically proceeding, we will also discuss briefly utilizing the single-band models in the analysis of full spectrum energy propagation, where the behaviour of the propagation depends on the frequency (sec. 3.4).

Concluding this chapter, we identify the limitations of the simplified theory with some suggestions on possible extensions to overcome them (sec. 3.5).

3.1 Environment

Radiation propagation requires an environment, and consequently, modelling requires an environment. The modelling environment consists of the space of the environment, and objects in the space. The surfaces of the objects constitute the *modelling geometry*, or geometry in short. We will define these concepts formally in the following subsection.

General environments are not well-suited for computability. Therefore, we present two subclasses of geometries suitable for computation — polyhedral environments and discrete environments. These environments can be used to approximate general environments.

3.1.1 Fundamental Definitions

We begin the fundamental definitions with the formal definition of the modelling space and geometry, in addition to some closely related important concepts. We also define *visibility*, which is an important concept in radiation propagation modelling.

Definition 3.1.1 *Modelling Space*

For obvious physical reasons, we choose three orthogonal spatial dimensions and one orthogonal temporal dimension as the modelling space. The spatial dimensions are represented by $\mathbb{R}^3$ and the temporal dimension by $\mathbb{R}$.

The modelling space is subject to all typical definitions of Euclidean (or geometric) space. For positions, we use vectors with the common representation:

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \tag{3.1}$$

The elementary vector concepts are presented in Appendix section A.1.

Definition 3.1.2 *Direction, Direction Space, Solid Angle*

Directions are represented by unit vectors (def. A.1.3), and sets of directions are represented by sets of unit vectors. The direction space is thus the set of all unit vectors — the surface of the unit sphere.

Solid angle, by the very definition, is the area measure of a geometrical object projected on the unit sphere surface. However, it is a common practice to use the term "solid angle" for projected areas as well, not only for area measures. We shall also follow this practice.

A point in the unit sphere surface represents a direction from the centre of the sphere to the point in the surface. The point is freely exchangeable with the respective unit vector. A general area in the unit sphere represents a set of directions, and the full surface cover represents the set of all directions.

In this work, we shall use the following notations:

ω	solid angle, set of directions
Ω	an infinitesimal solid angle, or a single direction
θ	the elevation component of Ω
ϕ	the azimuth component of Ω

The solid angle notation is always absolute, whereas the elevation and azimuth components are relative to the reflection surface in reflection calculus. There, the elevation component θ is the angle between the surface normal and the direction Ω, and the azimuth component ϕ is the torsion angle of Ω and some chosen reference orientation in the reflection surface.

When the elevation components are used to parametrize the direction, we use the following notation:

$$\Omega = (\theta, \phi) \tag{3.2}$$

The parentheses are omitted in function parameters for clarity.

Solid angle integration is defined as integration over the surface of the unit sphere. The following notation is used:

$$\int_{\omega} f(\Omega)\mathrm{d}\Omega \tag{3.3}$$

where ω is the solid angle (area in the unit sphere surface), $f(\cdots)$ is a direction-dependent function and $\mathrm{d}\Omega$ is the surface area measure.

For notational convenience, $\omega = 4\pi$ denotes the full sphere, and $\omega = 2\pi$ denotes the hemisphere on the *outerior* surface side.

See Appendix section A.2 for elementary techniques of solid angle integration.

Definition 3.1.3 *Modelling Geometry*

The environment geometry, denoted by $\mathcal{G} \subset \mathbb{R}^3$, represents all surface points in the environment. For almost everywhere[1] the following two related functions must be defined:

The reflection function:

$$f_{\mathrm{r}}(\Omega_{\mathrm{i}}, \Omega_{\mathrm{e}};\, a), \quad a \in \mathcal{G} \tag{3.4}$$

The unit normal:

$$\mathbf{n}(a) \in \mathbb{R}^3, \quad |\mathbf{n}(a)| = 1, \quad a \in \mathcal{G} \tag{3.5}$$

The unit normal points out from the *outerior* surface side. The reflection function will be defined later in definition 3.2.8.

The geometry is said to be *static*, if $\mathcal{G}$ and consequently $\mathbf{n}(\cdots)$, and f_{r} are invariant under time.

Definition 3.1.4 *Visibility Function*

[1] f is defined almost everywhere in $\mathcal{G}$ if f is defined in A, and $\int_{\mathcal{G}-A}$=0. f can thus be safely extended to all $\mathcal{G}$ by setting $f = C$ for all $\mathcal{G} - A$ where C is some bounded value, for example 0.

The visibility function $\nu(x_1, x_2)$ defines visibility of two points in space:

$$\nu(x,y) = \begin{cases} 1, & \nexists \tau \in (0,1) : \ (1-\tau)x + \tau y \in \mathcal{G} \\ 0 & \text{otherwise} \end{cases} \tag{3.6}$$

x is visible to y if no point in the geometry intersects with the segment of line between x and y. Note that $\nu(x,y) = \nu(y,x)$ (trivial to proof).

Definition 3.1.5 *Visible Geometry (Inverse Visibility Function)*

The inverse visibility function (per point) evaluates all points in the geometry visible at point x:

$$\nu^{-1}(x) = \{y \in \mathcal{G} \mid \nu(x,y) = 1\} \tag{3.7}$$

Definition 3.1.6 *Visible Point at Direction (Inverse Projection)*

The visible-point-at-direction function $\nu_p(x, \Omega)$ evaluates the visible point that resides in direction Ω from point x:

$$\nu_p(x,\Omega) = \nu^{-1}(x) \bigcap \{\tau\,\Omega |\ \tau > 0\} \tag{3.8}$$

In closed geometries ν_p is always defined in all points inside the geometry. Generally, ν_p is not necessarily defined in all points for all directions.

***

The three visibility-related functions ν, ν^{-1}, and ν_p are required in the derivation of the propagation models. The inverse projection function ν_p may be awkward to use inside non-closed geometries. However, every bounded geometry can be closed by inserting a large enough spherical surface around the geometry, and setting full absorption ($f_\mathrm{r} \equiv 0$) to the enclosing surface, which makes the surface "invisible" to any reflections.

3.1.2 Polyhedral Environments

Any sufficiently smooth geometry can always be approximated by polyhedra [6]. Surfaces of polyhedra consist of planar polygons. Because of this, polyhedral environments are suitable for *e.g.* the image source method.

The polyhedra approximation has, however, some important drawbacks. Polyhedra surfaces are continuous but non-smooth in the edges. This implies discontinuities in the surface normals, which may distort significantly the modelling with sharply angle-dependent reflections such as the mirror reflection (eq. 3.2.11). For example, [6] contains a more detailed insight.

3.1.3 Discrete Environments

When the reflection function is smooth enough everywhere, the surface geometry can be approximated by a finite set of points. A point represents a portion of the modelling geometry. The area of the represented geometry is the *weight* of the point.

Integration over the geometry is then reduced to a sum:

$$\int_{\mathcal{G}} f(a)\mathrm{d}a \approx \sum_i w_i f(a_i) \tag{3.9}$$

where w_i:s denote the weights and a_i:s the discretized surface points.

Discrete environments resemble somewhat the Monte Carlo integration with predetermined random function. The concept of discrete environments in acoustic ray tracing is presented in [10].

Discrete environments do not apply well in geometry approximation where mirror-reflective surfaces are present. Nevertheless, they offer often good approximation with diffuse reflective surfaces. Patch geometries, commonly used in the radiosity methods, are essentially discrete environments.

3.1.4 Visibility Computation

Fast algorithms exist for implementing the visibility functions, which is of importance in geometries of at least medium complexity. Examples of implementations are the binary space partition (BSP) algorithm and the octree algorithm. [30]

For a superficial example of a visibility algorithm implementation, we take the BSP algorithm. In that algorithm the space is hierarchically divided into subspaces. The subspace division is performed by defining a plane which splits the greater space in two subspaces. The whole modelling space is represented by the top node of a BSP tree, and subspaces by the child nodes. It is quite straightforward to browse the neighbouring subspaces for visibility calculations, or precalculate visibility between subspaces. In the latter case, complexity class $\mathcal{O}(\log N)$ or even $\mathcal{O}(1)$ for the evaluation is achievable, where N is the triangle count in the geometry.

3.2 Radiation and Reflection

In this section, we shall examine radiation and reflection in linear medium. In the following subsections, some important concepts are defined — energy flow in subsection 3.2.1, energy source in 3.2.2, observation and detection in 3.2.3 and 3.2.6. The observation is defined by intensities and the detection of propagated energy by radiances, hence the split-up. Radiance and reflection are examined in subsections 3.2.4 and 3.2.5.

In the examination, we ignore the effects of diffraction, which are generally important, especially with longer wave lengths. We return briefly to the subject in Section 3.5.1. Before rushing into the definitions, let us first glance at the fundamental ideas behind the definitions.

***

Let us submit into a small thought game. Imagine a small omni-directional energy source in free space, which emits energy waves continuously. Picture there a small particle riding a wave. The riding particle travels along the wave front, and therefore, it has the direction and speed of the propagating wave. The path of the particle is called a ray.

Next, picture a great number of small particles emerging simultaneously from the source — each in a different direction. The riding particles form a spherical shell, which grows in time due to the particles distancing from the source. Again, the particles follow straightforward paths — rays.

The small energy source uses energy to create the wave motion. If the energy source stops creating the waves, the consumption of energy stops. When it restarts, the energy consumption restarts also. The spherical shells of propagating wave fronts carry the energy *emitted* by the source.

Consider now that the source emits a short omni-directional burst of propagating waves. As the spherical shell formed by the burst is homogenous in form, it is justified to think that the energy is distributed homogenously in the shell. In time, the shell grows as the wave front distances, but no additional energy is gained or lost in the shell. The only logical conclusion is that the energy density must decline such that the energy of the shell is constant. Thus, the distancing wave front is subject to propagation attenuation, or decay.

Further, figure that the short burst emitted was populated by riding particles. In the beginning the shell is small in radius and the particle density is high. In time, that density lessens. It is not hard to picture or show that the particle density is connected to the energy density in a linear fashion.

Assume yet a small observer in free space. The observer can measure the strength of the passing wave just as one can measure the level of the water surface by using a float. If the observer so wishes, it can also observe the number of riding particles passing a small area to estimate the strength of the wave.

***

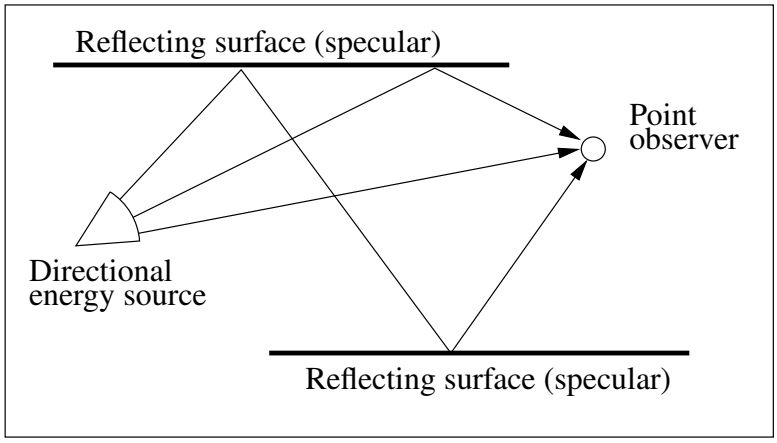

Figure 3.1: Possible ray paths in specular reflecting environment

In our imaginary space one can exchange wave fronts with particles, and call the particle paths rays. In free space, this exchange is readily valid. The exchange was also justified in Section 2.1.2, and by that justification we shall deduct that the exchange is valid even in a more general fashion. However, generally in environments with non-empty geometries there exists also diffraction which bends the paths of the particles. As previously stated, we shall ignore this now and discuss it briefly in subsection 3.5.1.

Let us consider an emitting source, and two smooth, stiff, reflective walls as illustrated in figure 3.1. The particle path represents the propagation of the small piece of the wave front — a ray, that is. When the particle strikes the wall and reflects back in the same elevation angle as in the incidence but to the opposite azimuth angle — just as a rubber ball bounces when thrown hard at a stiff smooth wall — the reflection is called *specular*, or mirror reflection.

In polyhedral environments with only specular reflections, the possible ray paths from the source to the observer are quite easy to find by using the image source method (sec. 2.4.1), and importantly, the number of ray paths is finite when the maximum number of successive reflections is finite. Thus, if we know how the ray is changed in the reflections, we can calculate the observation by applying

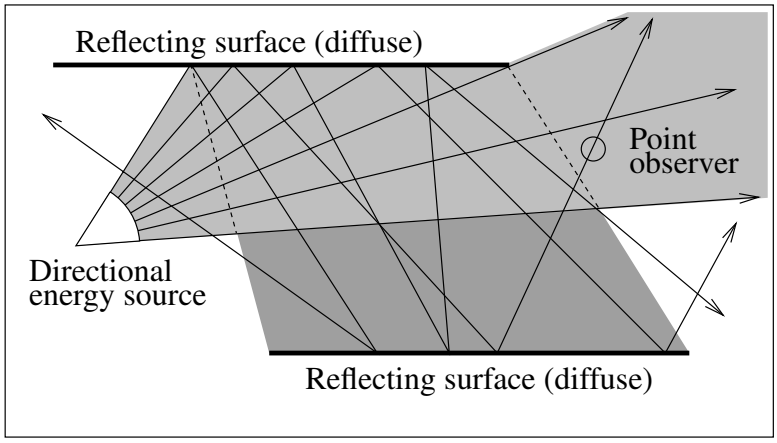

Figure 3.2: Possible ray paths in diffuse reflecting environment

distance attenuation and propagation delay per ray, and sum the rays together.

The *diffuse* reflection is very different from the specular reflection. When a particle strikes a diffuse reflective surface, it may reflect into any direction. For an analogy, consider a rubber ball thrown at a bumpy stiff surface. The diffuse reflective surface has an exitant ray distribution for the incident ray. The ray distribution can be considered as a probability distribution for the exitant particle direction.

Because of this one-to-infinite reflection behaviour, a single particle may have infinite number of paths from the source to the observer even in a simple setup, see figure 3.2. This suggests other means for calculating the observation, instead of the simple ray path sum — especially as any single ray path carries infinitesimal energy intensity.[2]

If the reflection is truly diffuse, the exitant ray does not remember the previous path in reflection, *i.e.* the exitant direction does not depend on the incident direction. It is easy to calculate the intensity of the incident energy flow at any point in the reflective surface caused by a direct source (see fig. 3.3). Further, this infor-

[2]This is easily seen by the one-to-infinite relation. The average intensity of a single ray is $\lim_{k\to\infty} \frac{I_0}{k} = 0$. However, the sum remains: $\lim_{k\to\infty} \sum_{i=1}^{k} \frac{I_0}{k} = I_0$.

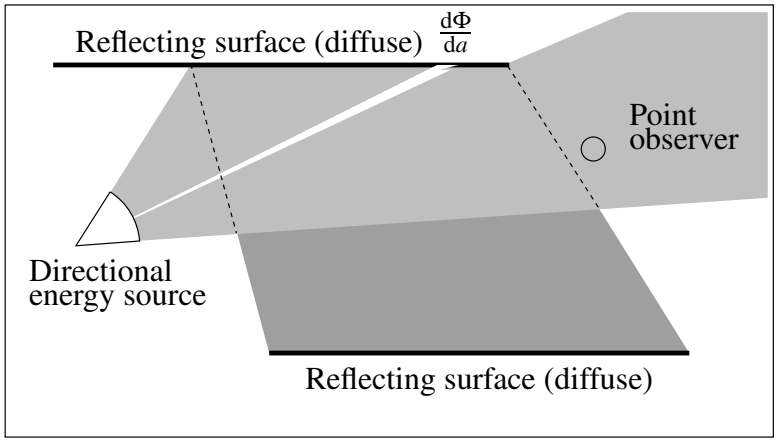

Figure 3.3: Incident energy from a point source to a small surface patch

mation is sufficient to calculate the energy intensity in the other surfaces caused by the reflected energy flow because of the memoryless nature of the diffuse reflection.

As noted above, the exitant flow is a distribution in diffuse reflections. This leads to a certain symmetry in calculation of the incident energy flow in other surfaces caused by the reflection. Namely, that the reflected flow is a distribution dictates that the incident flow in any surface, caused by the reflection, is also a distribution. The total incident energy flow can be evaluated by integrating over the incident directions, see figure 3.4.

When generalizing a bit, one may further define the exitant intensity distribution as a function of the incident intensity distribution for arbitrary reflections. Specifically, specular reflections can be expressed also as such distributions. *Radiance* is one form of energy distribution (defined in sec. 3.2.4), and BRDFs (defined in sec. 3.2.5) define exitant distribution per incident distribution in radiances.

Using terms of reflected radiance, one may straightforwardly construct an energy propagation model (as we do in sec. 3.3.3) to calculate the energy propagation from source to observer. By knowing the energy propagation, we can calculate the detection.

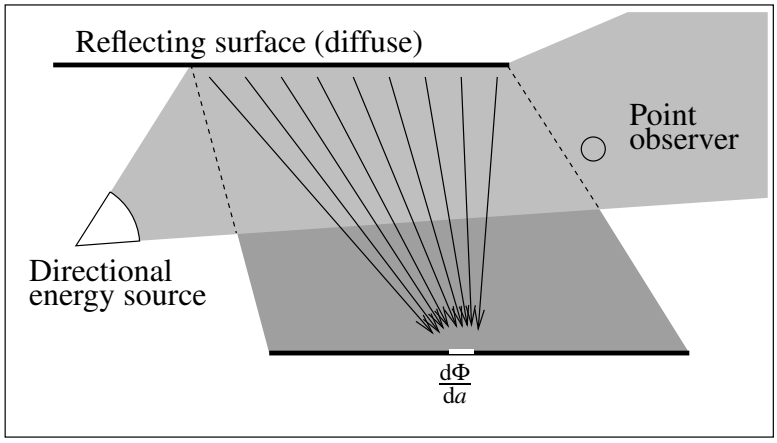

Figure 3.4: Radiant incident energy to a small surface patch

In addition to constructing a propagation model, one may use the definitions to instruct a balance equation. Whereas the straightforward construction gives straightforward result, the solution to the balance equation specifies the energy propagation.[3]

One such balance equation is the time-independent rendering equation by James T. Kajiya [13] transformed in radiance form [7]. The general requirement in Kajiya's rendering equation is such that, for each surface point, the incident and produced intensity (per point and direction) must equal to the absorbed and exitant intensity. We extend the rendering equation for time dependency in Section 3.3.4.

3.2.1 Energy Flow

In this section we define the energy transfer concepts. These elementary definitions are used in later constructions.

[3] Compare with the wave equation

Definition 3.2.1 *Energy Flow* Φ

Energy flow is the rate of energy transfer between two objects. The unit is $\left[\frac{\mathrm{J}}{\mathrm{s}}\right] = [W]$.

Definition 3.2.2 *Speed of Wave Fronts* **c**

The unit of the speed is $\left[\frac{\mathrm{m}}{\mathrm{s}}\right]$.

In homogenous linear medium, the speed of wave fronts is constant. This can be verified from the wave equation [11].[4]

The speed of the wave fronts also defines the speed of the energy flow, consider the particles riding waves in the thought game in Section 3.2. The speed of the energy flow is fundamental in the construction of the energy propagation models.

Definition 3.2.3 *Intensity of the Energy Flow*

The intensity of the energy flow defines the energy flow per surface area:

$$\mathrm{I} = \frac{\mathrm{d}\Phi}{\mathrm{d}a} \tag{3.10}$$

The unit is $\left[\frac{\mathrm{W}}{\mathrm{m}^2}\right]$.

Remark 3.2.4 *Intensity in Non-orthogonal Incidence*

If one measures an energy flow that has intensity I_0 in the orthogonal cross-section of the energy flow, one measures intensity

$$\mathrm{I} = \mathrm{I}_0 \cos\theta \tag{3.11}$$

in a planar surface, where θ is the angle between the normal vector of the surface and the direction of the energy flow. This can be verified as follows. If the planar

[4] Any $f(x+vt)$ where $|v| = \mathbf{c}$ is a solution to the linear homogenous medium wave equation.

surface has area A, the projection on the cross-section has area $A\cos\theta$. By the energy conservation principle and by noting that the intensity is homogenous, one finds that $\mathrm{I}A = \mathrm{I}_0 A\cos\theta$. After mutual division by A, equation 3.11 is obtained. Thus, the intensity of the ray is "stretched" along the non-orthogonal surface.[5]

For a surface, there are two types of energy flows — incident and exitant. The incident energy flow is the energy flow coming into the surface. The exitant energy flow is the flow emerging from the surface, such as the energy flows emitted and reflected by the surface. The small Latin letters $_{\mathrm{i}}$ and $_{\mathrm{e}}$ in the subscripts of other symbols accentuate this direction. For example, Φ_{e} denotes the exitant energy flow.

3.2.2 The Energy Source

An energy source emits radiation. Some energy sources emit energy homogenously in every direction, while some energy sources emit different amounts of energy in different directions. We assume the latter for generality.

In our model, the energy source is a point somewhere in the modelling space, and it does not have inherent geometry. However, real-life sources have always some geometry. We consider such geometry as a part of the environment geometry.

This said, we can form the definitions very straightforwardly. For simplicity, we do not assume any orientation of the energy source. Instead, the orientation of the source is embedded into the directional emittance pattern.[6]

Definition 3.2.5 *Energy Source*

The energy source is at position denoted by x_s. The emittance pattern p_{e} and the total emitted energy flow Φ_s determine the intensity I_0 of the energy at the unit

[5] More detailed considerations are found in *e.g.* [5] and [34].

[6] To allow rotational orientation, one simply has to apply a rotational transformation to the emittance pattern.

distance, such that

$$\mathrm{I}_0(x) = \frac{\Phi_s}{4\pi}\, p_\mathrm{e}(x - x_s) \tag{3.12}$$

where x is at the unit distance from the source, and the measurement plane is orthogonal to the radiation, *i.e.* the normal of the surface points to the source.

Next we shall derive the necessary requirements to the emittance pattern. The exitant energy flow is obtained by integrating the intensity of the energy flow over a surface enclosing the source (Gauss's law [34]). We choose the surface of a sphere with radius r as such enclosing surface. By requiring energy conservation, we get the left and the middle side equality of

$$\Phi_s = \int_{|a-x_s|=r} \mathrm{I}_0(a)\,\mathrm{d}a = \int_{|a-x_s|=r} \Phi_s p_\mathrm{e}(\frac{a-x_s}{|a-x_s|})\frac{1}{4\pi r^2}\,\mathrm{d}a \tag{3.13}$$

By substituting the intensity by equation 3.12, and by accounting the surface area change compared to the surface area of the unit sphere, we get the right side. Multiplying the left and the right sides by $\frac{4\pi}{\Phi_s}$, we obtain:

$$\int_{|a-x_s|=r} p_\mathrm{e}(\frac{a-x_s}{|a-x_s|})\frac{1}{r^2}\mathrm{d}a = 4\pi \tag{3.14}$$

By reparametrizing the integral (rem. A.2.1) the requirement may be written as:

$$\int_{|a|=1} p_\mathrm{e}(a)\mathrm{d}a = 4\pi \tag{3.15}$$

Thus, for example, for homogenous (omni-directional) source $p_\mathrm{e} \equiv 1$.

The examination above yields also that the intensity of the energy from a point source is subject to distance attenuation:

Remark 3.2.6 *Distance Attenuation*

$$\mathrm{I}(x) = \frac{\Phi_s p_\mathrm{e}(\frac{x-x_s}{|x-x_s|})}{4\pi\,|x-x_s|^2} = \frac{\mathrm{I}_0(x_s + \frac{x-x_s}{|x-x_s|})}{|x-x_s|^2} \tag{3.16}$$

where I_0 is the intensity at the unit distance from the source.

3.2.3 The Observer

The observer is a small object in the modelling space, position denoted by x_o. The observer does not interfere with the energy propagation and if such effect is required, one must modify the geometry around the observer accordingly.

The observer may have direction-dependent sensitivity, *i.e.*, it may be more sensitive to energy flows arriving in some directions than in others. Further, the observer may even detect radiation from different directions differently, such as in different channels. Consider *e.g.* multi-channel recording, where microphones are directed towards different directions.

The detection is defined by the *directional transfer function*, denoted by dtf. dtf is a linear operator in respect to the intensity, mapping the intensity of the energy flow and direction into *detection*. What the detection exactly means, is observer-specific.

Definition 3.2.7 *Detection Transfer Function*

$$\begin{aligned} &\mathrm{dtf}(\Omega, \mathrm{I}) && \text{directional transfer function} \\ &\mathrm{dtf}(\Omega, \alpha_1 \mathrm{I}_1 + \alpha_2 \mathrm{I}_2) = && \text{linearity of detection} \\ &\quad \alpha_1 \,\mathrm{dtf}(\Omega, \mathrm{I}_1) + \alpha_2 \,\mathrm{dtf}(\Omega, \mathrm{I}_2) \end{aligned} \tag{3.17}$$

The detection of passing intensity will be extended to the detection of passing radiance in subsection 3.2.6. See also Section 4.2 for an example of the implementation of the detection transfer function.

3.2.4 Radiance

Whenever radiation strikes a surface, there is an incident energy flow to that surface. The strength of the energy flow is dependent on the intensity and the angle of the incident radiation. Moreover, surfaces often receive energy in multiple directions. We begin the examination of the incident energy flow in a single direction.

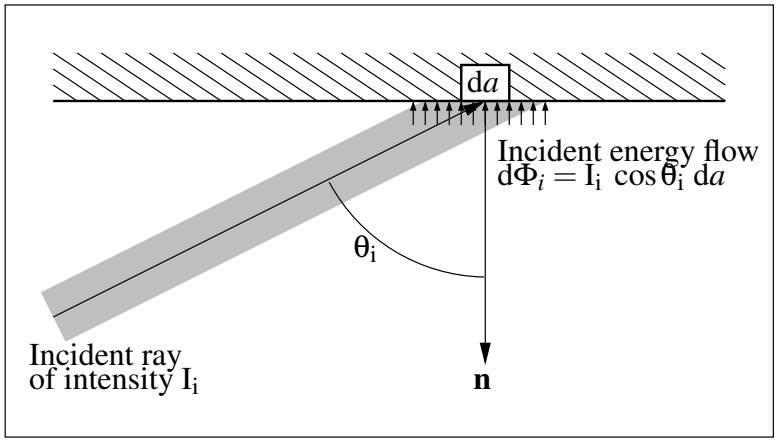

Figure 3.5: Energy flow of incident radiation

Using the examination as an introduction, we then form the concept of *radiance* which defines the "brightness" of the surface.

Let us assume that there is incident radiation to a planar surface. The intensity of the radiation in the surface is $I_i \cos\theta_i$ where I_i is the intensity of the radiation beam (rem. 3.2.4). Incident energy flow for a small piece of surface, patch, is proportional to the intensity and the size of the patch. Thus, we get equation (see fig. 3.5 and join def. 3.2.3 with rem. 3.2.4)

$$\frac{d\Phi_i}{da} = I_i \cos\theta_i \tag{3.18}$$

for incident energy per patch. The incident energy is also referred as irradiance. Further, the incident radiation does not need to come only in one direction. Instead, it may appear as a distribution of directions (see fig. 3.4). By direct distribution in directions, we get the following equation for patch irradiance:

$$I_i = \frac{d\Phi_i}{da} = \int_{2\pi} L_i(\Omega_i) \cos\theta_i \, d\Omega_i \tag{3.19}$$

There, Ω_i is the direction of the incident radiation, and θ_i the angle between Ω_i and the surface normal. L_i is the intensity distribution (see [5, 20, 7] for details). When observing the incident radiation in direction Ω_i, the integral equation becomes

after mutual division of $\mathrm{d}\Omega_\mathrm{i}$ and the cosine-term:

$$L_\mathrm{i}(\Omega_\mathrm{i}) = \frac{\mathrm{d}^2\Phi_\mathrm{i}}{\mathrm{d}a\ \mathrm{d}\Omega_\mathrm{i}\ \cos\theta_\mathrm{i}} = \frac{\mathrm{dI}_\mathrm{i}}{\mathrm{d}\Omega_\mathrm{i}\ \cos\theta_\mathrm{i}} \tag{3.20}$$

L_i as presented is called *incident radiance*.

The exitant radiance L_e is constructed similarly with equal definitions.

3.2.5 Reflection

Most surfaces do not absorb all incident radiation. Instead, they *reflect* a portion of the incident radiation back to the space. Depending on the material and the type of radiation, different amounts of radiation are reflected in different directions, when radiation strikes a certain point in the surface in a certain direction. In some materials, even the point of the exitant reflection may differ from the point of incidence.[7]

If the radiation does not propagate beneath the surface, the reflective material is called *rigid*, *i.e.*, the incident and exitant points are always equal in the reflection. The contrary is *non-rigid*, sometimes also referred as *porous*. We shall construct our models using strictly rigid materials. However, in Section 3.5.2, we shall discuss briefly extending the construction for non-rigid materials.

The reflection behaviour of a rigid material can be described with a BRDF (Bidirectional Reflection Distribution Function) [20, 7], denoted by symbol f_r. A BRDF maps an incident radiance distribution to an exitant distribution:

Definition 3.2.8 *Total Reflected Radiance, BRDF*

$$L_\mathrm{e}(\Omega_\mathrm{e}) = \int_{2\pi} f_\mathrm{r}(\Omega_\mathrm{i}, \Omega_\mathrm{e}) L_\mathrm{i}(\Omega_\mathrm{i}) \cos\theta_\mathrm{i}\ \mathrm{d}\Omega_\mathrm{i} \tag{3.21}$$

The BRDF (denoted by f_r) defines thus the reflected radiance distribution L_e for the incident radiance distribution L_i. When the material and the orientation of

[7] Consider a piece of almost transparent plastic rod. By illuminating the rod in one end with a powerful enough light source, the whole rod begins to gloom.

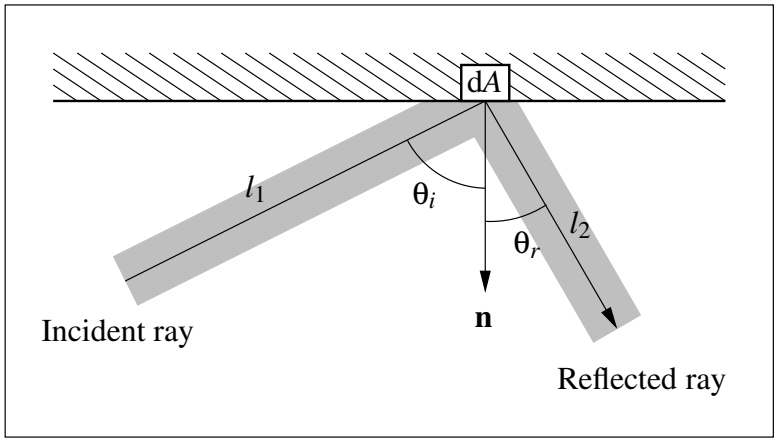

Figure 3.6: Reflection of planar beam

the surface are implicitly known, the defined form of f_r is adequate. If not, we explicitly display the surface point a as a parameter to the BRDF:

$$f_\mathrm{r}(\Omega_\mathrm{i}, \Omega_\mathrm{e};\ a) \tag{3.22}$$

Also, when discussing reflection of radiation at different frequencies, we explicitly add the frequency parameter f. Thus, the full position and frequency dependent form is:

$$f_\mathrm{r}(\Omega_\mathrm{i}, \Omega_\mathrm{e};\ a;\ \mathrm{f}) \tag{3.23}$$

The BRDF in our definitions is dependent on the surface orientation. The direction parameters Ω_i and Ω_e range in outerior surface hemisphere, and hence, surfaces of the same material in different orientations require different BRDFs. In real applications, however, it is usually better to define the material-dependent BRDF in some reference orientation, and pass the direction parameters through a rotation operator.

From the definition of the BRDF and the irradiance equation (eq. 3.19) one can derive *biconical reflectance factor* (in directional form): [20, 7]

Definition 3.2.9 *Biconical Reflectance Factor*

$$\rho(\omega_i,\ \omega_e) = \frac{1}{\int_{\omega_i} \cos\theta_i d\Omega_i} \cdot \int_{\omega_e} \int_{\omega_i} f_r(\Omega_i,\ \Omega_e) \cos\theta_i \cos\theta_e \ d\Omega_e d\Omega_i \quad (3.24)$$

The biconical reflectance factor evaluates the average reflectance, when radiation is incident in a set of directions ω_i and is reflected in a set of directions ω_e.

The biconical reflectance factor may be used to calculate the absorption properties of the BRDF. Set $\omega_e = 2\pi$ to get *total reflectance* $\beta(\omega_i)$ for a set of directions ω_i. The absorption factor for incident radiation is readily calculable: $\alpha(\omega_i) = 1 - \beta(\omega_i)$.

The total reflectance also sets the necessary requirements for BRDF of any realistic material. The total reflectance $\beta(\omega_i)$ must be in range $[0,1]$ for all ω_i. Reflectance below zero would imply that some reflected radiation has *negative* energy. Reflectance above one would imply that the material reflects more energy than it receives. This is possible only by using amplifiers that receive additional energy from some other source.

The biconical reflectance factor may also be used in indirect measurement of the BRDF, as the reflectance factors for different incidence and reflection direction sets are straightforward to measure — in principle, at least. It is then an optimization problem to find the most plausible BRDF that fits into the measurements [20]. A description of an exemplary BRDF measurement setup utilizing reflectance measurements is found in [24].

***

We will define yet the two most used reflection idealizations — diffuse and specular. They were informally introduced in the introduction to radiation and reflection. In diffuse reflection (fig. 3.2) we assume that any incident radiation for a small surface area is reflected with the same radiance in each direction. This means that the detected surface brightness does not depend on the orientation.

Specular reflection (fig. 3.1) is the opposite in this aspect, as the mirror reflection is assumed there. Mirror reflection dictates that the radiance incident in direction (θ, ϕ) is reflected only in direction $(\theta, \phi \pm \pi)$. We begin the mathematical formulation with the ideal diffuse reflection.

Because the exitant radiance is invariant in respect to the exitant direction, diffuse reflection is said to be memoryless. That is, the shape of the exitant radiance distribution does not depend on the shape of the incident radiance distribution. This quickly leads to that the BRDF of the diffuse reflection must be separable into the incident and the reflected parts:

$$f_{\mathrm{r,d}}(\Omega_\mathrm{i}, \Omega_\mathrm{e}) = f_{\mathrm{r,inc}}(\Omega_\mathrm{i}) f_{\mathrm{r,refl}}(\Omega_\mathrm{e}) \tag{3.25}$$

The requirement that all incident directions are equivalent leads to that

$$f_{\mathrm{r,inc}} \equiv C_1 = \beta \tag{3.26}$$

because

$$\frac{\mathrm{d}\Phi_\mathrm{e}}{\mathrm{d}a} = \beta \frac{\mathrm{d}\Phi_\mathrm{i}}{\mathrm{d}a} \tag{3.27}$$

$$\int_{2\pi} f_{\mathrm{r,inc}}(\Omega_\mathrm{i}) L_\mathrm{i}(\Omega_\mathrm{i}) \cos\theta_\mathrm{i} \mathrm{d}\Omega_\mathrm{i} = \beta \int_{2\pi} L_\mathrm{i}(\Omega_\mathrm{i}) \cos\theta_\mathrm{i} \mathrm{d}\Omega_\mathrm{i} \tag{3.28}$$

and, if $f_{\mathrm{r,inc}}$ was non-constant, radiation in certain directions would be more important than radiation in other directions (see also eq. 3.11 and eq. 3.19). By similar reasoning, we deduct that $f_{\mathrm{r,refl}} \equiv C_2$ is constant. To calculate the value of C_2, consider the energy flow equation:

$$\frac{\mathrm{d}\Phi_\mathrm{e}}{\mathrm{d}a} = \int_{2\pi} f_{\mathrm{r,refl}}(\Omega_\mathrm{e}) \beta \frac{\mathrm{d}\Phi_\mathrm{i}}{\mathrm{d}a} \cos\theta_\mathrm{e} \mathrm{d}\Omega_\mathrm{e} = \beta \frac{\mathrm{d}\Phi_\mathrm{i}}{\mathrm{d}a} \int_{2\pi} C_2 \cos\theta_\mathrm{e} \mathrm{d}\Omega_\mathrm{e} \tag{3.29}$$

Combine this with equation 3.27 to get: [8]

$$C_2 = \frac{1}{\int_{2\pi} \cos\theta_\mathrm{e} \mathrm{d}\Omega_\mathrm{e}} = \frac{1}{\pi} \tag{3.30}$$

Now, by combining equations 3.26 and 3.30 into 3.25 we get:

[8] $\int_{2\pi} \cos\theta \mathrm{d}\Omega = \int_0^{\frac{\pi}{2}} \left[\int_0^{2\pi} \cos\theta \mathrm{d}\phi \right] \sin\theta \mathrm{d}\theta = 2\pi \int_0^{\frac{\pi}{2}} \cos\theta \sin\theta \mathrm{d}\theta = 2\pi \Big/_0^{\frac{\pi}{2}} \frac{1}{2} \cos^2\theta = \pi$

Definition 3.2.10 *BRDF of Ideally Diffuse Reflection,* $f_{\mathrm{r,d}}$

$$f_{\mathrm{r,d}} \equiv \frac{\beta}{\pi} \tag{3.31}$$

where $\beta \in [0,1]$ is the reflection coefficient. It is easy to verify that β is really the total reflectance for any incident radiance distribution (def. 3.2.9).

The ideal specular reflection assumes that the reflected radiance is the same (with reflectance β accounted) with the incident radiance, where the incident direction is *mirror*-transformed. We use the symbol M to denote the transformation operator (def. A.1.6). Formally:

$$L_{\mathrm{e}}(\omega) = \beta L_{\mathrm{i}}(M\omega) \tag{3.32}$$

The angular form of the mirror operator M:

$$M(\theta,\phi) = (\theta, \phi \pm \pi) \tag{3.33}$$

Note that $L_{\mathrm{e}}(\omega) = \beta L_{\mathrm{i}}(M\omega)$ is equivalent to $L_{\mathrm{e}}(M\omega) = \beta L_{\mathrm{i}}(\omega)$. This is because $M = M^{-1}$. [9]

From equation 3.21 it is obvious that the BRDF of the specular reflection must contain the Dirac delta functional. Unfortunately, the delta functional is dependent on the parametrization of the integral. Thus, different definitions for different parametrizations. We formulate the BRDF for angular and solid angle parametrizations, and to do that, we require the following properties of the Dirac delta functional: [32, 20] (see also rem. A.2.3 in Appendix section A.2)

$$\delta(x) = 0, \quad \text{for any } x \neq 0 \tag{3.34}$$

$$\int_A \delta(x)\mathrm{d}x = 1, \quad \text{when } 0 \in A \tag{3.35}$$

$$\int_A \delta(x-a)f(x)\mathrm{d}x = f(a), \quad \text{when } a \in A \tag{3.36}$$

$$\delta(f(x)) = \sum_{\mathrm{i}} \frac{\delta(x-x_{\mathrm{i}})}{|f'(x_{\mathrm{i}})|}, \quad \text{where } x_{\mathrm{i}}\text{:s are the roots of } f \tag{3.37}$$

[9] $L_{\mathrm{e}}(\omega) = \beta L_{\mathrm{i}}(M\omega) \Leftrightarrow L_{\mathrm{e}}(M\omega) = \beta L_{\mathrm{i}}(MM\omega) = \beta L_{\mathrm{i}}(\omega)$

Let us go back to the requirement equation 3.32 and parametrize the hemisphere surface in equation 3.21 by angular parameters (eq. A.11) to obtain condition:

$$L_e(\theta_e,\phi_e) = \beta L_i(\theta_i,\phi_i \pm \pi) = \int_0^{\frac{\pi}{2}} \left[\int_0^{2\pi} f_{r,s}(\theta_i,\phi_i,\theta_e,\phi_e) L_i(\theta_i,\phi_i) d\phi \right] \cos\theta \sin\theta d\theta \quad (3.38)$$

One possible BRDF that satisfies the condition is: [20]

$$f_{r,s}(\theta_i,\phi_i,\theta_e,\phi_e) = 2\,\beta\,\delta(\sin^2\theta_e - \sin^2\theta_i)\,\delta(\phi_i - \phi_e \pm \pi) \quad (3.39)$$

This is straightforward to verify.[10] Remark, that there exists other equivalent forms because of the property revealed in equation 3.37.[11]

The solid angle (or directional) form of the BRDF is easier to derive. Directly from equations 3.21 and 3.32 we get:

$$f_{r,s}(\Omega_i,\Omega_e) = \frac{\beta}{\cos\theta_i}\,\delta(\Omega_i - M\Omega_e) = \frac{\beta}{\Omega_i \cdot \mathbf{n}(a)}\,\delta(\Omega_i - M\Omega_e) \quad (3.41)$$

We are now ready to state the definition:

[10] By direct substitution of f_r as in eq. 3.39 to eq. 3.21 we get

$$\begin{aligned} L_e(\theta_e,\phi_e) &= \iint 2\beta\delta(\sin^2\theta_e - \sin^2\theta_i)\delta(\phi_e - \phi_i \pm \pi) L_i(\theta_i,\phi_i)\cos\theta_i \sin\theta_i d\theta_i d\phi_i \\ &= 2\beta \int \delta(\sin^2\theta_e - \sin^2\theta_i) L_i(\theta_i,\phi_e \pm \pi)\cos\theta_i \sin_i\theta_i d\theta_i \\ &= \beta \int \frac{\delta(\theta_e - \theta_i)}{\cos\theta_e \sin\theta_e} L_i(\theta_i,\phi_e \pm \pi)\cos\theta_i \sin\theta_i d\theta_i \\ &= \frac{\beta}{\cos\theta_e \sin\theta_e} L_i(\theta_e,\phi_e \pm \pi)\cos\theta_e \sin\theta_e \\ &= \beta L_i(\theta_e,\phi_e \pm \pi) \end{aligned}$$

showing the satisfaction of eq. 3.32.

[11] For example, [7] displays the specular BRDF as:

$$f_{r,s} = \frac{\delta(\cos\theta_i - \cos\theta_e)}{\cos\theta_i}\delta(\phi_i - \phi_e \pm \pi) \quad (3.40)$$

By noting that $\cos\theta_i = \cos\theta_e$ when $\delta(\cos\theta_i - \cos\theta_e) \neq 0$ it is easy to show that this BRDF is also a solution.

Definition 3.2.11 *BRDF of Ideally Specular Reflection,* $f_{\mathrm{r,s}}$

In the angular form:

$$f_{\mathrm{r,s}}(\theta_\mathrm{i},\phi_\mathrm{i},\theta_\mathrm{e},\phi_\mathrm{e}) = 2\,\beta\,\delta(\sin^2\theta_\mathrm{e} - \sin^2\theta_\mathrm{i})\delta(\phi_\mathrm{i} - \phi_\mathrm{e} \pm \pi) \tag{3.42}$$

In the solid angle form:

$$f_{\mathrm{r,s}}(\Omega_\mathrm{i},\Omega_\mathrm{e};\ a) = \frac{\beta}{\Omega_\mathrm{i}\cdot\mathbf{n}(a)}\,\delta(\Omega_\mathrm{i} - M\Omega_\mathrm{e}) \tag{3.43}$$

where M is the mirror reflection transformation. Again, it is easy to verify that β really is the total reflectance (def. 3.2.9).

In real life applications it is traditional to use the idealized reflection models instead of the measured. The reason is that the special properties of the idealizations allow the use of fast, specific modelling methods: the image source method for specular-only reflections (sec. 2.4.1) and the radiosity method for diffuse-only reflections (sec. 2.4.2).

3.2.6 Detection

Finally, we shall discuss briefly the observation of the surrounding radiance, thus extending the detection of intensities passing through the infinitesimal observer. Referring to definition 3.2.7, the detection transfer function is linear in respect to the second parameter (intensity). To derive the detection for radiance, we differentiate the detection transfer function against directions — or point solid angles:

$$\frac{\partial\,\mathrm{dtf}(\Omega,\mathrm{I}(\Omega))}{\partial\Omega} = \frac{\mathrm{dtf}(\Omega,\partial\mathrm{I}(\Omega))}{\partial\Omega} = \mathrm{dtf}(\Omega,L(\Omega)) \tag{3.44}$$

The partial derivative operator can be moved inside the $\mathrm{dtf}(\cdots)$ because of the linearity of the operator. Note specifically the absence of $\cos\theta$ term. The surface of the observation object can always be considered orthogonal against the incident radiation. If not, the cosine term would appear with the intensity ($\mathrm{I}_0\cos\theta$, non-orthogonal intensity) but the radiance-irradiance relation would cancel the $\cos\theta$ factor, however.

Integrating over all directions, we get:

Definition 3.2.12 *Total detection*

$$d = \int_{4\pi} \mathrm{dtf}(\Omega, L(\Omega))\mathrm{d}\Omega \tag{3.45}$$

3.3 Energy Propagation Equations

Based on the reasoning in Section 2.3, we shall concentrate strictly on modelling the propagation of the unit energy impulse. The reasoning assumes linear system response. That linearity of the system will be shown after the construction of the propagation equations in subsection 3.3.5.

Before the construction, some common definitions are formulated. In subsection 3.3.1, a small algebra for impulse propagation delay and medium absorption is defined, and in subsection 3.3.2 is defined the *reflection kernel*. After the common definitions, the propagation equations are built: the reflection-iterative construction in subsection 3.3.3, and the acoustic rendering equation in subsection 3.3.4. The equivalence between the propagation equations is shown in subsection 3.3.5.

The section is concluded by a brief remark on radiance propagation in subsection 3.3.6. The remark reminds us that the propagation equations consider only the detection of the reflected but not the direct energy.

3.3.1 Temporal Intensity Algebra

Temporal intensity algebra (TIA) is a simple tool for propagation delay and linear medium absorption analysis. The temporal intensity algebra is derived by using intensities. Afterwards, we straightforwardly state its usability in radiances. The derivation begins with the examination of propagating planar wave fronts.

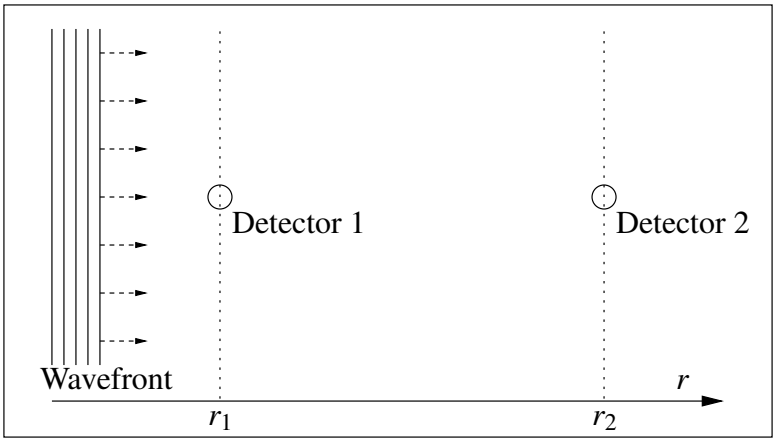

Figure 3.7: Intensity measurement of planar propagating wave front

A planar propagating wave front does not change shape in propagation, and thus it is not subject to the propagation attenuation because of growing, as happens to spherical wave fronts. It is, nevertheless, affected by the medium absorption. For examination, let there be a setup of two intensity measurement points along the path of the wave front, as in figure 3.7.

If one measures the time-dependent intensity $\mathrm{I}_1(t)$ at detector 1, one can predict the intensity measurement at detector 2 simply by shifting the intensity curve by propagation delay, and by scaling the curve down by propagation absorption, resulting in

$$\mathrm{I}_2(t) = \mathrm{I}_1(t - \frac{r_2 - r_1}{\mathbf{c}})\mathrm{e}^{-\alpha(r_2 - r_1)} \tag{3.46}$$

where $\mathbf{c}$ is the speed of the propagating wave front and α denotes the absorption coefficient characteristic to the medium.[12] One could also define an operator for propagation, such that:

$$\mathrm{I}_2(t) = \hat{S}_{r_2 - r_1}\mathrm{I}_1(t) \tag{3.47}$$

It would be easy to see that $\hat{S}$ is linear and that $\hat{S}_{r_1}\hat{S}_{r_2} = \hat{S}_{r_1 + r_2}$ holds as obviously is required. However, for our purposes, this is not yet the most optimal form

[12] Proportional propagation absorption in homogenous medium is of form $\mathrm{e}^{-\alpha r}$ where α is the absorption coefficient and r is the propagation distance [34].

achievable. Remember that it is adequate to model the propagation of the impulse, and everything else can be derived from it by convolution.

Let us then assume that a planar wave impulse is created at a planar source at time 0. So, at the source:

$$\mathrm{I}(t) = \delta(t) \tag{3.48}$$

From some distance r at the source the intensity is measured as

$$\mathrm{I}(t) = \delta(t - \frac{r}{\mathbf{c}})\mathrm{e}^{-\alpha r} = \mathrm{e}^{-\alpha r} S_r \delta(t) = \mathrm{e}^{-\alpha t \mathbf{c}} S_r \delta(t) \tag{3.49}$$

where S is as $\hat{S}$ but without the medium attenuation term.[13] We see that the propagation absorption term may be written in a form where it is dependent only on time. This suggests that the propagation absorption can be analyzed separately.

Based on this brief introduction, we now formulate the algebra and then show the necessary properties.

Definition 3.3.1 *Temporal Intensity Algebra*

The primitive element in the algebra is $\hat{\ell}(t)$ which represents the intensity of the propagating planar wave impulse in a non-absorptive linear medium. In addition, we define the propagation operator S_r which stands for propagation effects for propagated distance r.

The $+$ operator and scalar multiplication are defined as

$$(\beta_1 \hat{\ell}_1 + \beta_2 \hat{\ell}_2)(t) = \beta_1 \hat{\ell}_1(t) + \beta_2 \hat{\ell}_2(t) \tag{3.50}$$

and the propagation operator S_r is defined by:

$$S_r \hat{\ell}(t) = \hat{\ell}(t - \frac{r}{\mathbf{c}}) \tag{3.51}$$

This yields immediately the necessary additivity property in the propagation:

$$S_{r_1} S_{r_2} \hat{\ell}(t) = S_{r_1} \hat{\ell}(t - \frac{r_2}{\mathbf{c}}) = \hat{\ell}(t - \frac{r_1 + r_2}{\mathbf{c}}) = S_{r_1 + r_2} \hat{\ell}(t) \tag{3.52}$$

[13] Note that $r = \mathbf{c}t$. [34]

The impulse energy detection from the source via a simple reflection can be represented as

$$\hat{\ell}(t) = S_{r_2}\beta S_{r_1}\hat{\ell}_s(t) = \beta S_{r_1+r_2}\delta(t) \tag{3.53}$$

where r_1 is the distance from the source to the reflection point, β is the reflection coefficient, r_2 is the distance from the reflection point to the detection point, and $\hat{\ell}$ is the detected intensity in a non-absorptive medium.

The intensity $\hat{\ell}(t)$ in a non-absorptive medium can be transformed to intensity $\mathrm{I}(t)$ in an absorptive linear medium by the *medium absorption operator* H^α as

$$\mathrm{I}(t) = H^\alpha\hat{\ell}(t) = \mathrm{e}^{-\alpha t\,\mathbf{c}}\hat{\ell}(t) \tag{3.54}$$

where α is the medium absorption coefficient such that

$$\alpha = -\frac{\ln\frac{\mathrm{I}(r)}{\mathrm{I}(0)}}{r} = \frac{\ln\frac{\mathrm{I}(0)}{\mathrm{I}(r)}}{r} \tag{3.55}$$

where $\mathrm{I}(r)$ is the measured intensity after propagation r and $\mathrm{I}(0)$ is the measured intensity at the reference point.

***

The necessary linearities are straightforward to show. In essence, it is required that any combination of impulse responses produce linear detection response. This is shown in two parts:

1. The linearity of detection with simultaneous responses:

$$\begin{aligned} &\mathrm{I}_1(t) = H^\alpha\hat{\ell}_1(t), \qquad \mathrm{I}_2(t) = H^\alpha\hat{\ell}_2(t) \\ &\qquad\Rightarrow \\ &\mathrm{I}_1(t) + \mathrm{I}_2(t) = H^\alpha\hat{\ell}_1(t) + H^\alpha\hat{\ell}_2(t) = H^\alpha(\hat{\ell}_1 + \hat{\ell}_2)(t) \end{aligned} \tag{3.56}$$

   This shows that one can postpone the medium absorption into the evaluation of the detected intensity with concurrent responses.

2. The validity of medium absorption postponing when the energy impulse is propagated in the medium and reflected:

$$\begin{aligned} \mathrm{I}(t) = B\delta(t) &= \beta \mathrm{e}^{-\alpha t\,\mathbf{c}}\delta(t - \frac{l}{\mathbf{c}}) \\ &= H^{\alpha}\beta S_r \delta(t) \end{aligned} \tag{3.57}$$

where B is a reflection operator. Also, after two reflections B_1 and B_2 the time-dependent intensity is:

$$\begin{aligned} \mathrm{I}(t) = B_1 B_2 \delta(t) &= \beta_1 \beta_2 \mathrm{e}^{-\alpha t\,\mathbf{c}}\delta(t - \frac{r_1 + r_2}{\mathbf{c}}) \\ &= H^{\alpha}\beta_1 \beta_2 S_{r_1} S_{r_2} \delta(t) \end{aligned} \tag{3.58}$$

The medium absorption effects may again be postponed, and inductively, after any number of reflections, the medium absorption can always be postponed into the evaluation of the detected intensity.

Thus, one can always separate the medium absorption and propagation analysis in impulse response analysis in linear acoustics with simple reflections.

Remark 3.3.2 *Temporal Intensity Algebra and Radiance*

As radiance is a distribution of intensity in directions, the temporal intensity algebra is valid also with radiances. In the radiance case, the algebra element is denoted by $\ell(t)$ — the intensity element without the circumflex.

Later in this work, ℓ is actually a function of location (x), direction (Ω), and time (t). Parameters are often omitted and assumed implicit when the explicity is not required.

3.3.2 Geometry Term and Reflection Kernel

We define here the time-invariant and time-dependent geometry terms and the reflection kernel. The definitions are justified in the following section (construction of def. 3.3.6). The geometry terms represent the effects of propagation and

non-orthogonal incidence and exitance. The reflection kernel is a two-point reflection transport operator which contains the time-dependent geometry term, visibility, and the BRDF. The reflection kernel is used both in the reflection-iterative construction (sec. 3.3.3) and in the acoustic rendering equation (sec. 3.3.4). The time-invariant geometry term is a part of the time-invariant rendering equation.

Definition 3.3.3 *Geometry Terms g and $\hat{g}$*

$$\hat{g}(a,x) = \left[\mathbf{n}(a)\cdot\frac{x-a}{|x-a|}\right]\left[\mathbf{n}(x)\cdot\frac{a-x}{|a-x|}\right]\frac{1}{|a-x|^2} \tag{3.59}$$

$$g(a,x) = \left[\mathbf{n}(a)\cdot\frac{x-a}{|x-a|}\right]\left[\mathbf{n}(x)\cdot\frac{a-x}{|a-x|}\right]\frac{S_{|a-x|}}{|a-x|^2} = S_{|a-x|}\hat{g}(a,x) \tag{3.60}$$

The geometry term $\hat{g}$ is the Kajiyan geometry term [13, 7] used in the time-invariant radiosity methods. The time-dependent geometry term g is the Kajiyan geometry term appended with the propagation delay operator S. In the geometry terms, a and x are points in the geometry: a is the point in the geometry which conducts the incident radiation, and x is the reflection point of the incident radiation. The $[\mathbf{n}(a)\cdot\ldots]$ and $[\mathbf{n}(x)\cdot\ldots]$ terms[14] take into account the non-orthogonality of the incident and reflected rays (rem. 3.2.4). The distance attenuation and propagation delay term $\frac{S_{|a-x|}}{|a-x|^2}$ incorporates the effects of radiation propagation in a lossless medium. Note that g is reciprocal: $g(a,x) = g(x,a)$.

Definition 3.3.4 *Reflection Kernel* R

The reflection kernel consists of the reflection function f_r, the visibility function ν, and the geometry term g, and is written as

$$\mathrm{R}(a,x,\Omega) = \nu(a,x)g(a,x)f_r(\frac{a-x}{|a-x|},\Omega;\, x) \tag{3.61}$$

where a is the surface point of the incident radiance, x is the *reflection* point, and Ω is the direction of the reflected radiance, see figure 3.8. The reflection kernel

[14] Graphics literature typically uses the *cosine* term instead. Keep in mind the following equivalency: $\cos\angle(\vec{x},\vec{y}) = \frac{\vec{x}\cdot\vec{y}}{|\vec{x}||\vec{y}|}$. We prefer the dot-product form because 1) it is closer to the fundamental definitions of the space, 2) due to computability, and 3) to avoid the pole singularities of the angular form (rem. A.2.3).

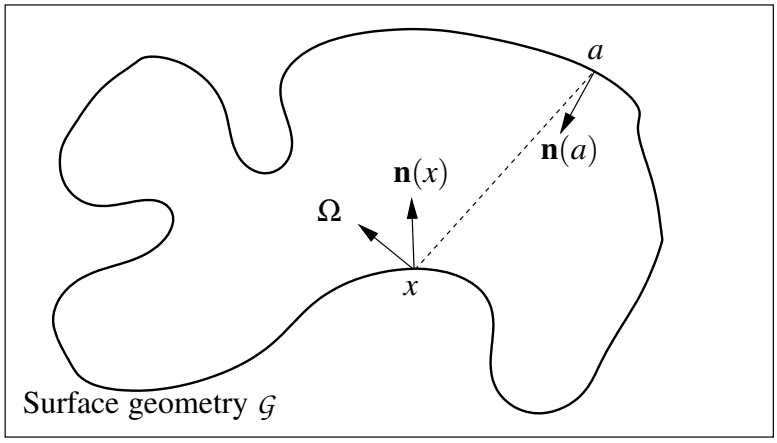

Figure 3.8: Parameters of the reflection kernel

evaluates the first-order reflected radiance at the surface point x when using the integration equation

$$\ell_1(x,\Omega) = \int_{\mathcal{G}} \mathrm{R}(a,x,\Omega)\ell_0(a,\frac{x-a}{|x-a|})\mathrm{d}a \tag{3.62}$$

where ℓ_0 is the exitant primary radiance at surface point a. ℓ_1 denotes the time-dependent reflected radiance. This equation will be constructed in detail in the following section.

There is a seeming issue with the distance attenuation in the reflection kernel and specular reflection. Assume distance r between the point source and the reflection point, and the same distance between the reflection point and the energy detector. It is obvious that the distance attenuation should be of form $\frac{1}{(2r)^2}$ but, quickly looked, it seems to be of form $\frac{1}{2r^2}$. We look into this in Section 4.3.1.

3.3.3 Reflection-iterative Construction

In the construction, we consider one stationary source and one observer placed in a static environment.[15] By the linearities of the wave field, as considered in Section 2.1.2, one may analyse each source separately. The multiple observers case is also straightforward: each detection can be performed independently, as the detection does not interact with the wave field. Medium absorption is ignored in the construction and applied afterwards, as suggested in Section 3.3.1. The positions of the source and observer are denoted by x_s and x_o.

For a primer, we begin with the detection of the direct sound. The impulse intensity response observation equation may be readily written in TIA form:

$$d_0(t) = \mathrm{dtf}(\frac{x_o - x_s}{|x_o - x_s|},\ H^{\alpha}\nu(x_o, x_s)\frac{S_{|x_o - x_s|}}{4\pi\,|x_o - x_s|^2}p_{\mathrm{e}}(\frac{x_o - x_s}{|x_o - x_s|})\delta(t)) \qquad (3.63)$$

The equation was constructed straightforwardly from equation 3.16, definition 3.2.7, and the temporal intensity algebra. The content should be quite obvious: H^{α} stands for medium absorption, ν for visibility, $\frac{S_{|\cdots|}}{4\pi|\cdots|^2}$ for propagation delay and attenuation, p_{e} for directional intensity of source, δ for the impulse, and finally, dtf for the detection transfer function.

As pointed out in the introduction to Section 3.3, plain intensity considerations are not general enough but radiances are. The equation above is nevertheless valid for the detection of the direct radiation emitted by a point source.

***

To refresh memory, let us summarize some radiance concepts. Radiance is the element of radiation distribution in different directions, and the incidence of radiance has to be corrected with the projection term $\cos\theta$ to obtain the correct energy flux.

[15] In a dynamic environment with a non-stationary source and observer one has to apply the Doppler-effects for below wavespeed object motion. This is straightforward, albeit not necessarily trivial because of the frequency shifts. The general effects of faster-than-wavespeed motion, however, are more complex.

Observation is special, however, and does not require that correction because there is no surface at the observer.[16] Observation is also direction-dependent, and the detection (per direction) is something observer-specific defined by the detection transfer function dtf.

Our construction of the radiance form begins by noting that radiation to the observation point may come in all directions. The total observed radiation is a sum of all observed radiation in all directions — or an integral when the set of directions is continuous, as in our case.

Now, let us assume that we know the time-dependent radiance of all surrounding geometry. This is denoted by $\ell(a,\Omega)$, where a is a point in the geometry and Ω is the direction of exitant radiance. Thus, collecting the radiance detection in different directions at the observation point x_o, one obtains

$$d_r(t) = \int_{4\pi} \mathrm{dtf}(\Omega,\, H^{\alpha} S_{\left|\nu_p(x_o,\Omega)-x_o\right|} \ell(\nu_p(x_o,\Omega),\, -\Omega))\, \mathrm{d}\Omega \tag{3.64}$$

where $\nu_p(\cdots)$ is the inverse projected surface point in the geometry, and ℓ is the radiance emission in TIA form in the projection point towards the observer. Note that the distance attenuation is *implicitly* included because the more distant radiance sources are "scanned faster" by the inverse projection.

The equation 3.64 can be transformed from the directional gathering form into the visible environment contribution form:

$$d_r(t) = \int_{\nu^{-1}(x_o)} \mathrm{dtf}(\frac{a-x_o}{|a-x_o|}, \\ H^{\alpha} \frac{S_{|a-x_o|}}{|a-x_o|^2} \ell(x_o, \frac{x_o-a}{|x_o-a|}) \left[\mathbf{n}(a)\cdot\frac{x_o-a}{|x_o-a|}\right])\mathrm{d}a \tag{3.65}$$

The integral surface has changed but notice that the new surface can still be projected on the unit sphere around x_o. For a small surface area da, the contribution to the result must be the same as when the same small surface area was enumerated in 3.64 through the inverse projection. This requires adjusting the measure

[16] Alternatively, we can think the observation point as a small sphere, where a small area in the surface of the sphere detects only incident radiation orthogonal to the surface.

because of the changed surface area ($|a-x_o|^2$ too big) and the non-orthogonality to the observation direction ($\frac{1}{\cos\theta}$ too much surface area). Compensating for these, we achieved the integral above.

The integral can be further transformed for the full geometry by adding the visibility function and noting that $\nu^{-1}(x_o) \subset \mathcal{G}$. The visibility function evaluates 1 if $a \in \nu^{-1}(x_o)$ and 0 otherwise. By this transformation we achieve:

Definition 3.3.5 *Detection of the Radiance in the Geometry*

$$d_r(t) = \int_{\mathcal{G}} \mathrm{dtf}(\frac{a-x_o}{|a-x_o|}, \\ H^{\alpha}\nu(x_o,a)\frac{S_{|a-x_o|}}{|a-x_o|^2}\ell(x_o,\frac{x_o-a}{|x_o-a|})\left[\mathbf{n}(a)\cdot\frac{x_o-a}{|x_o-a|}\right])\mathrm{d}a \quad (3.66)$$

The full geometry form has two advantages. First, the geometry does not need to be closed, as is the case with equations 3.64 and 3.65. Second, the surface area to integrate is invariant in respect to the observation point. These make operator analysis easier.

***

For notational convenience, from now on, we denote the reflected radiance caused by the direct source exposure with symbol ℓ_0, surface radiance via one reflection with ℓ_1, and radiance via k reflections with ℓ_k. Let us begin with the formulation of ℓ_0.

A point source conducts direct incident energy flow to the surface. For a small surface area, the energy flow is

$$\mathrm{I_i} = \frac{\mathrm{d}\Phi_\mathrm{i}}{\mathrm{d}a} = \nu(a,x_s)\frac{\Phi_s p_\mathrm{e}(\frac{a-x_s}{|a-x_s|})}{4\pi|a-x_s|^2}\left[\mathbf{n}(a)\cdot\frac{x_s-a}{|x_s-a|}\right] \quad (3.67)$$

where x_s is the location of the point source and p_e is the emittance pattern. The $\mathbf{n}(a)\cdot|\cdots|$ term accounts for the possible non-orthogonality to the surface. By

equation 3.19 we conclude that the incident radiance must be

$$L_{\mathrm{i}}(a,\Omega) = \nu(a,x_s)\frac{\Phi_s p_{\mathrm{e}}(\frac{a-x_s}{|a-x_s|})}{4\pi\,|a-x_s|^2}\delta(\Omega - \frac{x_s-a}{|x_s-a|}) \tag{3.68}$$

and the exitant radiance is by the total reflected radiance (eq. 3.21):

$$\begin{aligned} L_{\mathrm{e}}(a,\Omega_{\mathrm{e}}) &= \int_{2\pi} f_{\mathrm{r}}(\Omega_{\mathrm{i}},\Omega_{\mathrm{e}})L_{\mathrm{i}}(a,\Omega_{\mathrm{i}})\cos\theta_{\mathrm{i}}\mathrm{d}\Omega_{\mathrm{i}} \\ &= \nu(a,x_s)\frac{\Phi_s p_{\mathrm{e}}(\frac{a-x_s}{|a-x_s|})}{4\pi\,|a-x_s|^2}\int_{2\pi} f_{\mathrm{r}}(\Omega_{\mathrm{i}},\Omega_{\mathrm{e}})\delta(\Omega_{\mathrm{i}} - \frac{x_s-a}{|x_s-a|})\cos\theta_{\mathrm{i}}\mathrm{d}\Omega_{\mathrm{i}} \\ &= \nu(a,x_s)\frac{\Phi_s p_{\mathrm{e}}(\frac{a-x_s}{|a-x_s|})}{4\pi\,|a-x_s|^2}\left[\mathbf{n}(a)\cdot\frac{x_s-a}{|x_s-a|}\right] f_{\mathrm{r}}(\frac{x_s-a}{|x_s-a|},\Omega_{\mathrm{e}}) \end{aligned} \tag{3.69}$$

Now, assuming that the point source emits unit energy at time 0, and accounting for the propagation delay, we transform the exitant radiance into TIA form:

$$\begin{aligned} \ell_0(a,\Omega_{\mathrm{e}}) = \nu(a,x_s)\Phi_s p_{\mathrm{e}}(\frac{a-x_s}{|a-x_s|})\frac{S_{|a-x_s|}}{4\pi\,|a-x_s|^2}\cdot \\ \left[\mathbf{n}(a)\cdot\frac{x_s-a}{|x_s-a|}\right] f_{\mathrm{r}}(\frac{x_s-a}{|x_s-a|},\Omega_{\mathrm{e}})\delta(t) \end{aligned} \tag{3.70}$$

We call ℓ_0 the *primary* or 0-order reflected radiance which is analogous to radiance emitted by the surface. In a way we may consider the radiation emitted by the point source as radiance emitted by the surfaces exposed to the point source. Indeed, the propagation of the primary radiance is actually examined.[17]

***

Equation 3.66 tells us how to detect radiation in the environment — if we know the time-dependent radiance of the surfaces. We will next examine how to calculate reflected radiance of a small surface area conducted by the other surfaces. Essentially this means calculating the reflected radiance of the next order. Finally, we

[17] It would also have been possible to express the point source in terms of radiance. However, this would have added complexity into the analysis without any practical benefits. Then, ℓ_1 would have been exactly of the present form of ℓ_0.

sum all orders of radiance to obtain the time-dependent radiance via all numbers of reflections.

The radiance of the first order reflection (ℓ_1) can be calculated from the primary radiance (ℓ_0). By following the derivation pattern used in the derivation of the detection, we start with the incident radiation from all directions to surface point a_1:

$$\ell_1(a_1,\Omega_1) = \int_{2\pi} f_{\mathrm{r}}(\Omega,\Omega_1;\ \nu_p(a_1,\Omega)) \cdot S_{|\nu_p(a_1,\Omega)-a_1|}\ell_0(\nu_p(a_1,\Omega),-\Omega)\cos\theta\ \mathrm{d}\Omega \tag{3.71}$$

where Ω is the direction of the incident radiance, Ω_1 is the direction of the reflected radiance, a_1 is the reflection point, and θ is the angle between a_1 and Ω. The equation is easily reached by starting with the total reflected radiance (eq. 3.21), inverse projecting the point for incident radiation, and adding the appropriate propagation delay. Notice, however, that the distance attenuation is again implicit due to the inverse projection (see note at eq. 3.64). Following the derivation pattern, we transform the equation into the full geometry form:

Definition 3.3.6 *Reflected Radiance after one Reflection*

$$\begin{aligned}\ell_1(a_1,\Omega_1) &= \int_{\mathcal{G}} \nu(a,a_1) f_{\mathrm{r}}(\frac{a-a_1}{|a-a_1|},\Omega_1;\ a_1)\frac{S_{|a_1-a|}}{|a_1-a|^2}\ell_0(a,\frac{a_1-a}{|a_1-a|}) \cdot \\ &\qquad \left[\mathbf{n}(a_1)\cdot\frac{a-a_1}{|a-a_1|}\right]\left[\mathbf{n}(a)\cdot\frac{a_1-a}{|a_1-a|}\right]\mathrm{d}a \\ &= \int_{\mathcal{G}} \nu(a,a_1) f_{\mathrm{r}}(\frac{a-a_1}{|a-a_1|},\Omega_1;\ a_1)\ell_0(a,\frac{a_1-a}{|a_1-a|}) \cdot \\ &\qquad g(a,a_1)\mathrm{d}a \\ &= \int_{\mathcal{G}} \mathrm{R}(a,a_1,\Omega_1)\ell_0(a,\frac{a_1-a}{|a_1-a|})\mathrm{d}a\end{aligned} \tag{3.72}$$

The $\left[\mathbf{n}(a_1)\cdots\right]$ term replaces $\cos\theta$ of equation 3.71 and $\left[\mathbf{n}(a)\cdots\right]$ is analogous to the respective term in equation 3.65. R is the reflection kernel (eq. 3.61).

The first order radiances can be used to calculate second order radiances, and so forth. The kth order radiance is simply written as:

Definition 3.3.7 *Reflected Radiance after k Reflections*

$$\ell_k(a_k,\Omega_k) = \int_{\mathcal{G}} \mathrm{R}(a,a_k,\Omega_k)\ell_{k-1}(a,\frac{a_k-a}{|a_k-a|})\mathrm{d}a \tag{3.73}$$

Finally, when the observer detects the radiation originated from the source, it makes no difference whether the radiation is direct or reflected one or more times. Thus, we define the sum radiance:

Definition 3.3.8 *Total Propagated Radiance*

$$\ell(a,\Omega) = \sum_{k=0} \ell_k(a,\Omega) \tag{3.74}$$

The convergence is considered in Section 3.3.5.

3.3.4 Acoustic Rendering Equation

The time-independent rendering equation, first presented by Kajiya [13], has succeeded as a generalization for several modern global illumination models in computer ray graphics. Having a generalization simplifies greatly the development and analysis of new models. For example, if one can show the convergence to the rendering equation, the validity of the model in addition to any general property of the rendering equation follows immediately as a corollary. In computer ray graphics, the rendering equation is a proven valuable tool.

The success of the rendering equation in graphics drives us to present the time-dependent rendering equation as an acoustic analysis tool. The time-dependent form shall be later referred as the *acoustic rendering equation* (ARE). We begin by presenting the time-independent rendering equation in the radiance form: [7]

$$\begin{aligned}&\forall a' \in \mathcal{G}:\\ &\quad L(a',\Omega') = L_0(a',\Omega') + \\ &\qquad \int_{\mathcal{G}} f_{\mathrm{r}}(\frac{a-a'}{|a-a'|},\Omega';\ a')L(a,\frac{a'-a}{|a'-a|})\hat{g}(a,a')\mathrm{v}(a,a')\mathrm{d}a\end{aligned} \tag{3.75}$$

where

$L(a,\Omega)$ is the exitant radiance in a towards Ω

$L_0(a,\Omega)$ is the exitant primary in a towards Ω

By the remark in context of equation 3.70, the direct radiation of a point source can be transformed into primary radiance. For details on the rendering equation, consult the original work [13] and additionally *e.g.* [7].

The rendering equation is a balance equation. The solution L represents the state where each small surface area in the geometry receives and produces as much energy as they emit and absorb. The time-dependency is simply extended.[18] We do that by introducing the concept of time and delay into the equation, and by requiring that the equation holds at all times:

$$\begin{aligned}&\forall a' \in \mathcal{G} : \quad \forall t \in \mathbb{R} : \\ &\quad L(a',\Omega',t) = L_0(a',\Omega',t) + \\ &\qquad \int_{\mathcal{G}} f_{\mathrm{r}}(\frac{a-a'}{|a-a'|},\Omega';\, a') L(a,\frac{a'-a}{|a'-a|},t-\frac{|a'-a|}{\mathbf{c}})\hat{g}(a,a')\nu(a,a')\mathrm{d}a\end{aligned} \tag{3.76}$$

The form is simplified a bit by using TIA (sec. 3.3.1):

Definition 3.3.9 *Acoustic Rendering Equation (in TIA form)*

$$\begin{aligned}&\forall a' \in \mathcal{G} : \\ &\quad \ell(a',\Omega') = \ell_0(a',\Omega') + \\ &\qquad \int_{\mathcal{G}} f_{\mathrm{r}}(\frac{a-a'}{|a-a'|},\Omega';\, a') \ell(a,\frac{a'-a}{|a'-a|})g(a,a')\nu(a,a')\mathrm{d}a\end{aligned} \tag{3.77}$$

and with the reflection kernel:

$$\begin{aligned}&\forall a' \in \mathcal{G} : \quad \forall t \in \mathbb{R} : \\ &\quad \ell(a',\Omega') = \ell_0(a',\Omega') + \int_{\mathcal{G}} R(a,a',\Omega')\ell(a,\frac{a'-a}{|a'-a|})\mathrm{d}a\end{aligned} \tag{3.78}$$

The solution ℓ can be detected by the observer as in definition 3.3.5. The existence of the solution is considered in the next section.

[18] Acknowledgment notice: The idea of the time-dependent formulation was originally presented to the author by professor Lauri Savioja.

3.3.5 Equivalency of Reflection-iterative Construction and Acoustic Rendering Equation

It is obvious that if the energy propagation equations described in Sections 3.3.3 and 3.3.4 are valid, their results must be equivalent. The equivalence will be shown in this section. The discussion begins by noting that the acoustic rendering equation (eq. 3.78) is of form *Fredholm integration equation of second kind* (sec. A.3, eq. A.25), when using the reflection kernel.

We make the redenotations $K(x,\Omega,\tau) = \mathrm{R}(\tau,x,\Omega)$, $\phi = \ell$, and $f = \ell_0$ for the familiarity with the notation often used with the Fredholm integration equation. Now, the acoustic rendering equation is written as:

$$\phi(x,\Omega) = f(x,\Omega) + \int_{\mathcal{G}} K(x,\Omega,\tau)\phi(\tau,\frac{x-\tau}{|x-\tau|})\mathrm{d}\tau \tag{3.79}$$

That the ARE is of the Fredholm form is now easy to see. Now, let us denote the integration operator with B such that:

$$(B\phi)(x,\Omega) = \int_{\mathcal{G}} K(x,\Omega,\tau)\phi(\tau,\frac{x-\tau}{|x-\tau|})\mathrm{d}\tau \tag{3.80}$$

Equation 3.79 may now be written in the abstract form:

$$\phi = f + B\phi \tag{3.81}$$

We should note for further analysis that B is a linear operator in respect to addition, scalar multiplication and shifting:

$$\begin{aligned} &B(\alpha\phi_\alpha + \beta\phi_\beta) = \alpha B\phi_\alpha + \beta B\phi_\beta \\ &B(S_k\phi) = S_k B(\phi) \end{aligned} \tag{3.82}$$

The linearities can be easily verified by using equation 3.80. The abstract form of

the ARE (eq. 3.81) has the Neumann series solution (thm. A.3.3):

$$\begin{aligned}
\phi_0 &= B^0 f = f \\
\phi_1 &= Bf \\
\phi_2 &= B^2 f \\
&\cdots \\
\text{and } \phi &= \sum_k \phi_k = \sum_k B^k f
\end{aligned}$$

It is straightforward to verify that $\phi_k(x,\Omega) = \ell_k(x,\Omega)$, which shows that the reflection-iterative construction is actually a Neumann series solution to the acoustic rendering equation.[19]

Of solution existence and unambiguity. If the Neumann series solution converges, it is an unambiguous solution to the acoustic rendering equation. A natural norm for the convergence analysis is the *energy* norm: $\|\phi\|$ evaluates the total exitant energy in the system, and is defined as:

$$\|\phi\| = \int_0^\infty \int_{\mathcal{G}} \int_{4\pi} \phi(a,\Omega)(t)\,[\mathbf{n}(a)\cdot\Omega]\;\mathrm{d}t\,\mathrm{d}a\,\mathrm{d}\Omega \tag{3.83}$$

The energy norm is constructed by integrating the time-dependent irradiance over time and the surface geometry. The energy norm is then used to define the reflection operator norm: [15]

$$\|B\| = \sup_{\|\phi\|=1} \|B\phi\| \tag{3.84}$$

It should be easy to show that

$$\|B\| \leq \sup_{\substack{a\in\mathcal{G} \\ \omega_i\subset 2\pi}} \rho(\omega_i, 2\pi;\, a) \tag{3.85}$$

because the reflectivity supremum sets the upper limit for any reflected energy such that $\|B\phi\| \leq \|\phi\| \sup\ \rho(\cdots)$. However, weaker conditions for convergence might prove considerably more difficult to derive. We shall present one more condition.

[19] Similar considerations are present in *e.g.* [13].

If the geometry is irregular enough[20], all energy patterns will become diffuse after some number of reflections, no matter how coherent they initially were.[21] For diffuse fields, it is enough that for any positive area subset of the geometry, the reflectance is below one when the upper limit for all reflectance is exactly one. This guarantees that the energy decreases in the reflection when the field is diffuse enough, implying the convergence. More formally:

$$\begin{aligned} &\mathcal{G} \text{ irregular enough} \bigwedge \exists A \subset \mathcal{G}\,,\ \text{area}(A) > 0 :\ \rho(2\pi, 2\pi;\ a \in A) < 1 \\ &\qquad \Downarrow \\ &\exists k :\ \left\| B^k \right\| \leq \eta_0 < 1 \Rightarrow \left\| B^{nk} \right\| \leq \left\| B^k \right\|^n = {\eta_0}^n \Rightarrow \lim_{n\to\infty} \|B^n\| \to 0 \end{aligned} \tag{3.86}$$

3.3.6 Remark on Detection

In Sections 3.3.3 and 3.3.4 we considered the primary radiance as the source energy. The primary radiance is, however, reflected radiation of the point source (eq. 3.70). Using the sum of all orders of reflected primary radiance in the detection is otherwise valid and sufficient but it lacks the direct contribution of the point source (eq. 3.63). The detection can be completed by adding the direct detection of the energy source to the detection of the total propagated primary radiance:

Definition 3.3.10 *Complete Detection*

$$d(t) = d_0(t) + d_r(t) \tag{3.87}$$

where d_0 is as in equation 3.63 and d_r is as in definition 3.3.5.

3.4 Radiation at Various Frequencies

Behaviour of radiation in reflections is dependent on the frequency of the radiation, in general. Many surfaces reflect radiation very differently at different fre-

[20] *i.e.* not most regular. All natural geometries are *irregular enough* unless very specifically designed not to be so.

[21] This is a result of *billiard* theory. See *e.g.* [23] for further information.

quencies. In addition, the medium absorption is also generally dependent on the frequency.

Until now in this work, the assumption has been that both the propagation and reflection are independent of the frequency — or that all radiation in the examination has the same frequency. We shall now drop this requirement and say instead that everything depends on the frequency.

The brief examination begins with the definition of the frequency band of interest which consists of all frequencies between 0 and some f_{max}. This band is called here the *full spectrum*. The full spectrum is divided into sub-bands, and the propagation and reflection behaviour is assumed homogenous inside a sub-band.

In a linear system, such as our model, a sub-band filtered source signal produces identical response to a sub-band filtered response of the unfiltered source signal.[22] Because of this frequential separability, the sub-bands are orthogonal if they do not share frequency components. Furthermore, because of the orthogonality, we can examine the responses of the separate sub-bands, and sum the results together for the full spectrum response.

One should note that, unlike the unfiltered impulse responses, the sub-band filtered responses cannot have peaks because of the finite band requirement. To remedy this, we will use *representation* responses instead, where we assume the filtering at the realization of the representation. Then, the impulse energy propagation equations introduced in the previous sections are readily usable. The formal presentation follows.

3.4.1 Mathematical Discussion

From Section 3.3 we know how to calculate the total propagation of primary radiance (eq. 3.74 and eq. 3.78). The equations hold valid also when computing the propagated primary radiance for sub-band radiation, as noted in the previous

[22] This can be straightforwardly verified by *e.g.* Fourier transform analysis of a linear system.

section. Now, noting that the actual energy cannot be peak-formed in time – not in the observed energy response but also not at the source — we speak rather of *representations* of the energy responses instead of actual energy responses.

How would the representation scheme work? Let us assume that we have an operator which maps the source-emitted energy pattern to the observed response. Let us denote the operator by $\widehat{\overline{B}}$ and remark that it is defined as

$$\widehat{\overline{B}} = \sum_{n=0}^{\infty} B^n \tag{3.88}$$

where B is the reflection operator defined in Section 3.3.5. As the sum is a linear operator, operator $\widehat{\overline{B}}$ inherits the linearities of operator B (eq. 3.82).

One reason for tracking the impulse response, where the impulse was emitted at time 0, was the easy inclusion of the linear medium absorption. The medium absorption, however, is dependent on the propagation distance. The propagation distance was assumed to be a function of time, which is obviously misdirected if we allow the time-shift linearity. We will address this soon.

Ignoring the small medium absorption hardship, we note that because of the linearity, the detected energy response for a signal is equal to the energy response convoluted by the energy profile of the sub-band filtered source signal:

$$\widehat{\overline{B}}\,(\ell(t) * \sigma(t)) = \widehat{\overline{B}}(\ell(t)) * \sigma(t) \tag{3.89}$$

where the convolution is along time. Thus, if sub-band filtering the source-emitted energy impulse produces a smooth energy profile (represented by $\sigma(t)$), the representation of the sub-band filtered response may be realized by convolution with the smooth energy profile.

As promised above, we now include the medium absorption. In general, the medium absorption in a linear medium is a function of the propagation distance. The convolution in equation 3.89 broadens the response in time, but not in propagation distance. The representation (ℓ) originates from the impulse energy, and therefore, it is valid to apply medium absorption as:

$$H^{\alpha}\,\widehat{\overline{B}}(\ell(t)) \tag{3.90}$$

If we now convolute the energy emitted by the source, it is exactly the response above which gets also convoluted. Thus, we have

$$\left[H^{\alpha}\,\widehat{\widehat{B}}\left(\ell(t)\right)\right] * \sigma(t) \tag{3.91}$$

for energy profiles of non-impulse source signal responses.

In Section 4.2, the relationship between the representation and the audible impulse response is briefly discussed.

3.5 Considerations on Extensions

In this section, we shall briefly discuss some known limitations of the constructed energy propagation theory and offer some possible treatments. The limitations are caused by simplificative assumptions. In Section 3.3, we assumed that rays propagate straightforwardly, and in Section 3.2.5, we assumed that the reflections can be represented by BRDFs. In general, these assumptions do not hold. In subsection 3.5.1, we discuss extending the theory for diffraction — non-straight ray paths — and in subsection 3.5.2, we discuss extending the theory for sub-surface propagation.

3.5.1 Extending for Edge Diffraction

Whenever a wave front confronts a surface that bends inwards or outwards, the wave front tends to bend with the surface — see figure 3.9 for illustration. This bending effect is called edge diffraction. The energy propagation equations ignore edge diffraction. Edge diffraction is stronger at greater wave lengths, and therefore, the proper modelling of the diffraction is important especially with low frequency waves.

One possibility of including edge diffraction is to modify the reflection kernel in the propagation equations. Because the diffraction bends ray paths, radiance arriving in a specific direction in the incident point may actually be originated from

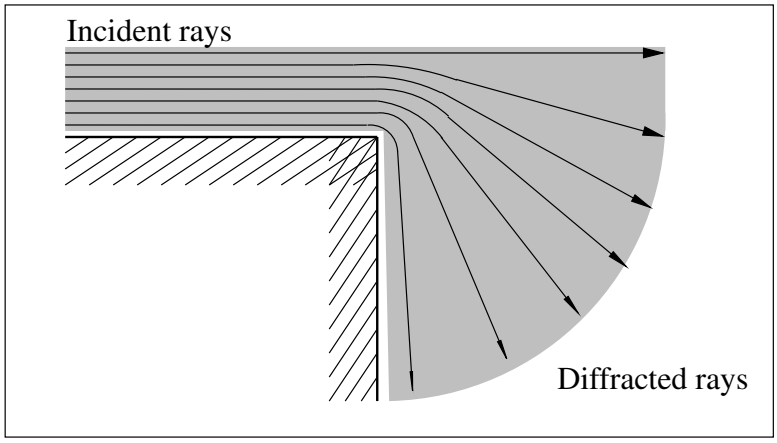

Figure 3.9: Diffractional bending of rays

a wider surface area than only an inverse projected point, as assumed by the current formulation. This could be remedied by redefining the concept of visibility. It is a boolean relation but extendable to a distribution, similar to unidirectional intensity that is extended to radiance.

The computation of visibility distributions requires solving the edge diffraction for ray path bending and broadening. Some edge diffraction models exist, see *e.g.* [28] for one. However, general edge diffraction models for ray paths are less known (at least to the author) but they are derivable from the wave equation, in principle.

3.5.2 Extending for Sub-surface Scattering

In some materials, the incident radiation can propagate beneath the surface. There, albeit a radiation beam is incident only in a specific part of the surface, the adjacent surface areas might emit back radiation. Specifically, the incident and exitant points for the reflected radiation need not be the same. In addition, there might be delay between the incidence and exitance.

Such reflection behaviour is called sub-surface scattering due to the famous model

introduced by Marcel Minnaert in lunar photometry [18]. In acoustics, sub-surface propagation exists in soft[23] materials, such as mineral wool.

It is obvious that a BRDF cannot model such scattering, as the BRDF is a function of only one point in the surface. However, the BRDF can be extended to a function of both incident and reflection points in the surface, see for example [20] for a definition. The extension is called *Bidirectional Sub-surface Scattering and Reflection Distribution Function* (BSSRDF). BRDFs are actually a subset of BSSRDFs.

As BRDFs are special cases of BSSRDFs, one might wonder why the reflection equations presented here are rather based on BRDFs and not BSSRDFs. The decision was made in favor of simplicity. The equations based on BRDFs are sufficient for many purposes and it should be straightforward enough to extend the equations for separate incident and exitant reflection points.

In essence, with BRDFs one integrates once over the surface geometry using the reflection kernel. With BSSRDFs one must integrate twice — once for the incidence and once for the exitance. A BRDF implemented by using a BSSRDF would consist of the Dirac delta functional and the embedded BRDF, much like the BRDF that embeds specular reflection.

Finally, a BSSRDF could imply delayed reflection due to sub-surface propagation. This delay does not, however, include the regular linear medium absorption. This prevents postponing the medium absorption into detection, and thus, the medium propagation absorption must be incorporated into energy propagation equations — say, into the geometry term, for example.

[23] *i.e.* non-rigid

Chapter 4

Acoustic Energy Propagation Theory

In this chapter we adapt the general energy propagation theory to the acoustic energy propagation theory (AEPT). The adaptation begins by specifying the general radiation propagation concepts as acoustic radiation propagation concepts in Section 4.1.

The general energy propagation theory evaluates energy responses per source and observer. In acoustics, however, we prefer the impulse responses — the pressure disturbance responses, that is. The response transformation process is called *auralization*, and is briefly discussed in Section 4.2.

The arguably three most common existing ray methods are shown to be specializations of the acoustic energy propagation theory in Section 4.3, thus providing the initial validation to the theory. Finally, a discussion on using the theory is given in Section 4.4.

4.1 Adaptation of the General Energy Propagation Theory

We adapt now the general energy propagation theory to acoustics. We do this by straightforward specification of the general concepts.

As stated in Chapter 2, we assume linear acoustics. The environment for the sound propagation may be described using the concepts introduced in Section 3.1. Sound field carries energy and propagating sound energy may be compared to propagating electromagnetic radiation. Because of the analogy, we may refer the propagating sound energy as *sound radiation*. The sound radiation may be separated into different frequency sub-bands. Section 2.1.2 contains the reasoning.

We assume homogenous behaviour of all sound radiation within a sub-band. This is required by Section 3.4. However, we do not fix the sub-bands, and thus, the sub-bands may be selected based on their approximation properties. We shall not harass ourselves with the question of the exact profiles of the sub-band filtered impulse energies (see sec. 3.4.1). For many problems, the sub-band energy profiles can be assumed as narrow peaks without remarkable loss of accuracy.

The sound radiation within a homogenous sub-band is an *energy flow*, which obeys all definitions in Section 3.2. Because of this, the sound radiation is subject to the energy propagation equations defined in Section 3.3. As pointed out in Section 3.5, the propagation models work only approximately. The lacks, sub-surface scattering and edge diffraction, are therefore inherited by the AEPT.

Finally, sound as radiation is not how we actually observe sound. Instead, we observe sound as pressure variations and arguably so do all current detectors, ultimately. Detectors may indicate energy or other derivative quantities, but the quantities always originate from the pressure variations. So, the question remains of how one can revert back from the energy quantities to the pressure quantities. We address this in the next section.

4.2 Auralization of Energy Response

Auralization stands for making something audible [14], and auralization of an energy response stands for making the energy response audible, respectively. In this section, we consider the auralization of sub-band energy impulse responses which are calculated by using the energy propagation equations described in Section 3.3. Auralization defines the detection transfer function dtf, which is used to transfer the surrounding time-dependent radiation into detection (sec. 3.2.6 and def. 3.3.5).

One used practice for auralization of energy responses is to consider the pointwise square root of an energy response as the envelope of the impulse response [12]. Now, when the envelope is *filled* with sub-band filtered white noise, one gets a signal with similar energy profile to the calculated response. However, in auralization of this kind one necessarily loses some phase information.

We begin the derivation by defining the intensity sensitivity patterns for both ears, say $\gamma_{\text{left}}(\Omega)$ and $\gamma_{\text{right}}(\Omega)$. The sensitivity patterns map the incident direction to the relative sensitivity factor. By multiplying the relative sensitivity factor by the incident intensity, one obtains the detection intensity.

In addition to the directional sensitivity, directional delay — temporal shift of the noise signal — should also be defined. Sound coming from the left side of a human observer comes a bit earlier to the left ear than to the right ear. Say that the delay patterns are $D_{\text{left}}(\Omega)$ and $D_{\text{right}}(\Omega)$.

The goal of the detection is to produce impulse responses from energy responses. For that, some special considerations are required. Let us say that $s(t)$ contains sub-band filtered white noise, which has the unit average intensity. Then, if we have energy response $\mathrm{I}(t)$, we can approximate the impulse response as $\hat{\imath}(t) \approx s(t)\sqrt{\mathrm{I}(t)}$. Adding delay D, smoothing the energy response by the sub-band impulse energy profile $\sigma(t)$ (sec. 3.4.1), and adding direction dependencies,

we obtain the detection transfer functions for the left and right ears:

$$\begin{aligned} \mathrm{dtf}_{\mathrm{left}}(\Omega, \mathrm{I}(t)) &= \sqrt{\gamma_{\mathrm{left}}(\Omega) \cdot (\mathrm{I} * \sigma)(t)} \cdot s(t + D_{\mathrm{left}}(\Omega)) \\ \mathrm{dtf}_{\mathrm{right}}(\Omega, \mathrm{I}(t)) &= \sqrt{\gamma_{\mathrm{right}}(\Omega) \cdot (\mathrm{I} * \sigma)(t)} \cdot s(t + D_{\mathrm{right}}(\Omega)) \end{aligned} \tag{4.1}$$

The intensity convolution with the sub-band energy profile σ eliminates possible singularities in the detection originated from the delta-peak intensity emitted by the source. This enables calculating the point-wise square root of the detection intensity. There is a problem, however, because the detection is not linear in respect to the intensity, as required in Section 3.2.3. Therefore, the direction distribution of intensity as defined in Section 3.2.6 requires additional justification.

First, the pressure disturbances arriving in different directions are assumed incoherent. Therefore, the energy of the incident disturbances is the sum of the energies of the individual disturbances. Second, because the incident sound is assumed random and thus, temporally incoherent, the total energy of multiple sounds incident in the same direction but at different times is the sum of the energies of the individual sound signals, regardless of that the sound signals may overlap.

The defined detection transfer functions $\mathrm{dtf}_{\mathrm{left}}$ and $\mathrm{dtf}_{\mathrm{right}}$ are head-related transfer functions (HRTF) [14]. By using them, the approximated impulse response per ear is obtained by the detection equations of the GEPT (eq. 3.63 and def. 3.3.5).

4.3 Specializations of the Theory

In this section we give a validative discussion on the acoustic energy propagation theory by showing that three common ray acoustic modelling methods are subclasses, or specializations, of the theory. We start the discussion with the image source method (sec. 2.4.1) and the radiosity method (sec. 2.4.2). That the methods are subclasses of the theory, is straightforward to see. Finally, we note that the ray tracing method (sec. 2.4.3) is also a specialization.

4.3.1 Image Source Method

The vanilla image source method considers only specular reflections. By substituting the BRDF for specular reflection (def. 3.2.11) into reflected radiance equation (eq. 3.72), we get:

$$\begin{aligned}\ell_1(a_1,\Omega_1) &= \int_{\mathcal{G}} \mathrm{R}(a_0,a_1,\Omega_1)\ell_0(a_0,\frac{a_1-a_0}{|a_1-a_0|})\mathrm{d}a_0 \\ &= \int_{\mathcal{G}} \nu(a_0,a_1)g(a_0,a_1)f_{\mathrm{r,s}}(\frac{a_0-a_1}{|a_0-a_1|},\Omega_1;a_1)\ell_0(a_0,\frac{a_1-a_0}{|a_1-a_0|})\mathrm{d}a_0\end{aligned} \tag{4.2}$$

The image source method works on polygon geometries. Now, remember that the BRDF for specular reflection is zero whenever $\frac{a_0-a_1}{|a_0-a_1|} \neq M\Omega_1$ because of the mirror reflection requirement. Remember also that for any polygon, there may exist at most one point where the inequality does not hold (sec. 2.4.1).

Obviously, for any polygon which does not have such potential point for specular reflection, the image source method and the equation above give the same result: zero contribution. Further, if the potential point of reflection is not visible to the observer or to the source, the visibility function ν evaluates zero, and again, the results are equivalent.

Showing the equivalency in the reflection — when the contribution is not zero — is a bit more difficult, and requires formal examination. In the examination, we perform our calculations with intensities instead of radiances. Later we note, however, that the results apply also in radiance calculus.

Assume a polygon with a visible potential reflection point. Then assume that the point source emits radiation omni-directionally, and that the reflected energy is detected by a small detection plane, see Figure 4.1.

By tracing back to the energy source from the detection plane A_2, one realizes that the energy must be reflected via reflection plane A_1. Energy conservation law sets relation to the beam intensity:

$$\mathrm{I}_2 = \mathrm{I}_1\frac{A_1}{A_2} \tag{4.3}$$

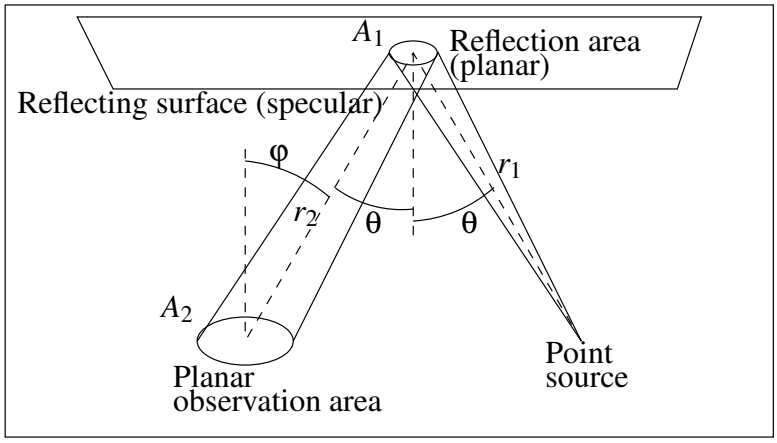

Figure 4.1: Reflecting beam

where I_1 is the average reflected intensity in the reflection plane, and I_2 is the average intensity incident to the detection plane.

Note especially, that there is no required knowledge of the source at all in calculation of I_2. It is enough to know the energy flow from A_1 to A_2. Now we verify that this equals to the direct irradiation with the same propagation distance.

Assume that the plane areas A_1 and A_2 are small compared to distances r_1 and r_2. Then the distance between any two arbitrary points in A_1 and A_2 are approximately the same, as well as the distance between an arbitrary point in A_1 and the source. The incident intensity in A_1 (or irradiance) is now easy to calculate:

$$I_1 = \frac{\Phi_S}{4\pi {r_1}^2} \cos\theta \tag{4.4}$$

Projecting the areas A_1 and A_2 on the cross-section of the beam (cross-sections intersecting the centers of the planes), and by magnification equation, we obtain the relation

$$\frac{A_1 \cos\theta}{{r_1}^2} = \frac{A_2 \cos\varphi}{(r_1 + r_2)^2} \tag{4.5}$$

where θ is the angle between the beam and the reflection surface normal, and φ is the angle between the normal of A_2 and the beam. Now combine 4.3 and 4.5 and

further 4.4 to get

$$\mathrm{I}_2 = \mathrm{I}_1 \frac{{r_1}^2 \cos\varphi}{(r_1 + r_2)^2 \cos\theta} = \Phi_S \frac{\cos\varphi}{4\pi (r_1 + r_2)^2} \tag{4.6}$$

which is the equation for the direct irradiance. Note that the areas of the planes A_1 and A_2 are eliminated from the equation, as well as the orientation of the reflection plane.

The point source conducts radiation according to equation 4.4 to plane A_1. This can be transferred to incident radiance by using equation 3.19:

$$\int_{2\pi} L_\mathrm{i}(\Omega, a) \cos\theta \mathrm{d}\Omega = \Phi_S \frac{\cos\theta}{4\pi {r_1}^2} \tag{4.7}$$

which is satisfied with

$$L_\mathrm{i}(\Omega, a) = \frac{\delta(\Omega - \frac{x_s - a}{|x_s - a|})}{4\pi |x_s - a|^2} \tag{4.8}$$

where x_s is the position of the source. By the mirror transformation requirement

$$L_\mathrm{e}(\Omega, a) = L_\mathrm{i}(M\Omega, a) \tag{4.9}$$

where Ω points at A_2 and $M\Omega$ points at the source. Thus, by integrating the exitant radiance from A_1 over directions, one obtains the exitant intensity directed at A_2. The energy flow from A_1 to A_2 equals to the energy flow from the source to A_1 by the previous examination, justifying the radiance calculus.

The ray analysis used with a single image source gives thus equivalent results with once reflected radiation in the GEPT. This can be extended to reflected radiation of any order by similar analysis. The propagation delay was neglected in the examination above but that analysis is trivial.

In polyhedral environments there may be multiple reflection paths but countable, nevertheless. For one path, the image source method and the reflection-iterative construction give equivalent results. For all paths per reflection order, the image source method simply sums the results of ray paths for total result. The same applies to the reflection-iterative construction. This is easy to see, as is also that the

reflection paths are equal. In integration over all polygons, the visible reflection points (because of the Dirac delta functionals) in integration are simply summed together. Thus, the image source method is a specialization of the reflection-iterative construction. □

4.3.2 Radiosity Method

The radiosity method is easily shown to be a specialization of the AEPT. The inheritance is so strong that the AEPT can be even considered as an extension to the radiosity method.

In the vanilla radiosity method (sec. 2.4.2) the environment consists of diffusely reflective patches. Patches are essentially discretization of the geometry. One calculates then the communication between each patch-patch pair such that if one patch receives an energy flow, the communication tells how that energy flow will be conducted to the other patches.

If the patches are numbered from 1 to n and the time-dependent exitant energy flows are gathered into vector P, the patch-patch communications can be gathered into a matrix A, such that

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & & a_{2,n} \\ \vdots & & \ddots & \\ a_{n,1} & & & a_{n,n} \end{pmatrix}, \qquad P = \begin{pmatrix} \Phi_1(t) \\ \Phi_2(t) \\ \vdots \\ \Phi_n(t) \end{pmatrix} \tag{4.10}$$

The matrix element $a_{i,j}$ describes the passing of energy flow from patch j to patch i. The element is defined as: [1]

$$a_{i,j} \approx \frac{\nu(i,j)g(i,j)\beta_j \text{area}(i)}{\pi} \tag{4.11}$$

where ν is the visibility, g is the geometry term, β_j is the diffuse reflectivity factor, and area$(\cdots)$ is the surface area represented by the patch.

[1] See Section V. *Form Factors* in [21]. The derivation follows from eq. (23) by assuming constant integrand.

If then there is a radiation source conducting a direct energy flow to the patches, this can be transferred to an energy flow emitted by the patches — compare with the primary radiance in the AEPT. Further, if the energy flow emitted by the patches is denoted by vector F, the reflected radiation via k reflections can be written as

$$P_k = A^k F \tag{4.12}$$

and via all reflection orders:

$$P = \sum_{k=0} A^k F \tag{4.13}$$

Alternatively, one may write the balance equation (see thm. A.3.1)

$$P = AP + F \tag{4.14}$$

and solve P.

There is a great similarity in the calculations to the AEPT, which already suggests a strong relationship. The calculus above is immediately derivable from the AEPT by assuming diffuse-only reflections and by approximating the surface geometry by patches. See [21, 7] for further details on radiosity.

4.3.3 Ray Tracing Method

The ray tracing method (sec. 2.4.3) is essentially a numerical solving method to the reflection-iterative construction or the acoustic rendering equation, depending on the flavor of the ray tracing method. For a short introduction, consider the equation for energy propagation via one reflection (def. 3.3.6):

$$\ell_1(a_1, \Omega_1) = \int_{\mathcal{G}} \nu(a, a_1) g(a, a_1) f_{\mathrm{r}}(\frac{a - a_1}{|a - a_1|}, \Omega_1;\, a_1) \ell_0(a, \frac{a_1 - a}{|a_1 - a|}) \mathrm{d}a \tag{4.15}$$

The incident radiance ℓ_0 can be thought as the incident particle density, and ℓ_1 as reflected particle density. The f_{r} defines the reflection probability distribution,

such that the biconical reflectance factor $\rho(\omega_i, \omega_e)$ (def. 3.2.9) defines the probability of a particle incident in ω_i to reflect towards ω_e. Reflectance of 0.6, for example, means that 60% of incident particles are reflected.

In brief, occupy the incident radiance with particles such that the energy of the radiance is approximated by the energy of the particles. At the reflection, for an individual particle, randomize a new direction or absorb the particle using the reflection distribution probability defined by the biconical reflectance factor. The reflected particles approximate the reflected energy.

A formal examination shall be omitted here, but can be found in *e.g.* [9].

4.4 Considerations on Using the Theory

The constructed theory is not necessarily the most suitable for direct numerical approximation, as after a straightforward discretation, the numerical calculations easily reach computational complexity beyond any realistic level. This is a price for generality, but the direct numerical approximation was not the purpose of the construction, anyway.

The possible value of the theory rises from the subclass methods. In Sections 4.3.1-4.3.3 it was shown that some commonly used ray acoustic methods are specializations of the constructed theory. The theory is not, however, restricted to be a base for only these three methods. Other linear ray acoustic methods can also be shown to be specializations, and more importantly, one can derive new methods as approximations or numerical solving methods to the theory.

For example, the radiosity method can be considered as a very straightforward approximation, assuming diffuse reflections and patch discretation of the environment. It is easy to see how the radiosity method can be extended for non-diffuse reflections using the theory. Immediately with almost no changes to the radios-

ity method, one can loosen the diffuse assumption to memoryless reflections[2] assumption. Further, one can extend the radiosity method even for full BRDF reflections. Then, however, the patch-patch communications as in vanilla method become patch-patch-patch communications. This increases greatly the complexity of the modelling but if one patch does not communicate directly with many other patches in average, the full BRDF radiosity might still be usable. An example could be the propagation modelling of sound radiation in a tunnel network.

The reflection kernel analysis might also yield new acoustic property calculation methods for rooms, such as estimates for reverberation times. Some currently used estimates assume diffuse or specular reflections but the reflection kernel is free of such assumptions. A possible method might, for example, use the average reflection response (the temporal distribution of the reflected energy) and the average absorption, and by these, approximate the average reverberation time.

One thing worth noticing is that the theory extends the time-independent ray analysis. Especially computer ray graphics is heavily based on that. One might even say that computer ray graphics is a subclass of ray acoustics. Nevertheless, the constructed theory is readily usable as a time-independent ray propagation theory by assuming infinite ray velocity. Then the theory is also simplified because the temporal intensity algebra can be replaced by a simpler algebra of the constant energy flows.

[2]Memoryless reflection: the shape of the exitant radiation distribution does not depend on the shape of the incident radiation distribution.

Chapter 5

Conclusion

The conclusion of this work is divided in two sections. In Section 5.1, the success of this work is critically reviewed, and in Section 5.2, the open issues and some suggestions for extensions are considered.

5.1 Discussion

In this work, an acoustic energy propagation theory (AEPT, Chapter 4) is constructed. The construction is based on a general energy propagation theory (GEPT, Chapter 3) which is also constructed in this work. The general energy propagation theory is based on well-accepted classical physics.

The general theory is based on time-dependent surface radiance analysis. Similar but time-independent theories exist in the field of computer graphics, and the time-independent theories are used to verify the most important constructions. For example, the time-dependent rendering equation (sec. 3.3.4) is straightforwardly extended from the time-independent rendering equation [13, 7]. The reflection-iterative construction (sec. 3.3.3) is then shown to be a Neumann series solution to the time-dependent rendering equation (sec. 3.3.5) — which should verify the correctness of the construction.

The acoustic energy propagation theory is constructed to form a base for ray acoustic modelling methods. The applicability of the theory as such a base is shown for the image source method, the acoustic radiosity method, and the acoustic ray tracing method. The theory unifies the analysis of the methods, as any general property shown to the AEPT may be immediately passed on the methods as a corollary.

The AEPT enables the derivation of new ray acoustic methods as approximations or numerical solving methods to the theory, similar to how the existing methods were shown to subclass the AEPT. The image source method and the radiosity method are approximations of the AEPT, and the ray tracing method is a numerical solving method. Having a reference theory helps understand the domain and properties of a newly derived method, and if the method can be shown to converge to the AEPT, no additional fundamental justification is required for the method, as the justification is inherited from the AEPT.

The AEPT has also uses beyond modelling. The reflection kernel analysis may be used to derive methods to estimate acoustic properties of environments. A proposition is made for a reverberation time estimator for a room or hall in Section 4.4.

The general energy propagation theory can also be used as a base for time-independent energy propagation analysis commonly used in computer ray graphics, when assuming infinite ray velocity. This is important because it allows any results and properties belonging to the GEPT to be applied in computer ray graphics — assuming that the infinite wavespeed does not introduce unsolvable singularities. Similarly, the results and properties known in computer ray graphics can be often extended straightforwardly into a time-dependent form, and thus to the GEPT. This possibility for the bidirectional property exchange interconnects the acoustic and graphic ray modelling methods.

The theory is incomplete, however. It lacks edge diffraction and sub-surface scattered reflection modelling. This limits somewhat the use of the theory. Edge diffraction was omitted in the very fundamental derivations. Generally known, this weakens the accuracy of the modelling with longer wave lengths. A possible

incorporation of edge diffraction is considered in Section 3.5.1. The lack of sub-surface scattering limits the use of the theory to rigid surfaces. The extension is straightforward, and demarcated in Section 3.5.2.

The strongest and the most fundamental base for the counter-arguments is perhaps the validity of the use of sound radiation propagation concept, instead of the more general propagation of pressure disturbations. The use of energies is because of two reasons: 1) realistic reflection modelling is much more difficult with pressure disturbations, and 2) to simplify calculus.

Admittedly, the use of energy radiation has some significant drawbacks. It inherently aggravates the problem of auralization (sec. 4.2), and requires special considerations for the analysis of radiation at multiple frequencies. The direct wave field analysis would yield full-spectrum results, for example. However, I consider that the drawbacks are outweighed by the advantages in simplicity and computational efficiency — otherwise we would not have the existing acoustic ray methods in the first place.

The most auspicious property of the AEPT is perhaps not in the original intention of the theory. The theory, if successful, may prove to hint the direction for the construction of new, more general theories. There are two principal paths: evolutionary and revolutionary. The evolutionary progression considers the lacks of the current theory and ways to overcome them by refining the theory. The revolutionary path seeks to construct a whole new theory. However, it is then important to identify the simplificative assumptions in the current theory because the revolutionary theory must give the exact same results with such simplificative assumptions — otherwise the theory contradicts with the presented AEPT, or is not a true extension. In this light, the AEPT itself is a revolutionary step over the three existing methods shown as specializations.

5.2 Further Work

As pointed out in the previous section, the acoustic energy propagation theory constructed in this work is incomplete. In addition to the obvious shortcomings, some possible related further work will be discussed in this section.

The theory lacks edge diffraction, and sub-surface scattering effects. These at least offer obvious further work. The extensions were discussed in Section 3.5 and subsections.

In some realistic setups there are multiple mediums wherein the sound radiation may propagate. Such extensions should be straightforward to implement in the theory where required.

Also, as speculated earlier, the theory could provide inspiration for new efficient ray methods. With Kajiya's rendering equation in graphics [13] such has happened. Many new global illumination models are indeed compared to the Kajiya's rendering equation for validation.

Some effort could be laid in the generalization of the multi-band radiation analysis presented in Section 3.4. When letting the number of equally sized sub-bands reach infinity, each sub-band becomes of infinitesimal size. The sum of them, the integral, has such resemblance to the Fourier transform that one must wonder whether such analysis could yield full-spectrum calculus without the need to address each sub-band separately, as is currently done.

The hard further work would be to study the possibility of porting the AEPT to an acoustic pressure disturbation propagation theory. The advantage of success would be immediate, as one would gain actual impulse responses as a result instead of the energy profiles of impulse responses. It would eliminate, or ease greatly at least, the problem of auralization.

I have identified two possible paths. The first one is to derive the disturbance propagation straight from the beginning — the wave equation and Huygens' Principle. However, there one confronts the issue of reflection definitions which are

somewhat problematic with pressures disturbations.

In the second path one would still calculate the wave front propagation with energies, and obtain energy responses. I conjecture that the energy responses can be transformed into impulse responses without loss of information. The conjecture is based on the intuition, when considering a sound field formed of particles, or wavelets, and each of them carry energy to the energy response. The particles at the source are each of the same form inside a homogenously behaving sub-band. The particles do not change form in reflection — again because of the homogeneity of the sub-band.

Whether the acoustic pressure disturbance propagation theory is achievable by either path, or possibly a third path yet unknown, I dare not speculate here. Before such further study, the success of the proposed AEPT should be weighed first.

Appendix A

Some Essential Mathematics

In this appendix, some mathematical concepts that are essential to this work are briefly presented.

A.1 Euclidean Space

Definition A.1.1 *Dot-product*

$$x \cdot y = x_1 y_1 + x_2 y_2 + x_3 y_3 \tag{A.1}$$

Definition A.1.2 *Vector Length (Length Norm)*

$$|x| = \|x\|_2 = \sqrt{x \cdot x} \tag{A.2}$$

Definition A.1.3 *Unit Vector*

Vectors of length 1 are called *unit* vectors.

Definition A.1.4 *Cross-product*

$$x \times y = \det \begin{pmatrix} \vec{\mathbf{i}} & \vec{\mathbf{j}} & \vec{\mathbf{k}} \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{pmatrix} = \begin{pmatrix} x_2 y_3 - x_3 y_2 \\ x_3 y_1 - x_1 y_3 \\ x_1 y_2 - x_2 y_1 \end{pmatrix} \tag{A.3}$$

Definition A.1.5 *Cosine and Sine of Angle Between Two Vectors*

$$\begin{aligned} \cos \angle(x,y) &= \frac{x \cdot y}{|x|\,|y|} \\ \sin \angle(x,y) &= \frac{|x \times y|}{|x|\,|y|} \end{aligned} \tag{A.4}$$

Definition A.1.6 *Mirror Transformation M*

The mirror transformation operator evaluates the specular (mirror) reflection direction Ω_e for the incident direction Ω_i.

In the angular form (def. A.2.2) the azimuth angle of the reflection is the opposite after the reflection, resulting to

$$M(\theta,\phi) = (\theta, \phi \pm \pi) \tag{A.5}$$

where θ is the elevation angle of the incident and exitant directions, and ϕ is the azimuth angle of the incident direction, and $\phi \pm \pi$ is the azimuth angle of the exitant direction [20].

In the solid angle form the mirror transformation can be represented as (see fig. A.1)

$$\begin{aligned} &\Omega_e + \Omega_i = 2(\Omega_i \cdot n)n, \quad \Omega_e = M\Omega_i \\ &\Rightarrow \Omega_e = M\Omega_i = 2(\Omega_i \cdot n)n - \Omega_i = 2n(n^T\Omega_i) - \Omega_i = (2nn^T - I)\Omega_i \end{aligned} \tag{A.6}$$

where $n = \mathbf{n}(a)$ is the surface normal at the reflection point. The operator M can be expressed as a matrix:

$$M = \begin{pmatrix} 2n_0n_0 - 1 & 2n_1n_0 & 2n_2n_0 \\ 2n_0n_1 & 2n_1n_1 - 1 & 2n_2n_1 \\ 2n_0n_2 & 2n_1n_2 & 2n_2n_2 - 1 \end{pmatrix} \tag{A.7}$$

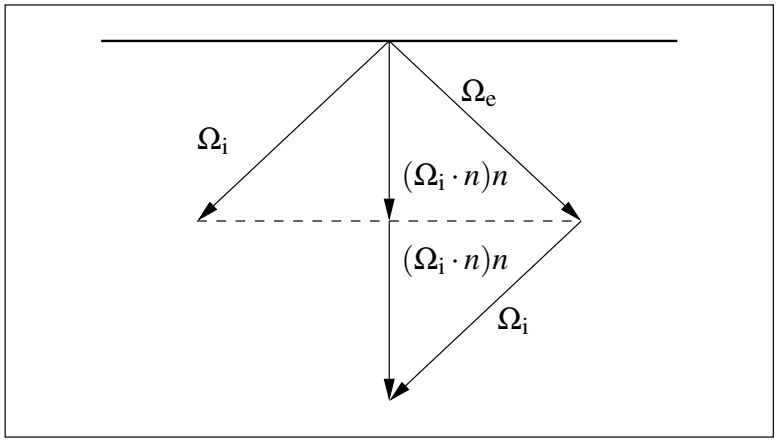

Figure A.1: Mirror reflection

A.2 Integration

Remark A.2.1 *Reparametrization of Integral*

Let us assume that we are calculating

$$\int \cdots \int_A f(x)\mathrm{d}x \tag{A.8}$$

where the integral is an n-dimensional integral over A. We can replace the integration parameter x with y as:

$$\int \cdots \int_A f(x_1,\ldots,x_n)\mathrm{d}x_1 \cdots \mathrm{d}x_n = \int \cdots \int_A f(x_1(y),\ldots,x_n(y)) \left| \det \begin{pmatrix} \frac{\partial x1}{\partial y1} & \frac{\partial x1}{\partial y2} & \cdots & \frac{\partial x1}{\partial y_n} \\ \frac{\partial x2}{\partial y1} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \frac{\partial x_n}{\partial y1} & \frac{\partial x_n}{\partial y2} & \cdots & \frac{\partial x_n}{\partial y_n} \end{pmatrix} \right| \mathrm{d}y_1 \cdots \mathrm{d}y_n \tag{A.9}$$

Typically, such transformation is used to simplify the integral. For example, sphere symmetrical functions are typically easier to integrate after reparametrization to the spherical coordinates.

Remark A.2.2 *Solid Angles, and Integration over Solid Angles*

The following notation for the solid angles is used: (see also def. 3.1.2)

ω	solid angle, set of directions
Ω	an infinitesimal solid angle, or a single direction
θ	the elevation component of Ω, ($\cos\theta = \mathbf{n}(a) \cdot \Omega$)
ϕ	the azimuth component of Ω

In addition, 4π denotes a full-covered sphere, and 2π denotes a hemisphere.[1] Unless otherwise noted, a hemisphere in the context of the surface reflection denotes always the outerior side hemisphere of the surface, *i.e.*, the side where the surface normal points out.

When calculating the actual integral, the solid angle must be parametrized somehow. In numeric integration, one can approximate a solid angle by a triangle cover, where each corner of the triangle lies in the sphere. One can evaluate the function to integrate in the corner points and calculate the average for each triangle. Then the integral over the solid angle can be approximated by multiplying the averages by respective triangle areas, and summing the products together. Formally:

$$\begin{aligned} &x_i, \text{such that} |x_i| = 1 \\ &T_j = (i_0, i_1, i_2) \\ &\int_\omega f(\Omega) \mathrm{d}\Omega \approx \sum_j \text{area}(T_j) \frac{f(x_{T_{j,0}}) + f(x_{T_{j,1}}) + f(x_{T_{j,2}})}{3} \end{aligned} \tag{A.10}$$

where the triangle set approximates the solid angle ($\bigcup_j T_j \approx \omega$) and area$(T_j)$ evaluates the area of a single triangle defined by the coordinates $x_{T_{j,0}}$, $x_{T_{j,1}}$, and $x_{T_{j,2}}$.

[1] Albeit a bit misleading, this practice is established in mathematics and in radiance analysis. The numbers come from the areas — 4π is the surface area of the unit sphere, and 2π of the unit hemisphere.

In analytical integration, the solid angle must also be parametrized. Typical is to parametrize the solid angle with two angular parameters such that:

$$\Omega(\theta,\phi) = \begin{pmatrix} \sin\theta\cos\phi \\ \sin\theta\sin\phi \\ \cos\theta \end{pmatrix}, \quad 0 \leq \theta \leq \pi, \quad 0 \leq \phi \leq 2\pi \tag{A.11}$$

When integrating by using a parametrization, one must adjust the integration with the "speed" of the parametrization — compare with $\int_0^1 f(x)\mathrm{d}x = \int_0^2 f(y/2)\frac{1}{2}\mathrm{d}y$ (see eq. A.9). For the spherical coordinates, the adjusted form is

$$\int_\omega f(\Omega)\mathrm{d}\Omega = \iint_\omega f(\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)\sin\theta\mathrm{d}\theta\mathrm{d}\phi \tag{A.12}$$

where $\sin\theta\mathrm{d}\theta\mathrm{d}\phi$ is the "speed" adjusted measure. We justify this by noting that parametrized surface integrals are calculated by

$$\int_A f(x)\mathrm{d}x = \iint_A f(x(\theta,\phi)) \left| \frac{\partial x(\theta,\phi)}{\partial\theta} \times \frac{\partial x(\theta,\phi)}{\partial\phi} \right| \mathrm{d}\theta\mathrm{d}\phi \tag{A.13}$$

where:

$$\left| \frac{\partial x(\theta,\phi)}{\partial\theta} \times \frac{\partial x(\theta,\phi)}{\partial\phi} \right| = \left| \begin{pmatrix} \cos\theta\cos\phi \\ \cos\theta\sin\phi \\ -\sin\theta \end{pmatrix} \times \begin{pmatrix} -\sin\theta\sin\phi \\ \sin\theta\cos\phi \\ 0 \end{pmatrix} \right| =$$

$$\left| \begin{pmatrix} \cos\theta\cos\phi \\ \cos\theta\sin\phi \\ -\sin\theta \end{pmatrix} \right| \left| \begin{pmatrix} -\sin\theta\sin\phi \\ \sin\theta\cos\phi \\ 0 \end{pmatrix} \right| =$$

$$\sqrt{(\cos\theta\cos\phi)^2 + (\cos\theta\sin\phi)^2 + (\sin\theta)^2}\sqrt{(\sin\theta\sin\phi)^2 + (\sin\theta\cos\phi)} =$$

$$\sqrt{(\cos^2\phi + \sin^2\phi)\cos^2\theta + \sin^2\theta}\sqrt{(\cos^2\phi + \sin^2\phi)\sin^2\theta} =$$

$$\sqrt{1}\sqrt{1\sin^2\theta} = |\sin\theta| = \sin\theta \tag{A.14}$$

The equivalency $|() \times ()| = |()||()|$ holds because the vectors are orthogonal. The orthogonality can be seen by taking the dot product:

$$\begin{pmatrix} \cos\theta\cos\phi \\ \cos\theta\sin\phi \\ -\sin\theta \end{pmatrix} \cdot \begin{pmatrix} -\sin\theta\sin\phi \\ \sin\theta\cos\phi \\ 0 \end{pmatrix} = \\ -\cos\theta\cos\phi\sin\theta\sin\phi + \cos\theta\sin\phi\sin\theta\cos\phi - \sin\theta\cdot 0 = 0 \quad \text{(A.15)}$$

The $|\sin\theta| = \sin\theta$ equivalency is obvious because $\sin\theta$ is always non-negative ($0 \leq \theta \leq \pi$).

However, (θ, ϕ)-parametrization has a weakness when integrating over delta functionals, which is briefly addressed next.

Remark A.2.3 *Dirac Delta Functionals in Solid Angle Integration*

The Dirac delta functional $\delta(x)$ is defined such that[2]

$$1)\ \int_A \delta(x)\mathrm{d}x = \begin{cases} 1 & \exists\varepsilon > 0 :\ B(0,\varepsilon) \subset A \\ 0 & \exists\varepsilon > 0 :\ B(0,\varepsilon) \subset A^C \end{cases} \quad \text{(A.16)}$$

$$2)\ \delta(x) = 0 \quad \forall x \neq 0$$

Further, it is not hard to show that

$$\delta(f(x)) = \sum_{x_0 \in f^{-1}(\{0\})} \frac{1}{|f'(x_0)|}\delta(x - x_0) \quad \text{(A.17)}$$

Unfortunately, δ is dependent on the parametrization, which can be seen from the equation above. Therefore, in reparametrization one needs to adjust the delta functionals also.

[2]It is also common to define the $\delta(x)$ in $\mathbb{R}^n$ as

$$\delta(x) = \lim_{k\to\infty} k^n \prod_j^n \left[\max\left\{0, 1 - \left|kx_j\right|\right\}\right]$$

However, it is worth noting that there exists many equivalent definitions.

The integral

$$\int_{\omega} \delta(\Omega - \Omega_0) \mathrm{d}\Omega \tag{A.18}$$

is well-defined. After (θ, ϕ) parametrization, it is of form

$$\iint_{\omega} \frac{1}{\sin\theta} \delta(\theta - \theta_0)\delta(\phi - \phi_0) \sin\theta \mathrm{d}\theta\mathrm{d}\phi = \iint_{\omega} \delta(\cos\theta - \cos\theta_0)\delta(\phi - \phi_0) \sin\theta \mathrm{d}\theta\mathrm{d}\phi \tag{A.19}$$

This form, however, has singularities when $\theta_0 \in \{0, \pi\}$.

Definition A.2.4 *Monte Carlo Integration Method*

Monte Carlo integration, also known as random sampling integration, may be generally defined as:

$$\int_A f(x)\mathrm{d}x \approx \frac{\int_A 1\mathrm{d}x}{\operatorname{card} X} \sum_{x \in X} f(x) \tag{A.20}$$

where f is a piece-wise continuous function and X is a finite set of uniformly distributed random samples from A. The equation can be transformed for non-uniform distribution as

$$\int_A f(x)\mathrm{d}x \approx \frac{1}{\operatorname{card} X} \sum_{x \in X} \frac{f(x)}{\rho(x)} \tag{A.21}$$

where $\rho(x)$ is the probability density of the distribution. Non-uniform distributions can be used to decrease variance, or to sample functions which contain delta functionals.

Further introduction to Monte Carlo integration can be found in *e.g.* [31] or [9].

A.3 Linear Operator Analysis

Theorem A.3.1 *Theorem of Inverse Operators*

Let A be a linear operator $X \to X$, where X is a complete norm space. If $\|A\| < 1$, then $(I-A)^{-1}$ exists and

$$(I-A)^{-1} = \sum_{k=0}^{\infty} A^k = I + A + A^2 + \ldots + \tag{A.22}$$

Proof. Because $\left\|A^k\right\| \leq \|A\|^k$, the series $\sum_k \|A_k\|$ converges which consequently leads that $\sum_k A^k$ converges in complete spaces. Now, noting that

$$(I-A)(I+A+\ldots+A^k) = (I-A^{n+1}) \tag{A.23}$$

and by multiplying both sides of equation A.22 with $(I-A)$, one obtains:

$$\begin{aligned} (I-A)(I-A)^{-1} &= (I-A)\lim_{k\to\infty}\sum_{j=0}^{k} A^j \\ I &= I - \lim_{k\to\infty} A^{k+1} \end{aligned} \tag{A.24}$$

Now, because $\left\|A^k\right\| \leq \|A\|^k$, and $\|A\| < 1$ we may conclude that $\|A\|^k \to 0$, and thus $A^k \to 0$ when $k \to \infty$. This proves that the sum equals to $(I-A)^{-1}$. □

The proof with the essential backgrounds can be found in more detail in *e.g.* [15].

Definition A.3.2 *Fredholm Integral Equation of Second Kind*

The Fredholm integral equation of second kind is of form

$$\phi(x) = f(x) + \int_a^b K(x,t)\phi(t)\mathrm{d}t \tag{A.25}$$

and it can be solved by:

Theorem A.3.3 *Integration Equation Neumann Series*

The integration equation Neumann series (ϕ_k) is defined as

$$\phi_k(x) = \sum_{j=0}^{k} u_k(x) \tag{A.26}$$

where

$$\begin{aligned} u_0(x) &= f(x) \\ u_1(x) &= \int_a^b K(x,t_1) f(t_1) \mathrm{d}t_1 \\ u_2(x) &= \int_a^b \int_a^b K(x,t_2) K(t_2,t_1) f(t_1) \mathrm{d}t_1 \mathrm{d}t_2 \\ &\cdots \\ u_k(x) &= \int_a^b \cdots \int_a^b K(x,t_k) K(t_k,t_{k-1}) \cdots K(t_2,t_1) f(t_1) \mathrm{d}t_1 \cdots \mathrm{d}t_k \end{aligned} \tag{A.27}$$

The limit $\phi(x) = \lim_{k\to\infty} \phi_k(x)$ is a solution to the Fredholm integral equation of second kind.

Proof. The function space, where $f(x)$ and $\phi(x)$ belongs, is a Banach space, denoted here by X. The integral operator $\int K(x,t) \cdot \mathrm{d}t$ is a linear operator $X \to X$. Denote the operator by B. Thus, $Bf(x) = \int K(x,t) f(t) \mathrm{d}t$. Let us also assume $\|B\| < 1$.

Now, the Neumann series limit can be written as

$$\phi(x) = \sum_{j=0}^{\infty} B^j f(x) \tag{A.28}$$

and thus, the equation A.25 can be written as

$$\begin{aligned} \sum_{j=0}^{\infty} B^j f(x) &= f(x) + B \sum_{j=0}^{\infty} B^j f(x) \\ &= f(x) + \sum_{j=1}^{\infty} B^j f(x) \\ &= \sum_{j=0}^{\infty} B^j f(x) \end{aligned} \tag{A.29}$$

And because $\|B\| < 1$, the sum converges. $\square$

Fredholm integral equation of second kind and Neumann series solution are introduced in greater detail in *e.g.* [11] and [33].

Appendix B

A Brief Note on BDRFs and Lambertian Diffuse Reflections

The Lambertian diffuse reflection model is used to define the *ideal* diffuse reflection. The model was initially constructed for heat flow analysis between blackbody objects in the field of thermodynamics. The model was derived using energies.

In acoustics, subject to the continuous physics, a surface reflection model satisfying the diffuse reflection requirement has many practical purposes. With the model, one can model diffuse reflections using arbitrarily shaped reflectors, even planar polygon reflectors. A reflection model exists for energy fields, but could one exist for pressure disturbances? If it existed, one could calculate impulse responses of diffuse reflections, as one can now calculate impulse responses of specular reflections.

The reflection model would be described using a BRDF for pressure fields.[1] Note that it is sufficient to search the solution in BRDFs instead of BSSRDFs, because of the reflector shape independency. Before rushing into applications, one question must be raised: How should the BRDF be defined?

[1] This BRDF operates with the amplitude and phase of the incident pressure disturbances.

It is obvious that the exitant emission pattern of the reflection may not depend on the incident pattern — except perhaps the phase shift. The reasoning is similar to the reasoning with the BRDFs for incident energy.

Finally, if we can define a BRDF that produces diffuse reflection pattern for planar waves incident in any direction, we can extend that BRDF for incident radiation in all directions because the exitant pattern may not depend on the incident pattern. Then, the extended BRDF would produce diffuse exitant fields for any incident pattern, or at least almost any incident pattern.

So, we ask what the exitant pattern should be for a planar wave arriving orthogonally to the surface. The BRDF producing diffuse reflections must be applicable for surfaces of any shape, as the Lambertian model implies. Specifically, huge planar surfaces should be no different.

But, observe any specific non-orthogonal direction of the reflected wave front to remark that the wave front is planar when the distance to the surface is small compared to the dimensions of the huge surface. Paradoxically, the wave front propagates along the direction of the surface normal and, in addition, the wave front propagates at subsonic speeds.[2] This is clearly in contradiction with the wave equation which predicts that the wave front must propagate exactly at the speed of sound. Therefore, the BRDF must be zero for any non-orthogonal exitant direction which certainly does not produce diffuse exitant wave fronts.

The brief examination shows that a general BRDF for diffuse reflections for pressure fields cannot be found, at least not in the space of conventional functions. It might still exist in a more general space, but that we will leave here open.

[2]To verify this, assume an infinite planar surface. Then, choose a direction and simultaneously, send a particle from every point in the surface in the chosen direction. Notice that the emitted particles form a plane — wave front — and that the plane moves in the direction of the surface normal. The speed of the wave front is $\mathbf{c}\cos\theta$ where θ is the angle between the surface normal and the chosen direction.

Bibliography

[1] Jont B. Allen and David A. Berkley. Image method for efficiently simulating small-room acoustics. *Journal of the Acoustical Society of America*, 65(4):943–950, April 1979.

[2] John C. Allred and Albert Newhouse. Applications of the Monte Carlo method to architectural acoustics. *Journal of the Acoustical Society of America*, 30(1):1–3, January 1958.

[3] John C. Allred and Albert Newhouse. Applications of the Monte Carlo method to architectural acoustics. II. *Journal of the Acoustical Society of America*, 30(1):903–904, October 1958.

[4] Jeffrey G. Borish. Extension of the image model to arbitrary polyhedra. *Journal of the Acoustical Society of America*, 75(6):1827–1836, June 1984.

[5] Robert W. Boyd. *Radiometry and the Detection of Optical Radiation*. John Wiley & Sons, 1983. ISBN 0-471-86188-X.

[6] Dietrich Braess. *Finite Elements*. Cambridge University Press, second edition, 2001. ISBN 0-521-01195-7.

[7] Michael F. Cohen and John R. Wallace. *Radiosity and Realistic Image Synthesis*. Morgan Kaufmann Publishers, 1993. ISBN 0-12-178270-0.

[8] Trevor J. Cox and Peter D'Antonio. *Acoustic Absorbers and Diffusers*. Spon Press, 2004. ISBN 0-415-29649-8.

[9] Philip Dutré, Philippe Bekaert, and Kavita Bala. *Advanced Global Illumination*. A K Peters, Ltd., 2003. ISBN 1-56881-177-2.

[10] Per-Anders Forsberg. Fully discrete ray tracing. *Applied Acoustics*, 18(6):393–397, 1985.

[11] Ronald B. Guenther and John W. Lee. *Partial Differential Equations of Mathematical Physics and Integral Equations*. Dover Publications Inc., 1996. ISBN 0-486-68889-5.

[12] Renate Heinz. Binaural room simulation based on an image source model with addition of statistical methods to include the diffuse sound scattering of walls and to predict the reverberant tail. *Applied Acoustics*, 38:145–159, 1993. Special Issue on Computer Modelling and Auralisation of Sound Fields in Rooms.

[13] James T. Kajiya. The rendering equation. In *Proceedings of Computer Graphics and Interactive Techniques (SIGGRAPH'86)*, volume 20(4), pages 143–150. ACM Press, 1986.

[14] Mendel Kleiner, Bengt-Inge L. Dalenbäck, and Peter Svensson. Auralization - An overview. *Journal of the Audio Engineering Society*, 41(11):861–875, November 1993.

[15] Erwin Kreyszig. *Introductory Functional Analysis with Applications*. John Wiley & Sons. Inc., 1978. ISBN 0-471-50731-8.

[16] Asbjørn Krokstad, Svein Strøm, and Svein Sørsdal. Calculating the acoustical room impulse response by the use of a ray tracing technique. *Journal of Sound and Vibration*, 8(1):118–125, 1968.

[17] Heinrich Kuttruff. Simulierte Nachhallkurven in Rechteckraumen mit Diffusem Schallfeld. *Acustica*, 25:333–342, 1971.

[18] Marcel Minnaert. The reciprocity principle in lunar photometry. *Astrophysics Journal*, 93:403–410, 1941.

[19] Sanjit K. Mitra. *Digital Signal Processing*. McGraw-Hill, second edition, 2001.

[20] Fred E. Nicodemus, J. C. Richmond, J. J. Hsia, I. W. Ginsberg, and T. Limperis. *Geometrical Considerations and Nomenclature for Reflectance*. National Bureau of Standards, October 1977.

[21] Eva-Marie Nosal, Murray Hodgson, and Ian Ashdown. Improved algorithms and methods for room sound-field prediction by acoustical radiosity in arbitrary polyhedral rooms. *Journal of the Acoustical Society of America*, 116(2):970–980, August 2004.

[22] Allan D. Pierce. *Acoustics – An Introduction to Its Physical Principles and Applications*. Acoustical Society of America, 1989. ISBN 0-88318-612-8.

[23] Jean-Dominique Polack. Playing billiards in the concert hall: The mathematical foundations of geometrical room acoustics. *Applied Acoustics*, 38:235–244, 1993.

[24] James E. Proctor and P. Yvonne Barnes. NIST high accuracy reference reflectometer-spectrophotometer. *Journal of Research of the National Institute of Standards and Technology*, 101(5):619–627, September–October 1996.

[25] Thomas D. Rossing, Richard F. Moore, and Paul A. Wheeler. *The Science of Sound*. Benjamin Cummings, third edition, 2002. ISBN 0-8053-8565-7.

[26] Walter Rudin. *Real and Complex Analysis*. McGraw-Hill Book Company, 1987. ISBN 0-07-100276-6.

[27] Lauri Savioja, Jyri Huopaniemi, Tapio Lokki, and Riitta Väänänen. Creating interactive virtual acoustic environments. *Journal of the Audio Engineering Society*, 47(9):675–705, 1999.

[28] U. Peter Svensson, Roger I. Fred, and John Vanderkooy. An analytic secondary source model of edge diffraction impulse responses. *Journal of the Acoustical Society of America*, 106(5):2331–2344, 1999.

[29] U. Peter Svensson and Ulf R. Kristiansen. Computational modelling and simulation of acoustic spaces. In *The Proc. of the AES 22nd International Conference*, pages 11–30. Audio Engineering Society, Inc., June 2002.

[30] Alan Watt. *3D Computer Graphics*. Addison-Wesley Publishing Company, second edition, 1993. ISBN 0-201-63186-5.

[31] Stefan Weinzierl. Introduction to Monte Carlo methods. *arxiv.org web*, June 2000. http://arxiv.org/abs/hep-ph/0006269/.

[32] Eric W. Weisstein. Delta function. From *MathWorld — A Wolfram Web Resource*. http://mathworld.wolfram.com/DeltaFunction.html.

[33] Eric W. Weisstein. Integral equation Neumann series. From *MathWorld — A Wolfram Web Resource*. http://mathworld.wolfram.com/IntegralEquationNeumannSeries.html.

[34] Hugh D. Young and Roger A. Freedman. *University Physics, Extended Version with Modern Physics*. Addison-Wesley Publishing Company Inc, ninth edition, 1996. ISBN 0-201-31132-1.