THE SONEL MAPPING ACOUSTICAL MODELING METHOD

by

BILL KAPRALOS

A DISSERTATION SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN COMPUTER SCIENCE
YORK UNIVERSITY
TORONTO, ONTARIO

September, 2006

Abstract

This work investigates the application of photon mapping to model environmental acoustics. The resulting acoustic *sonel mapping* technique is a Monte-Carlo based approach that can be used to model acoustic environments while accounting for diffuse and specular acoustic reflections as well as diffraction effects. This modeling is performed in an efficient manner in contrast to available deterministic techniques. The sonel mapping approach models many of the subtle interaction effects required for realistic acoustical modeling.

Acknowledgments

Wow, the dissertation is complete! Now comes the difficult task of remembering and thanking all those that have helped me throughout my Ph.D journey. First and foremost, I would like to thank my supervisor Professor Michael Jenkin for all his help, support and very, very generous financial assistance throughout my Ph.D (and M.Sc) studies at York University. Despite his very busy schedule, Professor Jenkin always made time for any questions I may have had regarding my work and always had an answer! His door was always open and his email responses were amazingly prompt! I also very much appreciate all the advice and guidance that he provided me with during my search for a faculty position as well as his advice and guidance after obtaining a faculty position. I would like to also thank my co-supervisor Professor Evangelos Milios for all his help, support and guidance throughout all of my academic career. Professor Milios actually provided me with my first research opportunity during my first undergraduate year (an in-depth search of "how animals climb") and opened the doors to a "whole new research world" for me. As they say, the rest is history! A thank you also goes out to the other members of my dissertation committee who were very accommodating in setting up a defense during the vacation month of July and for their comments and feedback: Professor Patrick Dymond, Professor Laurence Harris, Professor William Martens and Professor Richard Wildes. Professor Martens (external member) came to Toronto specifically for my defense. He provided me with plenty of useful feedback and suggestions on several occasions during the early stages of my work and also provided very useful comments and suggestions for future work. I also appreciate the help and advice (on a variety of topics - not only research related) throughout the years from Professor John Tsotsos. A thank you also goes out to Professor Wolfgang Stürzlinger for suggesting the photon mapping framework and for his feedback and advice on the negatives of radiosity-based approaches.

I would also like to thank everyone else who has assisted me with many aspects of my work - and there are of course many (hopefully I don't forget any-

one)! Dan Zikovitz and Kosta Derpanis have been especially helpful with various aspects of my research including experimental set-ups, statistical analysis (thanks Dan!), listening and commenting on various aspects of my research. Of course these discussions typically took place over coffee so I thank both of them for the caffiene addiction I have now developed! I shared an office with Dan for at least seven years and I am still trying to figure out how he put up with my typing (which was essentially slamming of the keyboard)! I would also like to thank Erich Leung, Nathan Mekuz, Andrew Hogue and Andrei Rotenstein for all their help and support throughout, Antonin Pribetic for the all his proof-reading help, Professors Rob Allison and Minas Spetsakis for their help with signal processing and physics related issues, Agnieszka Kopinska for all her support and encouragement and my family (parents, brother and sisters) for all their support during this endeavor, Franck van Breugel and Hamzeh Roumani for all their insights on teaching. Thanks also to all the administrative staff (Sandra Fyffe, Clara Masaro, Mel Poteck, Maria Szucs and Lora Zuech) and technical staff (Seela Balkissoon, Paul Griffith, Oana Ionita, Jason Keltz, Matt Robinson and Ulya Yigit) in the Department of Computer Science and Engineering at York University for all their help on many occasions.

Finally, I would like to thank York University President Lorna Marsden for providing me the York University President's Dissertation Scholarship, the Natural Sciences and Engineering Research Council of Canada (NSERC), for providing me a PGS B scholarship, the Government of Ontario for providing me the Ontario Graduate Scholarship in Science and Technology (OGSST) and the Institute for Robotics and Intelligent Systems (IRIS) for providing me with a Student Fellowship.

Contents

List of Figures

List of Tables

Abbreviations

3D	Three-Dimensional
BEM	Boundary Element Method
BRDF	Bidirectional Reflectance Distribution Function
BRIR	Binaural Room Impulse Response
EDP	Energy Discontinuity Percentage
FDTDM	Finite Difference Time Domain Method
FEM	Finite Element Method
HRTF	Head Related Transfer Function
ILD	Interaural Level Difference
ITD	Interaural Time Difference
IHL	Inside-the-Head Localization
KEMAR	Knowles Electronic Manikin for Auditory Research
RIR	Room Impulse Response
RT	Reverberation Time
SPL	Sound Pressure Level
UTD	Uniform Theory of Diffraction
VBAP	Vector Base Amplitude Panning
WFS	Wave-Field Synthesis

Notation

Greek Alphabet

α	Surface absorption coefficient
δ	Surface diffuse reflection coefficient
ϕ_s	Vertical angle of the secondary source within some Fresnel zone
λ	Wavelength
ω	Angular frequency ($\omega = 2\pi f$)
ω_d	Ideal diffuse (Lambertian) reflection direction
ω_i	Incoming direction
ω_s	Ideal specular reflection angle
ρ	Radius of the initial wavefront emitted from the sound source
ξ	Uniformly distributed random number
θ_{res}	Angular resolution of sampled positions within a Fresnel zone
$\theta_{receiver}$	Horizontal angle between the receiver and the sound source

Latin Alphabet (Capital)

$E_\circ$	Initial sound source energy
E_A	Energy of the secondary sources within a differential area dS
E_i	Total energy of all the secondary sources within the ith Fresnel zone
E_{sonel}	Sonel energy
E_{total}	Sum of the energy reaching the receiver from each Fresnel zone
I_m	Monte-Carlo approximation of integral I
$K(\theta)$	Obliquity factor
L_{direct}	Level of the direct sound reaching a receiver
L_e	Self-emitted radiance
L_i	Incoming radiance
$L_{i,c}$	Incoming radiance via caustics
$L_{i,d}$	Incoming radiance via diffuse reflections
$L_{i,l}$	Incoming radiance that comes directly from light sources
$L_\circ$	Outgoing radiance
L_{reverb}	Level of the reverberant sound reaching a receiver
L_r	Reflected radiance
L_s	Sound source level
N_{vis}	Total number of positions sampled within a Fresnel zone to determine its visibility relative to the receiver
N_{freq}	Number of frequency bands considered by the simulation
N_{sonel}	Number of sonels emitted by a sound source
N_{zones}	Total number of Fresnel zones for a particular sound source/receiver configuration
RT_{est}	Reverberation time as estimated by sonel mapping
RT_{pre}	Reverberation time as predicted by Sabine's formula
Z_i	ith Fresnel zone
Z_{init}	Initial Fresnel zone

Latin Alphabet (Non-Capital)

dE	Energy reaching the receiver from the secondary sources from a differential area dS within a Fresnel zone
dS	Ring-shaped differential area within a Fresnel zone
$\%diff$	Percent difference
dir_{sonel}	Incoming incidence sonel direction (θ, ϕ) at the point of intersection between the sonel and the surface
f_c	Center frequency of a particular frequency band
f_l	Lower frequency of a particular frequency band
f_r	Bidirectional reflectance distribution function
$f_{r,S}$	Bidirectional reflectance distribution function for the specular component
$f_{r,D}$	Bidirectional reflectance distribution function for the diffuse component
f_{sonel}	Sonel frequency
f_u	Upper frequency of a particular frequency band
k	Wave-number ($k = 2\pi\lambda$)
$\vec{n}$	Surface normal vector
n_{actual}	Actual number of sonels used by the density estimation algorithm
n_{max}	Maximum number of sonels considered by the density estimation algorithm
n_{vis}	Number of sampled positions on a Fresnel zone visible to the receiver
$p_\circ$	Intersection point between the sphere representing the initial wavefront and the line between the receiver and the sound source
p_1	Position of a secondary source within the first Fresnel zone
p_{edge}	Position of a sonel on an edge when it is to be diffracted
$r_\circ$	Distance between the receiver and the initial wavefront
r_{direct}	Direct distance between sound source and receiver
r_{init}	Distance between the receiver and the secondary source within the initial Fresnel zone
r_k	Radius of the sphere representing the receiver
r_{perp}	Perpendicular (shortest) distance between the sonel incident position and an edge
r_{ray}	Distance an acoustic visibility ray has traveled from the time it was emitted, to the time of the last interaction point
r_{reverb}	Reverberation distance

Latin Alphabet (Non-Capital) Continued

r_{sonel}	Distance a sonel has traveled from the time it was emitted, to the time of the last interaction point
r_s	Radius of the circle centered about the sonel incidence point used in the sonel map density estimation algorithm
r_{sr}	Distance between the sound source and receiver
t_{edp}	Time required to estimate a particular reverberation time using an EDP termination criterion
t_{rus}	Time required to estimate a particular reverberation time using a Russian roulette termination criterion
t_{ray}	Time for a visibility ray emitted by the receiver to reach some point p
t_{rir}	Echogram resolution (spacing between “bins”)
t_{sonel}	Time for a sonel emitted from a sound source to reach the receiver
v_i	Visibility of Fresnel zone i relative to a receiver
v_s	Velocity of sound in air

Chapter 1

Introduction

In our natural surroundings we hear sounds from different locations, from different distances and after they have interacted with a variety of objects. We are capable of distinguishing individual sounds by pitch, tone, loudness, and by their location in space. Our ability to extract spatial information from the sounds we hear provides us with detailed information about our surroundings, assisting us in determining both the distance to and direction of objects [190]. Furthermore, hearing serves to guide our more finely tuned visual attention system, thereby easing the burden on the visual system [43]. The spatial sounds that are present in our environment provide us with detailed information regarding our surroundings and at times are crucial to our survival.

Given the importance of spatial hearing to humans, incorporating spatial-

ized sound cues in realistic simulations such as immersive virtual environments seems obvious. In fact, doing so can be beneficial for a variety of reasons. Spatial sound cues can add a better sense of "presence" or "immersion", they can compensate for poor visual cues (graphics), lead to improved object localization and, at the very least, add a "pleasing quality" to the simulation [4, 164]. Martens and Woszczyk [115] describe a set of human-centered guidelines that focus on enhancing the user's sense of presence in an interactive virtual acoustical environment. As a consequence, sound has been incorporated into a variety of human-machine interfaces to convey alarms, warnings, status information and messages [35]. *Foley sounds*, associated with particular visual imagery such as footsteps, a door opening, glass breaking, a ball bouncing etc. are added to the post production of animations (see [55, 139]), films and computer games. Although the inclusion of such sounds can lead to greater realism and quality [55], in contrast to the sounds present in our natural surroundings such sounds typically lack spatial information. In fact, spatial sound cues are often overlooked by the majority of immersive virtual environments where historically, emphasis has been placed on the visual senses instead [37, 43]. Furthermore, when present, the spatial sound cues that are present do not necessarily reflect natural cues. Many systems that do convey sound information do so poorly, typically assuming that all interactions between a sound wave and objects/surfaces in the environment are specular reflections, despite that

in our natural settings, acoustical reflections may be diffuse and there may also be diffractive and refracted components to the sounds we hear as well. Failure to accurately model all these phenomena leads to a decrease in the spatialization capabilities of the system, ultimately leading to a decrease in performance and a decrease in presence or immersion [175].

Simulating the propagation of sound through an environment, from the time it is emitted from a sound source until the time it reaches a receiver while interacting with any objects/surfaces it may encounter, is known as *acoustical modeling*. In addition to virtual environments and virtual reality in general, there are many applications for which acoustical modeling is beneficial. These include the design or refurbishment of acoustical concert halls, auditoria and public buildings (e.g., offices), where the primary goal is to provide optimum speech intelligibility and sound quality [180]. For such applications, acoustical modeling allows the concert hall, auditorium etc. to be "heard" via virtual acoustical rendering and problems diagnosed and addressed prior to physical construction, potentially leading to tremendous cost savings and reduced construction time.

Collectively, "the process of rendering audible, by physical or mathematical modeling, the sound field of a sound source in space, in such a way as to simulate the binaural listening experience at a given position in the modeled space" is known as *auralization* [95]. The goal of auralization is to recreate

a particular listening environment, taking into account the acoustics of the environment and the characteristics of the listener. Auralization is typically accomplished by determining the *binaural room impulse response* (BRIR). The BRIR represents the response of a particular acoustical environment to sound energy and captures the room acoustics for a particular sound source and listener configuration. Once obtained, the BRIR can be used to filter, typically through a convolution process, the desired anechoic sound. When this filtered sound is presented to a listener the original sound environment is recreated.

The BRIR can be considered as the signature of the room response for a particular sound source and human receiver. Although interlinked, for simplicity and reasons of practicality, the room response and the response of the human receiver are commonly determined separately and combined via a post-processing operation to provide an approximation to the actual BRIR [95]. The response of the room is known as the *room impulse response* (RIR) and captures the reflection properties (reverberation), diffraction, refraction, sound attenuation and absorption properties of a particular room configuration (e.g., the environmental context of a listening room or the "room acoustics"). The response of the human receiver captures the direction dependent effects introduced by the listener due to the listener's physical make-up (e.g., pinna, head, shoulders neck and torso) and is known as the *head related transfer function* (HRTF).

This work develops an acoustical modeling algorithm, termed *sonel map-*

ping, that estimates the time and frequency dependent *echogram* (the temporal acoustical energy distribution) of a particular environment. The echogram can be converted to an equivalent room impulse response function through a post-processing operation [101]. Although the capture and modeling of HRTFs is a fascinating subject, it will not be addressed in any detail here. Greater details regarding the estimation of the HRTF can be found in [17, 37, 191].

1.1 Goals of this Dissertation

Acoustical modeling of even small, simple environments is a complex, computationally expensive and time consuming task for all but the simplest environments. This is due to the potentially large number of complex interactions encountered by sound waves as they propagate from the sound source to the receiver. Accurately recreating the interaction of sound with the objects/surfaces the sound may encounter as it propagates through the environment is extremely difficult and beyond our current analytical and computational reach except for simple environments. However, attempts have been made to simplify rendering in order to minimize computational needs while maintaining perceptual accuracy [112]. Given the similarities that exist between the fields of computer graphics (image synthesis) and acoustical modeling, this dissertation investigates the application of suitably modified computer graphics and optics-based modeling methods and techniques to accurately model environ-

mental acoustics. By accounting for the differences between the propagation of sound and light as well as differences in how propagating sound waves interact when they encounter objects in the environment, a sound synthesis method is developed that models many of the complex effects that propagating acoustical signals encounter in the environment.

The goal of *image synthesis* is the creation of realistic images, images that are indistinguishable from the environment they model [45]. In the field of image synthesis, two general models have emerged to solve for the illumination of a scene. The first model is known as the *local illumination model* where the shading of a particular point is determined by considering the light coming directly from a light source and the point itself only, ignoring the interaction of any light reaching it indirectly via reflections from other objects (surfaces) in the environment. Given the influence of indirect lighting in the majority of natural scenes, local illumination models do not provide an accurate simulation of a particular scene and the resulting renderings are easily noticed as "computer generated" [169]. Under the *global illumination model*, the indirect light resulting from the inter-reflection of light between surfaces in the environment is also considered to determine the shading at a point. Global illumination models produce far more accurate renderings and have thus received a great deal of research effort over the last couple of decades, especially with the introduction of the *radiosity* method [46, 70] in the early 1980s. Radiosity is a

view-independent technique based on the principle of energy conservation in a closed environment where all surfaces are assumed to be Lambertian diffuse. The model is sub-divided into small "patches" and the distribution of light is found by solving a set of linear equations for the exchange of light between all patches. Essentially, radiosity is one approach to solving Kajiya's rendering equation [86]. The rendering equation gives the necessary conditions for equilibrium of light transport and is used to calculate the outgoing radiance, the energy carried by electromagnetic radiation, per unit source area per unit solid angle, at any location within model [84, 86]. The rendering equation is itself a potentially complex and difficult to solve integral and rather than solving it analytically it is typically approximated using finite element methods (FEMs) in radiosity approaches. Radiosity methods have received a great deal of attention and for many years were the predominant method for computing global illumination. However, radiosity approaches are extremely computationally intensive and assume diffuse reflections only, ignoring any specular reflections and associated phenomena such as caustics. In an attempt to limit computational needs and time requirements, various alternatives to mesh-based finite element representations have been developed. Illumination maps, whereby a texture map with illumination values is used to represent and store the irradiance in a model, can be used. Once the map has been constructed, irradiance at view-independent locations in the model can be "looked-up" without the need

to re-compute the entire solution at run-time (however, view-dependent affects must still be modeled at run-time). Despite the potential computational savings, illumination maps suffer from various problems including: i) computing the resolution of the map, ii) computationally costly for complex models and iii) difficult to use on arbitrary surfaces [84]. Given the shortcomings associated with the radiosity-based techniques, several other global illumination techniques have been devised. One such technique is *photon mapping*, a two-pass "particle-based" probabilistic method developed by Jensen in 1995 [85].

Photon mapping has become very popular and is preferred over finite element techniques such as radiosity for a variety of reasons. In contrast to radiosity, photon mapping is independent of the scene geometry and thereby allows for the illumination of arbitrarily complex scenes to be computed. In contrast to the radiosity method, the scene does not have to be sub-divided. Furthermore, photon mapping relies on stochastic techniques such as Monte-Carlo integration methods and therefore, the solution can be made more accurate by increasing the number of samples at various points of the computation. Increasing the number of samples does increase storage requirements as well as the final rendering time however, the option is there and an accuracy vs. efficiency trade-off can nevertheless be made. Finally, photon mapping can handle complex interactions between light and a surface (e.g., specular and diffuse reflections and caustics (e.g., one or more specular reflections followed by a single

diffuse reflection)).

In the first pass of photon mapping, *photons* (named after the basic quantity of light) are emitted from each light source and traced through the scene until they interact with a surface. When photons encounter a diffuse surface, they are stored in a structure called a *photon map*. In the second stage, the scene is rendered using the information provided by the previously collected photon map to provide a quick estimate of the diffusely reflected illumination. Distribution ray tracing is employed to model specular effects (greater details regarding photon mapping are provided in Section 3.4).

Sonel mapping uses the same basic approach as photon mapping but takes into account the physical attributes of sound propagation, addressing the possible interactions when a propagating sound encounters a surface/object or obstruction in its path (e.g., specular or diffuse reflection, diffraction, absorption or refraction[1] as illustrated in Figure 1.1). Following the same strategy as used in photon mapping, rather than modeling the exact *mechanical wave* phenomena of sound propagation (e.g., particles in the medium as they move about in their equilibrium position), this process is approximated by emitting one or more "sound elements" from each sound source and tracing these sound elements through the scene until they encounter an object/surface (hereinafter, a

[1]Although refraction can occur, it is not as common as the other interactions when considering room acoustics since even regions in the medium with differing temperatures will eventually inter-mix into a single homogeneous region. Refraction can therefore typically be ignored by room acoustical modeling applications [48].

sound element will be referred to as a *sonel*, or *sonels* when considering more than one sound element). A sonel[2] can be viewed as a packet of information propagating from the sound source to the receiver, carrying the relevant information required to simulate the mechanical wave propagation. The information carried by each sonel includes the information used by photons in the photon mapping approach: position (x,y,z coordinates), direction to incidence and energy in addition to information specific to sound and sound propagation, including distance traveled and frequency.

Although the focus of this work is on acoustical modeling, sonel mapping is itself not specific to acoustical wave energy propagation. Rather, sonel mapping is a framework for energy propagation in general. Given the appropriate model parameters (for example, source emission functions, wave/surface interaction models etc.), sonel mapping can be used to model the propagation of any type of wave energy, be it sound, light etc.

1.2 Contributions of this Dissertation

The primary contribution of this dissertation is the development of the sonel mapping probabilistic (Monte-Carlo) based framework for modeling the propagation of wave energy in a particular environment. Sonel mapping allows

[2]Sonels are akin to Cohen and Koizumi's [42] "sound mixing elements" or *mixels* although the intended application of mixels is not for determining the acoustical room impulse response.

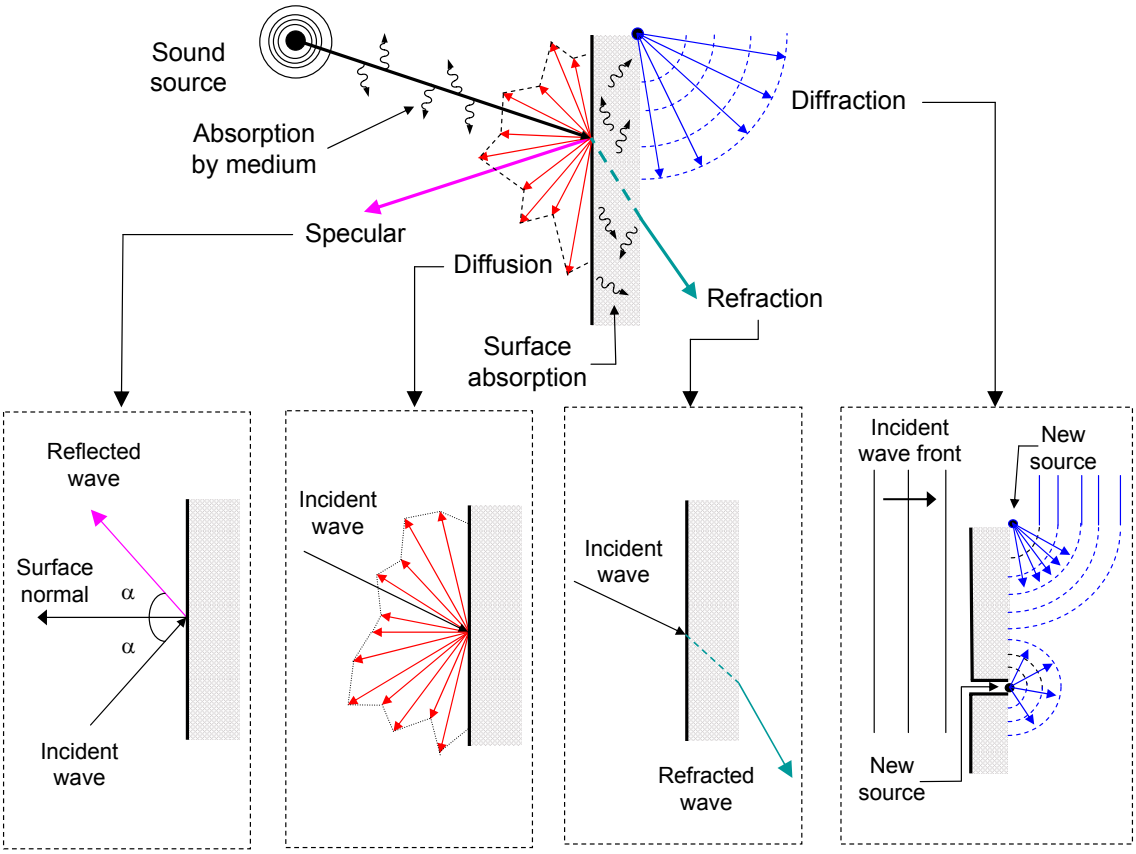


Figure 1.1: Acoustical reflection phenomena: specular reflection, diffuse reflection, refraction and diffraction.

for the modeling of the complex interactions between wave (sound energy for the purpose of this work) energy and objects/surfaces the sound energy may encounter as it propagates from a sound source to a receiver in a simple and efficient manner. Sonel mapping is independent of the scene geometry thereby allowing it to be used to model arbitrary complex scenes. In addition, it can handle any combination of interactions between a propagating sound element (sonel) and any objects/surfaces it may encounter including specular, diffuse, diffraction and their combination.

The probabilistic nature of sonel mapping and in particular, the application of a Russian roulette strategy to determine the type of interaction between a sonel and any object/surface it may encounter, allows for a quick and computationally feasible solution to room acoustical modeling. It allows the solution to be made more accurate by increasing the number of samples at various points of the computation. Although increasing the number of samples does increase storage requirements as well as the final rendering time, the option is there and an accuracy vs. efficiency trade-off can be made.

In addition to modeling specular and diffuse reflections, sonel mapping addresses the modeling of diffraction effects. Acoustical diffraction is accomplished using a modified version of the Huygens-Fresnel principle [79]. The Huygens-Fresnel principle assumes a propagating wavefront is composed of a number of secondary sources, fitting nicely into the sonel mapping probabilistic

framework whereby acoustical wave propagation is approximated by propagating sound "particles" (sonels) from a sound source and tracing them through the environment. Diffraction effects can be approximated in a very simple and efficient manner allowing them to be computed at interactive rates. Although the Huygens-Fresnel principle is a rather simple approach, it can describe a large number of diffraction configurations in an efficient manner.

1.3 Organization of this Dissertation

The remainder of this dissertation is organized as follows. Chapter 2 presents background information relevant to the rest of the dissertation including a brief overview on the physics of sound and sound propagation. Since any simulation of sound that will ultimately be presented to a human listener requires some underlying understanding of both the physics of sound and human sound localization, the chapter begins with a review of the physics of sound and sound propagation followed by brief review of human sound localization and perception. Emphasis is placed on the factors relevant to acoustical modeling. Chapter 2 ends with a detailed review of photon mapping. Chapter 3 also presents background information however, the focus of Chapter 3 is on acoustical modeling and in particular, methods and techniques used to approximate the room impulse response. In Chapter 4, the framework of the sonel mapping method is introduced. Details regarding the modifications made to the original photon

mapping method to allow for acoustical modeling are given in addition to features such as acoustical diffraction modeling. Simulations for various sound source, receiver and occluder configurations are presented in Chapter 5. The outcome of several of the simulations are compared to analytical results. A discussion of the outcomes of the simulations is presented in Chapter 6 where the strengths and limitations of the sonel mapping method are described. A summary and discussion of future work is presented in Chapter 7. Additional background information is provided in the Appendix. Appendix A provides details regarding the output of sound in a virtual display. Appendix B provides a review of Monte-Carlo methods including Russian roulette. Finally, mathematical details regarding the optics-based Huygens-Fresnel principle are provided in Appendix C. Included in this Appendix are several derivations specific to the acoustical diffraction method developed in this work.

Chapter 2

The Physics of Sound and Human Sound Localization

Spatial audio rests on the ability to control the auditory signals arriving at the listener's ears such that these signals are perceptually equivalent to the signals a listener receives in the environment being simulated [189]. When considering the design of any spatial sound system, it is therefore necessary to have some knowledge regarding the physics of sound and sound propagation in addition to human sound localization and perception [37]. As a result, although the focus of this dissertation is on the modeling of room acoustics for auralization, a brief introduction to both the physics of sound and human sound perception and localization is provided here. A complete discussion of both these topics is beyond the scope of this dissertation. Greater details regarding the physics of sound in general can be found in [75, 155] while information regarding human sound localization and perception can be found in [26, 134, 190, 196].

2.1 The Physics of Sound

Sound results from the rapid variations in air pressure caused from the vibrations of a vibrating object (e.g., a vibrating guitar string, human vocal chords etc.) [132]. As shown in Figure 2.1, sound waves consist of alternating regions of compression and rarefaction (e.g., "back and forth" motion) of the molecules comprising the medium [198] through which the sound propagates (typically air although sound can also propagate through other mediums as well, including water and metal). The molecules themselves do not move with the wave but rather oscillate about some fixed position. The wave itself propagates through the interaction of molecules in the medium. In other words, sound waves are a *mechanical wave* phenomenon and require a medium to propagate (e.g., sound waves cannot propagate in a vacuum). Considering a sound propagating through air, the air molecules surrounding the vibrating object will be compressed during forward motion and expanded during the object's backward motion. As these molecules are displaced, they "push or pull" the molecules neighboring them, causing these neighboring molecules to also be displaced from their resting position. This forward and backward movement of the molecules propagates throughout the entire medium, with each molecule displacing its neighbors. Sound waves may propagate in an *omni-directional* manner whereby the wave propagation is independent of direction (e.g., equal

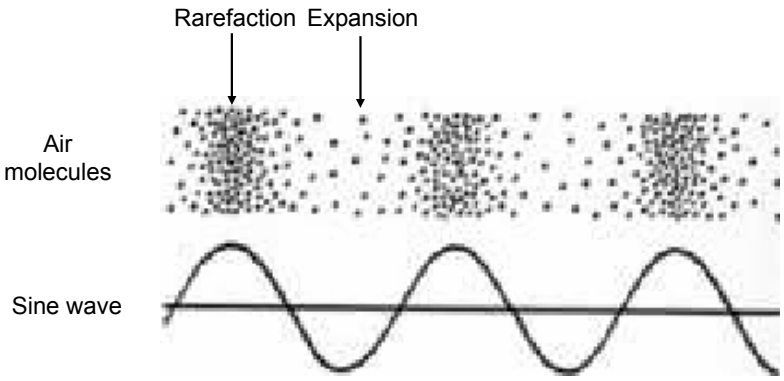


Figure 2.1: Sound waves consist of alternating regions of compression and rarefaction (e.g., "back and forth" motion) of the air molecules (top), corresponding to the "high" and "low" points of a "sine wave" (bottom).

in all directions) or may exhibit *directional* properties leading to wave propagation in a particular direction only.

A very simple type of sound wave is the sinusoid (sine wave) illustrated in Figure 2.1. Sinusoids are also known as *tones* or *pure tones* and result in simple auditory responses, producing a very "clean" sound [132]. Mathematically, a sinusoid $x(t)$ can be described as

$$x(t) = A\cos(2\pi ft + \phi).$$

Although sinusoids are very simple to analyze, they are not typically encountered in normal listening situations. Rather, the sounds we hear under normal listening conditions are much more complex and may not even be periodic. These complex waveforms can be "broken-down" into a series of sinusoids, each with its own frequency, amplitude and phase. A complex tone (periodic and non-periodic) can be described as the superposition of a number of sinusoids,

where the frequency of each sinusoid is an integral multiple of the *fundamental frequency*, the frequency of the lowest common "fundamental" component, that may not necessarily be present [132]. Frequencies other than the fundamental are known as the *harmonics*, where the first harmonic is the first multiple of the fundamental, the second harmonic is the second multiple of the fundamental and so on. For example, a square wave consists of a fundamental frequency sinusoid and the superposition of the odd harmonics of the fundamental (e.g., if the fundamental is 100Hz, the odd harmonics are 300Hz, 500Hz, 700Hz etc.). The amplitude of each harmonic is equal to the amplitude of the fundamental scaled by the inverse of the harmonic index.

2.1.1 Measuring Sound

As previously described, sound results from the variation of pressure arising when the molecules in the medium of propagation are compressed and expanded due to a vibrating object. Intensity is usually used to specify the magnitude of these variations (the compressions and expansions of the medium of propagation) and is defined as the sound energy transmitted each second through a unit area in a soundfield [132]. In the ideal scenario, sound source intensity is attenuated following the *inverse square law* $1 \cdot r^{-2}$ (e.g., by a factor of four for each doubling of distance the wave propagates). Although the inverse square law relates the intensity of sound waves to sound source dis-

tance, we perceive intensity as *loudness* [17, 201]. According to Moore [132], "loudness is defined as that attribute of auditory sensation in terms of which sounds can be ordered on a scale extending from quiet to loud". Loudness is a subjective measure and therefore cannot be measured directly. It may not always be an accurate representation of intensity [132] as the loudness of pure tone sounds is frequency and bandwidth dependent [62, 133, 152].

The range of intensity levels to which the human auditory system is sensitive is very large, and therefore, rather than giving direct intensity measures, a logarithmic scale is used instead. Given this logarithmic scale, the measures are referred to as *levels* (S_L) and are specified as a ratio with respect to some reference intensity measure ($I_\circ$) [132]

$$S_L = 10 \times log_{10}\left(\frac{I}{I_\circ}\right)$$

where, S_L is the sound level in Decibels (dB) corresponding to the ratio of intensities I and the reference intensity $I_\circ$. With a decibel scale, a 3dB increase in the intensity ratio corresponds to a doubling of the ratio of intensities.

Although the sound level ratio between two intensities can be determined, there may be times where a single measure of intensity is required. In such a situation, a standard reference intensity level is used. The standard reference level chosen is the threshold of human hearing for a 1000Hz tone and is equal to $10^{-12}\text{W}\cdot\text{m}^{-2}$ (Watts per square meter). Intensity levels given relative

to this particular reference level are known as *sound pressure levels* (SPL). As an example, a sound level of 3dB SPL represents an intensity twice that of the reference level, while a sound level of 0dB SPL represents an intensity equal to the reference level. Finally, intensity ratios can also be given as pressure ratios as well since there is a relationship between intensity and pressure (e.g., intensity is proportional to the square of pressure) [75]

$$\begin{aligned} S_L &= 10 \times log_{10}\left(\frac{I}{I_2}\right) \\ &= 10 \times log_{10}\left(\frac{P_1}{P_2}\right)^2 \\ &= 20 \times log_{10}\left(\frac{P_1}{P_2}\right) \end{aligned}$$

where, S_L is the level in Decibels and P_1 and P_2 are the two pressure measurements in Pascals (Pa). As with intensity levels, pressure levels can also be given relative to the standard measurement $P_\circ$, where the standard pressure measurement is 20μPa (e.g., $P_\circ = 20\mu$Pa) [75].

2.1.2 Near Field vs. Far Field

When describing the distance to a sound source, a distinction can be made between a sound source in the *near field* or in the *far field*. When the distance to the sound source is "very large", the sound source is in the far field and the

sound waves reaching a listener are *planar*. On the other hand, the sound waves reaching the listener from a sound source which is "very close" are not planar but rather spherical and therefore curved with respect to the listener's head.

The notion of the sound source distance being "very large" or "very small" relative to the listener can be ambiguous. Brungart and Rabinowitz [33], define the near field as "the region of space surrounding the listener within a fraction of a wavelength away from the sound source". Using this definition, the designation of a sound source in the near field or in the far field is frequency dependent given the inverse relationship between frequency and wavelength. However, for practical considerations, assuming propagation in the air, when the distance to a sound source is greater than approximately one meter, a far field sound source is assumed and the propagating waves are approximated by planar waves [33].

2.1.3 Coordinate Systems

When describing an auditory environment the position of a sound source and receiver must be given relative to some coordinate system. Various coordinate systems exist. In the *head-centered rectangular system* (Figure 2.2), the center of the head defines the origin and positions are specified in (x,y,z) coordinates. The positive x-axis (also known as the *interaural axis*) goes through the right

ear, the positive y-axis points directly in front of the head and the positive z-axis points directly upwards[1] (vertically). In this coordinate system, the axes form three planes. The y-z axis form the *median* (or *sagittal*) plane whereby any point on this plane is equidistant from the left and right ears. The x-y plane is known as the *horizontal* plane and is level with the listener's ears and finally, the x-z plane is referred to as the *frontal* plane.

Rather than specifying individual (x,y,z) axis components, a *spherical coordinate system*, in which coordinates are specified with an *azimuth*, *elevation* and *range*, may be used instead. In the "single pole" spherical system (Figure 2.3(a)), the center of the head defines the origin while azimuth (θ) and elevation (ϕ) are specified by lines of latitude and longitude respectively [37]. An azimuth angle of 0° is directly in front (e.g., median plane) while an angle of -90° is directly to the right (e.g., moving clockwise from 0° results in negative azimuth angles). The horizontal plane is at an elevation of 0° and moving upwards from this point, elevation increases positively, with $+90^\circ$ directly on top of the head. Range specifies the distance between the origin (center of the head) to the point of interest. The single pole system is intuitive and the most widely used coordinate system. However, it does have its problems. Most importantly, the length of an arc length (semi-circle) between two angles of azimuth is de-

[1]In a variant of this coordinate system, the role of the z and y-axis are reversed. In other words, the positive z-axis points directly in front of the head and the positive y-axis points directly upwards. This configuration is employed by the popular and widely used OpenGL computer graphics API [167] and is in fact the configuration used in this work although, the center of the listener's head does not define the origin.

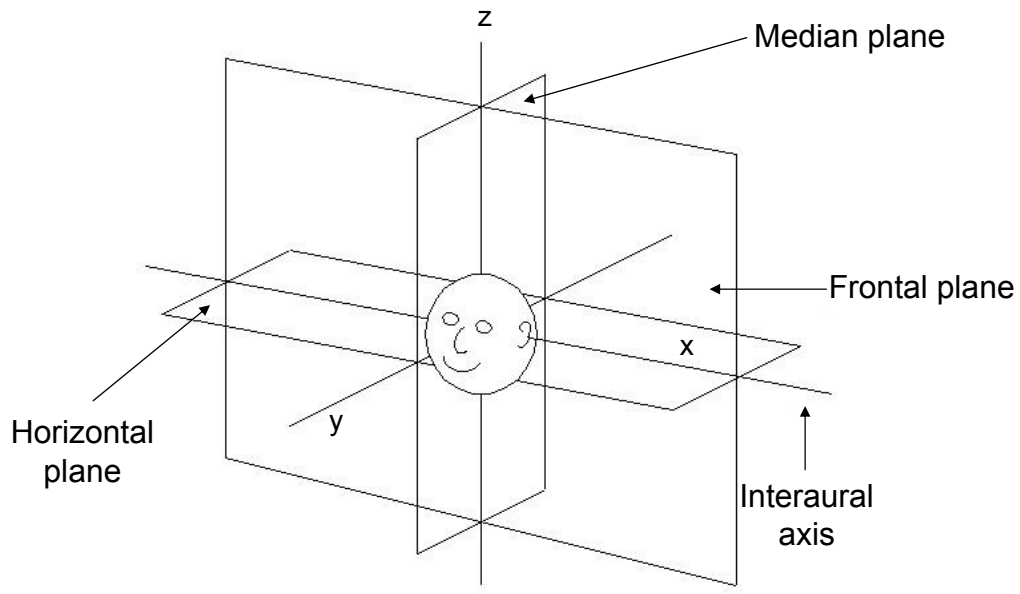
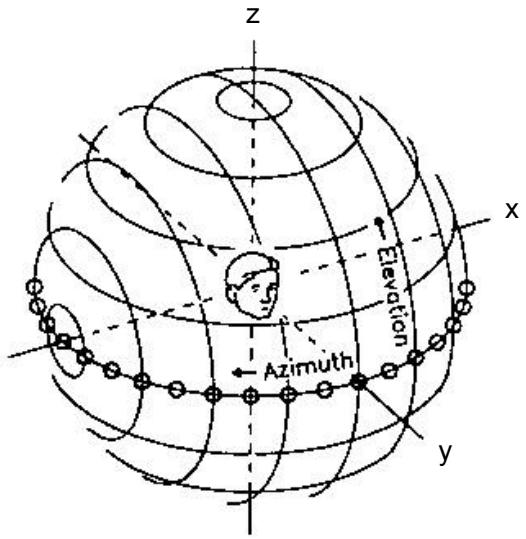


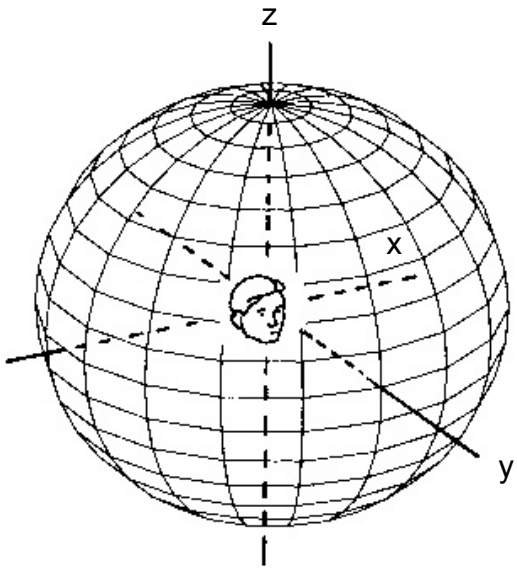
Figure 2.2: Coordinate systems: three planes of interest. The center of the head defines the origin, with the positive x-axis (also known as the *interaural axis*) going through the right ear, the positive y-axis pointing directly in front of the head and the positive z-axis pointing directly upwards (vertically). After [94].

pendent on elevation. For example, the arc length between 0° and 90° azimuth at an elevation of 0° is greater than the same arc at an elevation of 75°.

In the "double pole" system (Figure 2.3(b)), elevation is specified in the same way as in the single pole system however, azimuth is given as a series of rings which are parallel to the midline (the z-axis) and centered at the poles at each interaural axis [37]. In this system, the arc length between two angles of azimuth is independent of elevation.



(a) Single pole polar coordinate system.



(b) Double pole polar coordinate system.

Figure 2.3: Polar coordinate systems. (a) Single pole and (b) double pole. Reprinted from [37].

2.2 Room Acoustics and Sound Propagation

Various factors affect a propagating sound wave before it reaches the listener (receiver), including the medium itself. When the medium is air, the characteristics of the air (e.g., humidity level, temperature, etc.) have an effect on the propagating wave. As a wave propagates, a portion of it is absorbed by the air, modifying its sound spectrum. The amount of absorption is affected by the properties of the air itself, including its temperature and humidity level [74]. Absorption of sound waves is also a function of frequency. Higher frequency components are absorbed more readily than the lower frequency components.

In addition to the absorption by the medium (air), typical listening environments are *echoic*, as opposed to *anechoic*. In an ideal anechoic environment there are no reflections and the listener will only receive sound that propagates on a direct path from the sound source. Anechoic settings rarely occur in nature although a large open space or the top of a mountain summit does approach anechoic [17]. Anechoic chambers are artificially created environments where the walls, floor and ceiling are covered with sound absorbing material to prevent reflections of sound waves that would occur when a sound wave encounters a surface.

In contrast, in an *echoic* environment reflections result when the sound waves encounter obstacles/objects in the environment (e.g., walls of a room) on

the path from the sound source to the listener. Sound waves reach the listener both directly, via the straight line path between the sound source and receiver (assuming such a path is not occluded) and indirectly after being reflected off of the walls, floor, ceiling or any other obstacles and obstructions the waves may encounter. Sound waves may also reach the listener after being diffracted off of objects or after being refracted. The sound reaching the listener varies as a function of the geometry of the room relative to the listener [37], as well as the material of the room and the source spectrum (e.g., frequency components) and is rather irregular [65]. Collectively, the sound reaching the listener via a direct path, specular or diffuse reflections, diffraction or refraction over a period of time is described by the room impulse response (RIR) which provides a measure of the sound energy reaching the listener over time. Greater details regarding the nature of the indirect sound reaching a lister is provided in the following sections.

2.2.1 Reverberation: Specular and Diffuse Reflections

The collection of sound waves reaching the receiver over time either directly or indirectly, which may consist of several thousand waves reflecting specularly or diffusely from the various surfaces within a space, is known as *reverberation* [40]. The number of times a wave is reflected (either specularly or diffusely) is denoted by its *order* (e.g., a reflection of order n indicates the wave has been re-

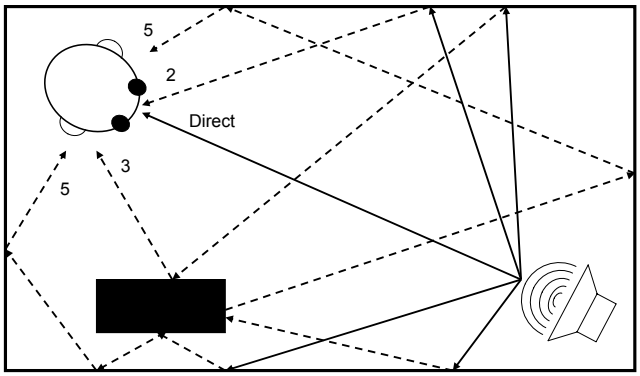


Figure 2.4: Direct and reflected sound waves reaching a listener. In addition to the sound waves reaching a listener via a "straight line path" directly from the sound source, reflected sound will also reach a listener. The number of times a wave is reflected before reaching the listener is known as its *order*. The wave order in this example is provided next to each ray reaching the receiver. The direct sound has an order of zero.

flected n times), where the order of the direct sound is zero. In many situations, a higher reflection order indicates a reduction in the intensity level due to the absorption of a portion of the wave's energy by the reflecting surfaces and the air [168]. An example illustrating the order of reflected waves is provided in Figure 2.4.

In addition to reflection order, reflections can be broken down into two categories: *early* and *late* reflections. Reflections arriving from the room boundaries (e.g., walls, floor and ceiling), are known as early reflections and typically arrive within 80ms of the direct sound. Reflections arriving after 80ms and with reflection orders greater than one are known as late reflections. Late reflections, arising from "reflected reflections" from one surface to another, are assumed to arrive equally from all directions and are described statistically as exponential decaying noise [65]. A graphical illustration of these concepts is provided in

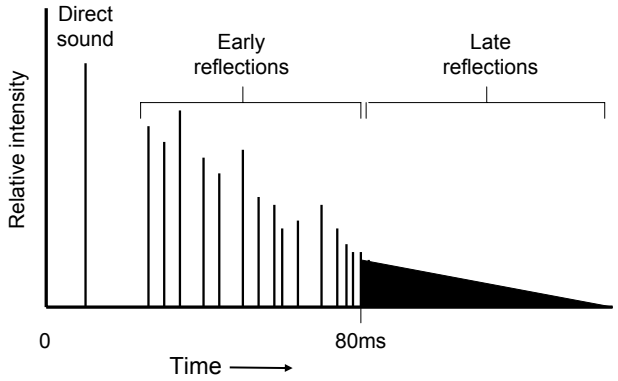


Figure 2.5: Theoretical echogram (distribution of sound energy arriving at a receiver over a period of time), consisting of the direct sound, early and late reflections. In addition to the sound waves reaching a listener by traveling from the sound source to the listener directly, indirect sound waves reflected from the walls, floor or other objects in the environment will also reach the listener, albeit after the arrival of the direct sound. Early reflections arrive within 80ms of the arrival of the direct sound. Late reflections arrive after 80ms of the direct sound and are typically described as exponentially decreasing noise. The echogram can be converted to a corresponding impulse response through a post-processing process.

Figure 2.5 where, the distribution of sound energy arriving at a receiver over a period of time (the echogram) is shown. An actual room impulse response measured from in a classroom at the right ear of a receiver is illustrated in Figure 2.6.

Other parameters used to describe reverberation include *reverberation time* and *reverberation distance*. Reverberation time (RT_{60}) is defined as the time required for the sound pressure level (SPL) to be attenuated by 60dB (e.g., by a factor of one million), independent of the intensity of the sound after a steady state sound is turned off. Reverberation time can be approximated by [65]

$$T_{60} \approx \frac{V}{6 \times \beta \times S} \tag{2.1}$$

where, V is the volume of the room (in m^3), β is the (frequency dependent)

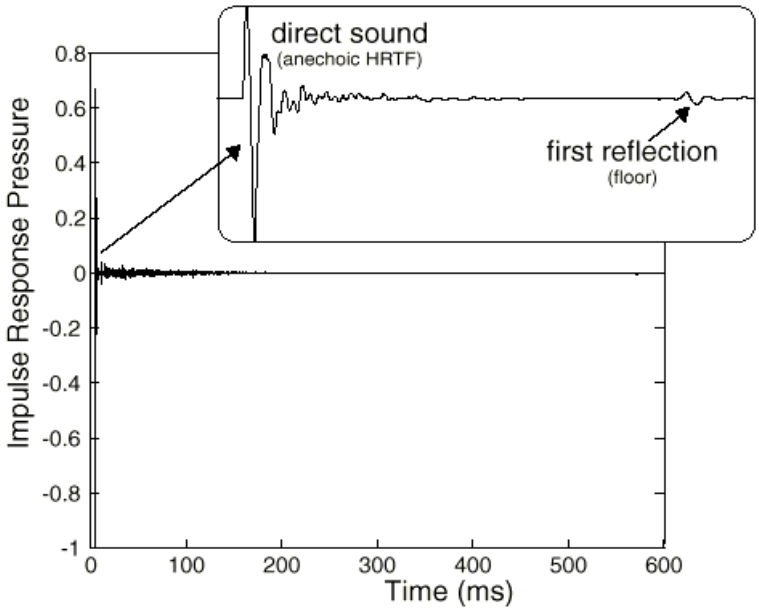


Figure 2.6: Actual impulse response measured in a classroom at the right ear of a receiver. The sound source was positioned at 45° azimuth and 0° elevation relative to the listener. Reprinted from [164].

average *surface absorption coefficient* of the room boundaries and S is the sum of the surface areas of the room in m^2.

When a sound wave in a room strikes a surface, the fraction of sound lost by the surface (either absorbed by the surface or transmitted through the surface), is defined by the surface absorption coefficient which can take values between 0 and 1 (1 being a perfect absorber and 0 a perfect reflector). The surface absorption coefficient indicates the fraction of incoming energy that is absorbed by the surface itself. Since the absorption coefficient of a surface typically changes with frequency, reverberation time is frequency dependent. As given, reverberation time is rather arbitrary and depends on the characteristics of the enclosure, including the material of the walls, floor and ceiling, number and type of objects in the room etc. Depending on the level of the background noise, it may be the case that reflections arriving after T_{60} are still considerably audible [17]. However, the choice of 60dB was made by considering a good "music making area", such as a concert hall. The loudest level reached for most orchestral music is typically 100dB (SPL), while the level of background noise is approximately 40dB. As a result, a reverberation time of 60dB is the time required for the loudest sounds of an orchestra to be reduced to the level of the background noise. Reverberation time is highly affected by the reflective surfaces encountered by the propagating waves. When a surface is highly reflective, very little energy is absorbed by the surface (e.g., the reflected

wave maintains most of its energy) leading to an increase in the reverberation time. In contrast, highly absorbing materials absorb much of the energy of a wave striking it, greatly reducing the energy in the reflected portion thereby reducing the reverberation time.

Late reflections arrive any time after 80ms of the direct sound. However, as the distance between the sound source and listener increases, the level of the direct sound L_{direct} decreases until the level of the direct sound equals the level of the reverberation L_{reverb}. Reverberation distance r_{reverb} is defined as the distance such that $L_{direct} = L_{reverb}$ and is given by [65]

$$\begin{aligned} r_{reverb} &= 0.25 \times \sqrt{\frac{\beta \times S}{\pi}} \\ &= 0.006 \times \sqrt{\frac{V}{T_{60}}}. \end{aligned}$$

2.2.1.1 Specular Reflection

As illustrated in Figure 2.7, and similarly to light waves encountering a mirror surface, an ideal specular reflection occurs when all of the reflected energy is reflected in one direction only. In such a situation, with respect to the surface normal vector, the angle of incidence is equal to the angle reflection. The roughness of many "large" smooth surfaces, such as glass, concrete, plastic or wood, that are encountered frequently in indoor environments are "smooth"

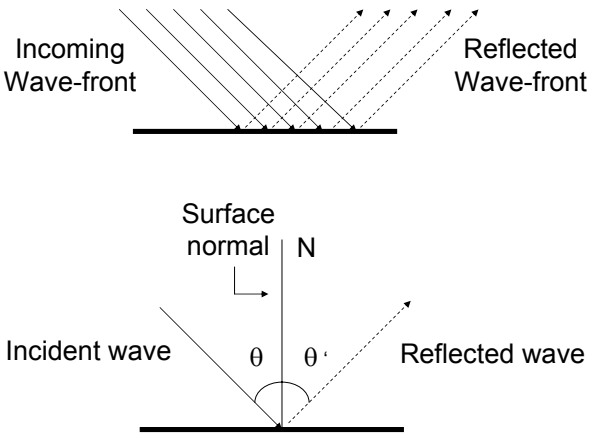


Figure 2.7: Specular reflection.

with respect to many of the sounds incident upon them, resulting in specular reflections, especially when considering lower frequency (longer wavelength) sounds [156]. The smoothness of a surface is defined relative to the wavelength of sound. The larger the wavelength of sound relative to the roughness of the surface, the greater the surface smoothness.

2.2.1.2 Diffuse Reflection

When the roughness of the surface is of the same order of magnitude as the incident sound, the incident sound is reflected in (almost) all directions as opposed to a single direction only (Figure 2.8). This "scattering" or "dispersion" of sound in all directions is referred to as diffuse reflection or *diffusion*. Surfaces that reflect incident (sound) waves diffusely are known as diffuse surfaces [51, 156]. Such surfaces include coffered ceilings, faceted walls, raw brick walls and the audience area of a concert hall [51].

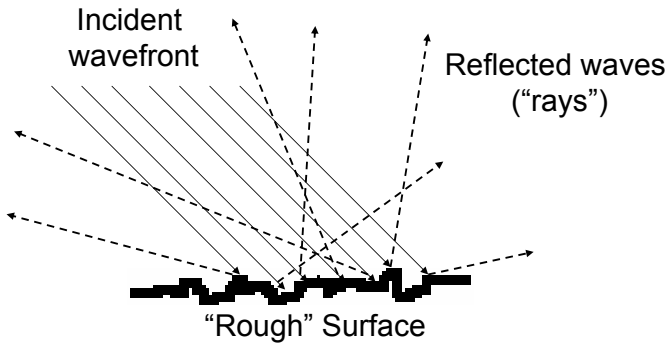


Figure 2.8: Diffuse reflection.

2.2.2 Diffraction

Diffraction refers to the "bending mode" of sound propagation whereby sound waves go ("bend") around an obstacle that lies directly in the line of straight propagation [48], allowing us to hear sounds around corners and around barriers (see Figure 2.9). Diffraction is dependent on both wavelength and obstacle/surface size, increasing as the ratio between wavelength and obstacle size is increased [48]. In other words, diffraction will typically be greater for lower frequency sounds and when the obstacles are small. The frequency spectrum of audible sound ranges from approximately 20Hz to 20kHz, corresponding to wavelengths ranging from 0.02m to 17m (with a velocity of $v_c = 343\text{m}\cdot\text{s}^{-1}$ [48] for sound in air and a frequency of f Hz, wavelength $\lambda = v_c \cdot f$). Since, the dimensions of many of the objects/surfaces encountered in our daily lives is within the same order of magnitude as the wavelength of audible sounds, diffraction is an elementary means of sound propagation especially when there

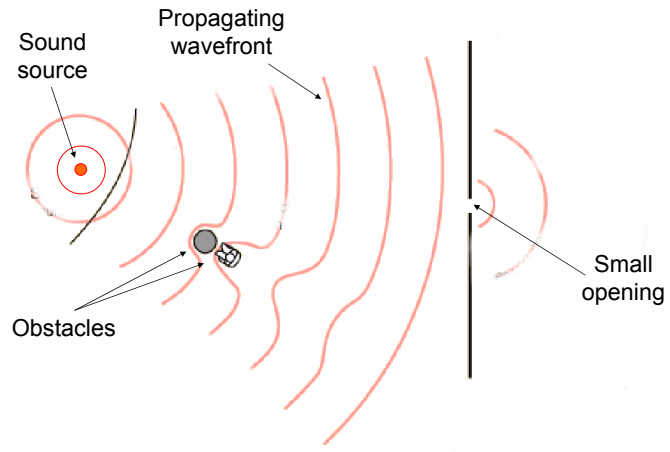


Figure 2.9: Diffraction.

is no direct path between the sound source and the receiver, such as in buildings [180]. Failure to account for diffraction can lead to a non-realistic auditory simulation. That being said, diffraction effects are typically ignored by spatial sound systems [178, 180].

2.2.3 Refraction

The changes in the direction of propagation of a sound wave as the wave encounters changes in the medium is referred to as refraction. The two media will have different characteristics and therefore result in different wave propagation speeds (see Figure 2.10). Examples of refraction include sound waves traveling in air encountering a large body of water or sound waves traveling through one particular region of air encountering another region of air with differing characteristics (e.g., temperature). Although refraction occurs, it is not as common as the other reflection types when considering room acoustics

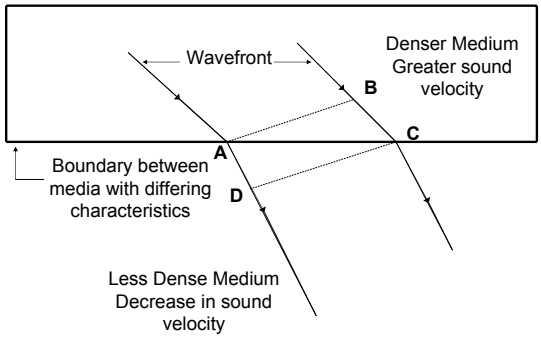


Figure 2.10: Refraction.

since even regions in the medium with differing temperatures will eventually inter-mix into a single homogeneous region [48]. Refraction is therefore typically be ignored by room acoustics applications and is ignored in this work.

2.3 Human Sound Localization

In humans, the perception of sound begins with the arrival of varying sound pressure at the ear drums. The range of frequencies to which humans are sensitive (e.g., can hear) is restricted to the range of 20Hz to 20kHz for a young healthy adult [37]. Frequencies below 20Hz are known as *subsonic* and can at times be felt rather than heard, while frequencies above 20kHz are known as supersonic [17]. Through the actions of the eardrum these oscillating mechanical variations of air pressure are passed through to the middle ear and converted (transduced) into electrical signals in the inner ear and ultimately coded into a pattern of neuronal spikes which are interpreted by the brain (a complete discussion of the physiology of the ear is beyond the scope of this

dissertation - see [25, 132] for greater detail).

2.3.1 Duplex Theory

The *duplex theory*, one of the earliest theories of human sound localization, was developed by Lord Raleigh in 1907 [172] and is premised on the observation that the two ears are separated by a featureless spherical head. Under the assumption of a perfectly spherical head with no external ears (pinnae), this theory explains many of the properties of human sound localization. Unless the sound source lies on the median plane (i.e., the plane where the points are equidistant from the left and right ears), the distance traveled by sound waves emanating from a sound source to the listener's left and right ears differs. This causes the sound to reach the *ipsilateral* ear (the ear closest to the sound source) prior to reaching the *contralateral* ear (the ear farthest from the sound source). The *interaural time delay* (ITD) is the difference between the onset of non-continuous (transient) sounds, or the phase of more continuous sounds, at the two ears. Similarly, since the ears are separated by the head, when the wavelengths of the sound waves are short relative to the size of the head, the head acts as an acoustical shadow attenuating the sound pressure level of the sound waves reaching the contralateral ear [196]. This difference in sound pressure level (SPL) between the waves reaching the ipsilateral and

contralateral ears is referred to as *interaural level difference* (ILD[2]).

ITDs are more prevalent primarily for low frequencies (< 1500Hz) [26], where the arriving sound wavelengths are long relative to the diameter of the head and the phase of the sounds reaching the ears is determinable. Low frequency sounds corresponding to wavelengths greater than the diameter of the head, experience diffraction whereby they "bend around" the head to reach the contralateral ear. As a result, ILD cues for low frequency sounds are very minuscule (although in some cases, they may be as large as 5dB [196]). However, for frequencies greater than approximately 1500Hz, where the size of the head is greater than the wavelength, the sound waves are too small to bend around the head, and are therefore shadowed by the head.

Although the duplex theory does explain sound localization on the horizontal plane, it is incomplete as it fails to account for many aspects of human auditory localization (see [37]). Human beings are capable of localizing a sound source even if the listener's hearing is impaired in one ear [171]. Furthermore, the duplex theory cannot differentiate between sound sources on the median plane directly in front of or in back of an individual (this is known as *front-back-confusion*) since both ITD and ILD in either of these cases is (ideally) zero. In fact, a sound source positioned anywhere on the surface of a cone centered on the interaural axis (known as the *cone of confusion*, see Figure 2.11), will have

[2]Also known as interaural intensity difference (IID).

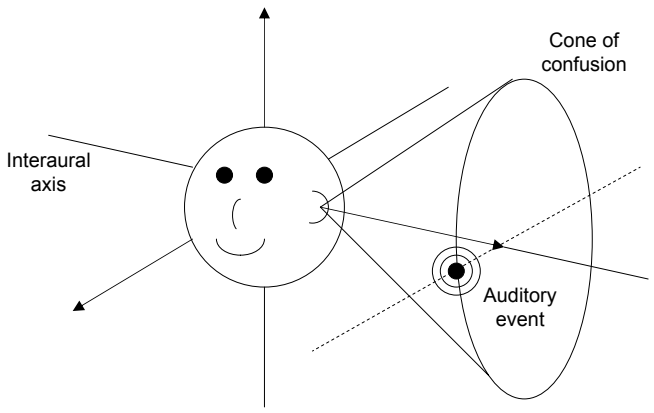


Figure 2.11: Cone of confusion. A sound source positioned anywhere on the surface of the cone will produce an identical ITD value. After [94].

identical ITD values [26].

Batteau's work in the 1960s on the filtering effects introduced by the pinna [10] was the next major advance in the study of human auditory localization. In fact, in addition to the filtering effects introduced by the pinna, the listener's head shoulders and upper torso will, perhaps to a lesser degree, also modify the sound reaching the listener's left and right ears. Collectively, these filtering effects are described by a complex response function known as the *head-related transfer function* (HRTF) or the anatomical transfer function (ATF) [77] and encompass various sound localization cues including ITDs, ILDs, and the changes in the spectral shape of the sound reaching a listener. HRTF-based sound localization models overcome some of the localization limitations inherent with models based on the use of ITD and ILD cues alone.

2.3.2 Head Related Transfer Function

Each person's head and upper torso, in addition to the physical structure of the pinna that consists of a series of asymmetric grooves and notches, modulate the mid and high frequency energy content of the sound spectrum. The HRTF modifies the spectrum and timing of a sound signal reaching the ears in a location-dependent-manner [17]. Studies have shown that the number of front-back confusions increases and localization accuracy decreases as the bandwidth of the source is decreased [26, 37, 127]. Furthermore, when a portion of the outer ear is occluded there is an increase in the number of front-back confusions [61, 140].

Following Zotkin et al. [202], the left and right ear HRTFs (H_L and H_R respectively) are defined as the ratio between the sound pressure level (SPL) present at the eardrum of the left and right ears, $\Phi_L(\omega,\theta,\phi,d)$ and $\Phi_R(\omega,\theta,\phi,d)$ respectively, and the *free field* SPL at the position corresponding to the center of the head but with the head absent ($\Phi_f(\omega,d)$)

$$H_L = \frac{\Phi_L(\omega,\theta,\phi,d)}{\Phi_f(\omega)}, \qquad H_R = \frac{\Phi_R(\omega,\theta,\phi,d)}{\Phi_f(\omega)}. \tag{2.2}$$

Here, ω is the angular frequency, θ and ϕ are the azimuth and elevation angles respectively, and d is the distance from the listener to the sound source (measured from the center of the listener's head). When considering a sound

source in the near field (e.g., sound source distance less than approximately one meter), displaced from the median plane, HRTFs (and in particular the ILD component of the HRTFs) are, in addition to direction, dependent on distance [32, 165]. In fact, as demonstrated by Brungart and Rabinowitz [32], the ILD associated with a near-field sound source can vary by as much as 30dB (see Duda and Martens [57] for greater details on the range dependence of HRTFs).

Far-field HRTFs are typically assumed to be independent of distance, despite the fact that the frequency spectrum of a sound source varies with source distance due to absorption effects by the medium [136]. This high frequency attenuation is of particular importance for larger distances (greater than approximately 15m), but provides little information for smaller distances [31].

The pinna of individuals differ; varying widely with respect to size, shape and general make-up, leading to variations to the resulting filtering of the sound source spectrum amongst individuals, particularly at higher frequencies. Higher frequencies (e.g., $>$ 10kHz) are more attenuated when the sound source is to the rear of the listener, as opposed to the front of the listener and in the $5-10$kHz frequency range. When considering higher frequencies, the HRTFs of individuals can differ by as much as 28dB irrespective of the individual's characteristics [195]. This high frequency filtering is an important cue to source elevation perception and in resolving front-back ambiguities [127, 191].

Example left and right ear HRTFs from three individuals, measured by

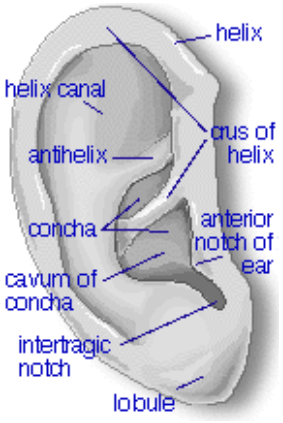
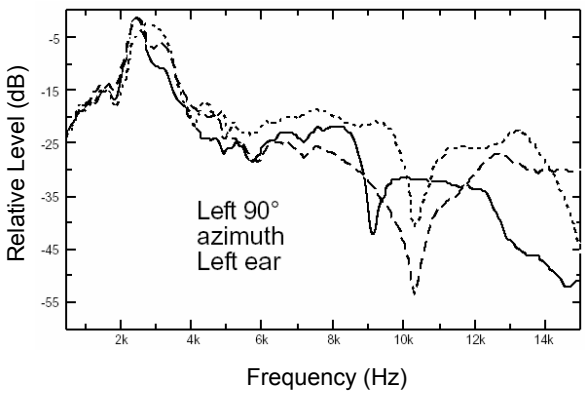


Figure 2.12: Human pinna.

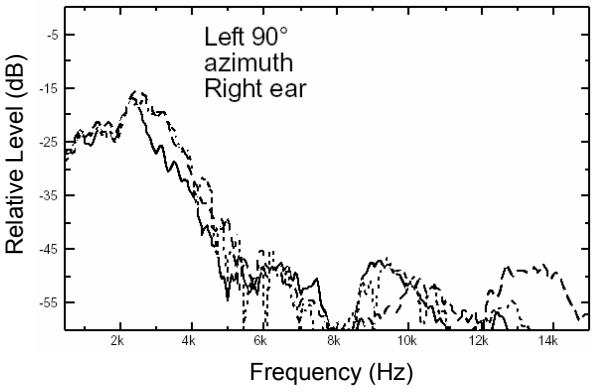
Fred Wightman and Doris Kistler [17] are shown in Figure 2.13 for a sound source located at $\theta = 90^\circ$ and $\phi = 0^\circ$ (e.g., directly to the left of the listener). The inter-subject differences in each plot are clearly evident (the HRTF for each individual is denoted by different line styles). A decrease in level of the right ear HRTF due to the shadowing of the head is also evident.

2.3.3 Other Factors Affecting Human Sound Localization

The sound localization cues described in the previous section arise due to our own anthropomorphic make-up (e.g., two ears separated by the head and the grooves and notches of our pinna). In addition to the sound localization cues arising due to our own physical make-up, various other "external" factors alter the sound reaching a listener providing further information to the location of a



(a) Left ear.



(b) Right ear.

Figure 2.13: Example HRTFs of three individuals for a sound source at $\theta = 90^\circ$ and $\phi = 0^\circ$. (a) Left ear and (b) right ear. These HRTFs were originally measured by Fred Wightman and Doris Kistler and are reprinted from [17].

sound source.

2.3.3.1 Reverberation

Reverberation is a useful cue to sound localization capable of providing sound source distance estimation and information with respect to the physical "make-up" of a room (e.g., size, types of materials on the walls, floor, ceiling etc.). Reverberation can be used to provide absolute sound source distance estimation independent of overall sound source intensity [37, 165], due to the variation of the direct-to-reverberant sound energy level as a function of source distance [17, 18, 31, 40, 47, 137, 165]. As the distance to a sound source increases, the ratio between the direct-to-reverberant levels

$$L_{ratio} = \frac{L_{direct}}{L_{reverb}}$$

decreases. When the direct distance to the sound source r_{direct}, is less than the reverberant distance r_{reverb} the intensity of the direct sound is greater than that of the reverberant sound (e.g., $L_{direct} > L_{reverb} \Rightarrow L_{ratio} > 1$). In contrast, when the reverberant distance is greater than the direct sound source distance ($r_{reverb} > r_{direct}$) the level of the reverberant sound dominates (e.g., $L_{reverb} > L_{direct} \Rightarrow L_{ratio} < 1$). Mershon has performed various studies examining the effect of the ratio of the direct-to-reverberant intensity with respect to sound source distance estimation [123, 124, 125]. These studies suggest that distance judgments are more accurate in the presence of reverberation than in an

anechoic environment and as previously described, that the ratio of direct-to-reverberant intensity is a cue to absolute sound source distance judgment.

Reverberation may be altered drastically with small changes to the objects in the environment themselves, changes in their positions, changes to the medium that the sound is propagating in (typically air) or with the introduction of new objects in the environment. Although in general the ratio between the direct and reverberant sound decreases/increases as the sound source distance is increased/decreased, this may not necessarily always be true. Furthermore, although evidence indicates that reverberation does provide a cue to absolute sound source distance, studies also indicate reverberation can have negative affects as well. In particular, it leads to a decrease in directional localization accuracy in both real and virtual environments [17], and although this effect is of small magnitude, it is nevertheless measurable [166].

Finally, in a reverberant environment, sound source distance is also affected by the background noise [123]. In the presence of background noise, we tend to underestimate the distance to a sound source. This is probably due to the fact that since noise masks part of the weaker indirect portion of the sound reaching the listener, we cannot detect the entire extent of the reverberation [136].

Precedence Effect: In a typical listening situation, the listener receives sound emitted by the sound source directly in addition to delayed and attenuated ver-

sions resulting from the reflection of the sound with objects in the environment. The reflected sounds reaching the listener may emanate from any direction in the environment, potentially creating a false impression of a sound source at the location of reflection. The ability of the auditory system to combine both the direct as well as reflected sounds such that they are heard as a single entity and localized in the direction corresponding to the direct sound has been termed the *precedence effect* by Wallach et al. [188] (it is also known as the *Haas effect* and the *law of the first wavefront*). The precedence effect allows one to localize a sound source in the presence of reverberation, even when the energy of the reverberant sound is greater than that of the direct sound [76, 132].

2.3.3.2 Head Movement

In any normal listening environment, we are not stationary but are rather free to move about. Head movements are a very important and natural component of sound localization and can greatly reduce front-back confusions and improve sound localization accuracy [177, 187, 197]. Head movements result in a change of position between the sound source and the listener, leading to changes in the ITD and ILD cues and to the sound spectrum reaching the ears (see Figure 2.14 for a graphical illustration). We are capable of integrating these changes as they occur over time in order to resolve ambiguous situations such as front-back confusions [16].

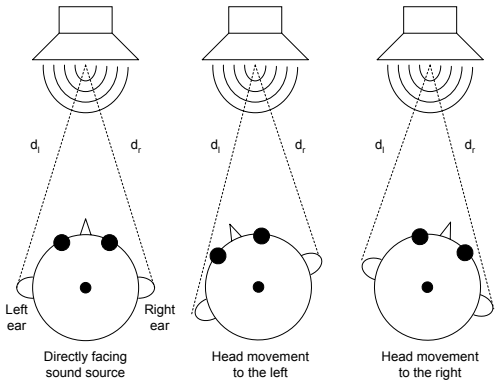


Figure 2.14: Head movements to resolve front-back ambiguities. When the sound source is directly in front of the listener, the distance between the left and right ears (d_l and d_r respectively), is the same. A head movement to the left will increase the distance between the left ear and the sound source d_l, while a head movement to the right will increase the distance between the right ear and the sound source d_r.

2.3.3.3 Sound Source Familiarity

A listener's prior experience with a particular sound source and the environment (e.g., the sound source transmission path) can greatly affect a listener's ability to localize a sound source. Prior information about a sound source or environment allows a listener to use their previous experiences and knowledge to provide a more accurate localization estimate or to overcome ambiguous situations. For example, from a very young age we engage in conversations with others. For normal listeners, speech has become an important aspect of life as it allows us to communicate with others and express our thoughts. As a result, we have become familiar with the characteristics of speech (e.g., how loud a whisper or shouting may be and who is speaking) and are capable of accurately judging the distance to a live talker under normal conditions, especially when the distances are within a few feet [34, 66].

2.4 Summary

Sound waves are a *mechanical wave* phenomenon that result from the variation of pressure arising when the molecules in the medium of propagation (typically air) are compressed and expanded due to a vibrating object. Various factors affect a propagating sound wave before it reaches the listener. As a wave propagates, a portion of its energy is absorbed by the air, modifying the sound spectrum in some manner. Furthermore, sound waves may encounter any number of surfaces/objects as they propagates through the environment and may be reflected either specularly of diffusely and they may be diffracted or refracted, further altering the sound spectrum. Collectively, the sound reaching the listener via a direct path, via specular or diffuse reflection, diffraction or refraction over a period of time is described by the room impulse response which provides a measure of the sound energy reaching the listener over time. In humans, the perception of sound begins with the arrival of varying sound pressure at the ear drums. Through a variety of sound localization cues that arise from our own anthropomorphic make-up (most notably interaural time and level differences and head-related transfer functions), in addition to localization cues that arise from various other "external" factors such as reverberation and sound source familiarity, humans are capable of accurately determining the location of a sound source in the presence of complex effects.

Chapter 3

Acoustical Modeling

Although computational acoustical modeling is the primary focus of this chapter, this chapter begins with a brief introduction to estimating the room impulse response using methods where the response is physically measured as opposed to being computationally modeled. This chapter ends with a review of the photon mapping method. Photon mapping forms the basis of the sonel mapping method. It should be noted that although focus of the chapter is on estimating the room impulse response, in sonel mapping it is the echogram (the distribution of acoustical energy over time) that is actually estimated. The equivalent room impulse response can be obtained by post-processing the echogram (see [59, 101]). For the reminder of this chapter, although mention is made to estimating the room impulse response, unless specified otherwise, it is the echogram that is actually estimated.

3.1 Physically Measuring the Room Impulse Response

The binaural room impulse response (BRIR) represents the response of a particular acoustical environment to sound energy and captures the room acoustics for a particular sound source and listener configuration. For a real environment, the BRIR can be measured by generating an impulsive sound with known characteristics through a loudspeaker positioned somewhere in the room and measuring the response of the arriving sound (with probe microphones), at the ears of an observer (either an actual human listener or an anthropomorphic "dummy head" positioned in the room). The recorded response can then form the basis of a filter that can be used to modulate source sound material (anechoic or synthesized sound) and when this filtered sound is presented to a listener, the reproduced sound will include the direct sound, the effects of the reflected sound in addition to directional filtering effects introduced by the HRTF [120]. In other words, the original sound environment and listener configuration is reproduced at the listener. Physically measuring the BRIR is very restrictive. The measured response is specific to the room configuration in which it was measured with the original sound source and listener positions hence, only that particular room and sound source/receiver configuration are "re-created". Movement of either the sound source or the receiver or changes to the room itself (e.g., introduction of new objects or movement of existing objects in the room), requires the room impulse response to be re-measured. Such an

approach is impractical for all but a few limited scenarios.

The BRIR can be considered as the signature of the room response for a particular sound source and receiver. Although interlinked and their interaction is very complex, for simplicity and reasons of practicality, the room impulse response and the response of the human receiver are commonly determined separately and combined via a post-processing operation to provide an approximation to the actual BRIR [95]. The response of the room is known as the *room impulse response* (RIR) and captures the reflection properties (reverberation), diffraction, refraction, sound attenuation and absorption properties of a particular room configuration (e.g., the environmental context of a listening room or the "room acoustics"). The response of the human receiver captures the direction dependent effects introduced by the listener due to the listener's physical make-up (e.g., pinna, head, neck, shoulders and torso) and is known as the *head related transfer function* (HRTF).

In a manner similar to measuring the BRIR, the RIR of a particular environment can be measured by generating a sound with known spectral characteristics from a source at a known position and measuring the response at a particular location with a microphone. Similarly, a listener or dummy head's left and right ear HRTF can also be measured by outputting a sound from a loudspeaker placed at known positions relative to the "head" and measuring the resulting response at the left and right ears using small probe microphones

inserted in the head's left and right ear canals [17]. As previously described, for a sound source in the near-field, HRTFs are, in addition to direction, dependent on distance and therefore, distance cannot be ignored [33]. Accounting for distance complicates both the task of HRTF measurement and the computational requirements of any auralization system. As a result, a far-field sound source is typically assumed (e.g., sound source distance is greater than approximately one meter) and distance is thus often ignored. Martens [106, 107] describes a binaural display that does account for sound source distance in simulated HRTFs at close range. Greater details regarding HRTF-based binaural synthesis techniques are provided by Martens [111].

Once both the RIR and HRTF have been measured/approximated, they can be combined via post-processing operation to provide an approximation to the actual BRIR. This post-processing operation is a non-trivial and a potentially computationally intensive task at the minimum requiring the following information [95]:

1. The calibrated band level of each reflection as a function of frequency.
2. The incidence angles of each reflection.
3. The data for late reverberation generation as a function of frequency.

Each reflection has its own direction of arrival and given the direction (position) dependent filtering introduced by the HRTF, each reflection should be filtered with an HRTF pair corresponding to that direction [17] as done in a

variety of auralization systems including NASA's SLAB [192, 193]. To limit the computational complexity and for simplicity reasons, rather than filtering each reflection with its corresponding HRTF, the HRTF can been used to filter the direct sound only. This is not completely correct and leads to some degree of perceptual degradation. For example, in the real world, the first reflection usually comes from sound reflected off of the floor. This should therefore be the case in any virtual simulation/environment. However, when the HRTF is used to process the direct sound only, such information may be lost [149].

3.1.1 Acoustical Scale Modeling

Rather than measuring the RIR in the actual room of interest, *acoustical scale modeling* can be used instead. In the acoustical scale modeling technique a three dimensional scaled down actual material model of the particular environment is built and used to measure the acoustical properties of the real environment and construct the RIR. This allows for the correct inclusion of all environmental effects including scattering and diffraction of the sound waves as they encounter surfaces in the environment provided the scaled model is an exact scaled version of the room being modeled (something impossible to achieve in reality [95]). Given the limitations inherent with physically measuring a set of RIRs for a particular environment (actual environment or a scaled version), as described in the following section, the RIR is often approximated

using *computational acoustical modeling approaches* instead.

3.2 Computational Acoustical Modeling

The two major approaches to computational acoustical modeling (e.g., estimating the room impulse response), are [64]: i) *wave-based modeling* whereby numerical solutions to the wave equation are sought and ii) *geometric modeling* whereby sound is approximated as a ray phenomenon and traced through the scene. Geometric modeling, and in particular ray-based approaches, such as ray-tracing, are the most widely used due to various factors, including simplicity and computational feasibility. Greater details regarding both approaches are provided below.

3.2.1 Wave-Based Methods

Here the objective is to solve the *wave equation*, also known as the *Helmholtz-Kirchoff* equation [179], to completely recreate a particular sound field. An analytical solution to the wave equation is rarely feasible and wave-based methods using numerical approximation, such as finite element methods (FEMs), boundary element methods (BEMs) and finite difference time domain methods (FDTDMs) are used instead [160]. These methods solve the complex integral and differential equations representing the wave equation by sub-dividing the domain of these complex functions into smaller units such that for each smaller

unit, the function can then be approximated using simpler functions that can be easily solved [14, 54]. Numerical approximations such as FEMs and BEMs essentially project the original complex function into a *finite function* space where the approximated function is characterized by a finite number of unknowns that can then be solved numerically [46].

One of the first numerical approximations was presented by Miles [128] who determined the sound field in a three-dimensional rectangular enclosure with non-uniform absorbent and diffusely reflecting walls. Shi et al. [163] describe an acoustical radiosity-based approach that combines both auditory and visual information to model both the acoustical and visual scene. Their acoustical radiosity method is obtained by modifying the original (light-based) radiosity method to account for the finite propagation speed of sound and attenuation of sound by air. Essentially, the exchange of energy between surfaces is calculated based on a form factor [44] that describes the fraction of energy leaving one surface that arrives at another. More recent acoustical radiosity approaches include the work of Nosal et al. [138] as well as a study by Le Bot and Bocquillet [119] where a comparison between radiosity based approaches and ray tracing methods was made. In addition, although not directly intended to model sound propagation, Rougeron et al. [156], describe a time-dependent radiosity method to simulate the indoor propagation of a 60GHz electromagnetic wave. Their technique can easily be applied to the modeling of sound propagation as well.

Although wave-based methods can account for non-specular reflection phenomena, they are computationally very expensive making them impractical for all but very simple, static environments. Furthermore, their computational complexity increases linearly with the volume of the room and the number of volume elements and is proportional to the fourth power of the frequency of interest [146]. Rabenstein et al. estimate that 4.2 Gigaflops (4.2 billion floating point operations each second) are required to simulate the propagation of a 3kHz sound in a $100m^3$ room using wave-based approximation techniques [146].

3.2.2 Geometric Acoustics - Ray-Based Methods

In a manner similar to geometric optics, with ray-based acoustical modeling it is assumed that sound acts as rays and ray-based computer graphics (image synthesis) rendering-like techniques are used to model the acoustics of an environment. Essentially, the propagation paths taken by the acoustical energy are found by tracing (following) these "acoustical rays". Tracing a ray involves following the ray while it propagates from the sound source and interacts with any number of objects/surfaces in the environment before reaching the receiver (listener). Mathematical models are used to account for source emission patterns, atmospheric scattering, absorption of the sound ray energy by the medium itself and interactions with any surfaces/objects a ray may en-

counter [64]. The receiver is typically modeled as a sphere of some radius r and the *echogram*, describing the distribution of incident sound energy (rays) reaching the receiver over time is calculated. Several of the more popular ray-based methods include *image sources* [2], *ray tracing* [97, 99], *beam tracing* [49, 63, 64, 80]. More recent techniques include sonel mapping [89, 90, 91, 92] and phonon tracing [23], both of which are inspired by photon mapping.

Although many different geometric acoustical modeling systems are currently available, given the potentially complex task associated with modeling the acoustics of all but the simplest environments, many of the existing systems have their limitations. They typically assume that all interactions between a sound ray (wave) and objects/surfaces in the environment are specular, despite that in our natural settings other phenomena influence a sound wave while it propagates through the environment as well (e.g., diffuse reflections, diffraction and refraction). These methods typically ignore the wavelength of sound and any phenomena associated with it (e.g., diffraction) [100] and only accurately model the early portion of the room impulse response. Furthermore, acoustical ray-based simulations become rather complicated for all but very simple, theoretically ideal cases [95]. Finally, ray-based methods are not particularly suited for dynamic environments where the sound source and receiver are free to move about freely.

Ray-based systems that do attempt to model non-specular phenomena do

so poorly [51, 100, 103]. Failure to accurately model various acoustical phenomena (in particular, diffractive [178, 180] and diffuse [51] components) leads to a decrease in the spatialization capabilities of the system. Failure to properly model diffuse reflections results in a geometric model that is not capable of capturing the true effects that result as a propagating sound interacts with objects/surfaces in the environment. This has several consequences, most notably, it leads to higher predicted reverberation times [81, 103]. In addition, the predicted room impulse response will be less steep than actual measured results [58] and generally, a decrease in the overall quality of the simulation results [50]. However, despite the importance of diffuse reflections, [51, 100, 103], geometric acoustical modeling systems have typically ignored their effects. Models that do attempt to account for diffuse reflections, according to Dalenbäck et al. [51], typically do so in a "crude manner" (see Martin [117] for a description of a hybrid method of simulating diffusion based on both physical and phenomenological modeling components). Typically, systems incorporating diffusion assume a *diffusion coefficient* δ for each surface to indicate the percentage of incoming energy that the surface will reflect diffusely. Although the use of such a diffusion coefficient can lead to improvements as opposed to modeling specular reflections alone, it is far from perfect. Furthermore, given the lack of a clear relationship between the diffusion coefficient and surface properties, there is no defined method of choosing an appropriate

coefficient value. Rather, the coefficient is chosen through trial and error and is therefore very application dependent [73].

3.2.2.1 The Image Source Method

The image source method [2] is a modeling mechanism used to estimate the path followed by acoustical energy subject to low order specular reflections [78]. A virtual sound source copy S_i of the original sound source S is created at a position obtained by *mirroring* the original sound source over each polygonal surface of a room [63]. Reflections up to any order can be produced by recursively repeating this procedure. A graphical illustration of the image source method is illustrated in Figure 3.1. After creating the first virtual sound source S_1 by mirroring the original source, a second order reflection can be created by treating S_1 as an "original" source and then mirroring it to create another virtual source S_2 and so on. For each virtual source, a visibility check is made to determine whether the virtual source is visible to the listener (the visibility check may be complex depending on the room being simulated). If the virtual source is visible to the listener, it is adjusted to account for the absorption of energy by the medium of propagation (e.g., air). Changes in the environment including movement of the sound source or listener or the introduction of any objects/obstructions in the environment may require the re-computation of all image sources as their visibility relative to the listener may change.

Although the image source method can find specular reflections up to a cer-

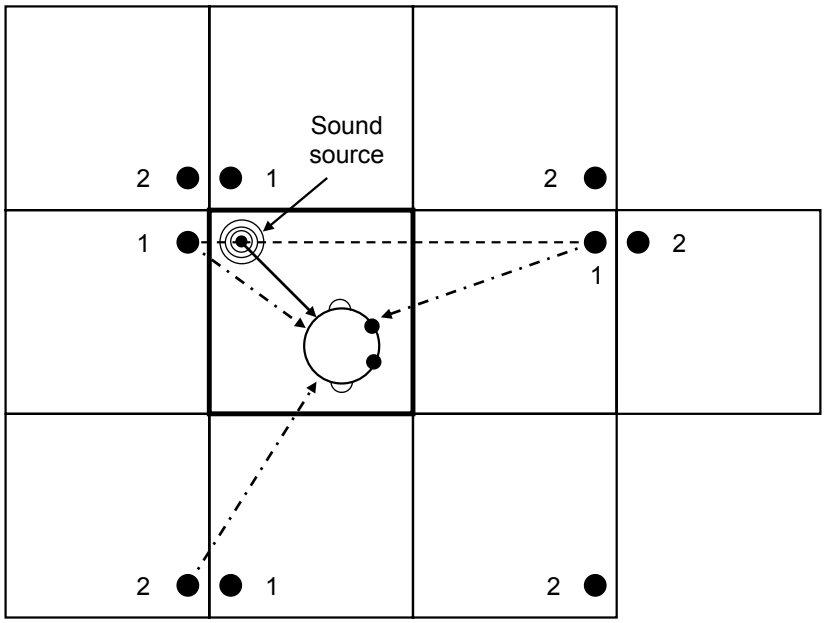


Figure 3.1: The image source method to determine low order specular reflections. The bold outlined rectangle represents the actual room with the listener and sound source. First order reflections are created by mirroring the sound source once (labeled with a 1). Multiple order reflections are created by mirroring the first order reflections (labeled with a 2) and so on.

tain order and the method is rather robust, it does have its shortcomings. Most importantly, as described by Funkhouser et al. [63], it can only model specular reflections and its computational complexity is exponential with respect to reflection order [63]. Given the potentially complex visibility checks that must be performed, the number of image sources that can be calculated is dependent on the available processing power. Due to their exponential running time, image source methods are of use only in simple environments to model a limited number of (e.g., low order) specular reflections [96].

3.2.2.2 Ray Tracing

Ray tracing methods are well known and widely used in computer graphics applications to render scenes. As with image source methods, ray tracing methods find the paths between a sound source and the listener. However, rather than mirroring the sound source, rays are emitted from the sound source in all directions and are followed through the environment until some pre-defined number of them reach the listener or until the ray's energy falls below some pre-defined level or the number of reflections exceeds some pre-defined number (see Figure 3.2). On their path from the sound source to the listener, the rays may encounter any number of surfaces (e.g., walls) or obstacles/obstructions. At this point, the rays are reflected once again (specular reflections are typically assumed, although diffuse reflection, diffraction and refraction can also

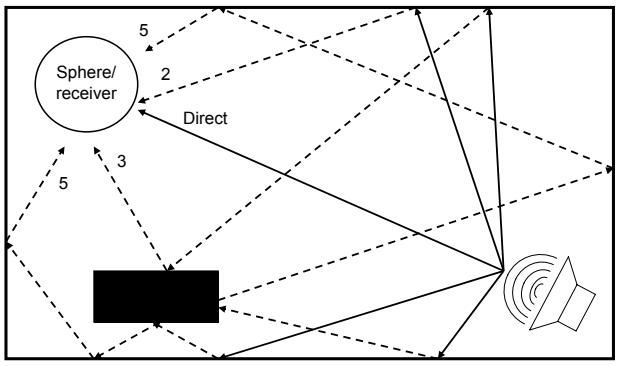


Figure 3.2: The acoustical ray tracing method. The acoustical rays are emitted from the sound source and traced through the environment while recording any interactions between any surfaces they may encounter. The room impulse response is approximated by recording the rays that reach the receiver (approximated by a sphere) over time.

be modeled). As with the image source method, the intensity of each reflection is reduced to account for absorption by both the medium of propagation (e.g., air) and the object it may encounter. Having rays emitted from the source in all directions is clearly impractical computationally as it will lead to a large number of reflections that must be followed. Rather, a subset of rays is emitted instead. Various methods can be used to choose this subset, including *stochastic* techniques which choose the paths followed by the rays randomly [63].

The ray tracing method has its share of advantages and disadvantages. Advantages include simplicity and manageable computational complexity, which increases sub-linearly with respect to the number of surfaces in the environment [63]. With respect to disadvantages however, given that a subset of the actual paths from the sound source to the listener are actually followed, certain paths may be missed altogether. Furthermore, ray tracing methods typically sample points on a regular grid [83] and therefore, aliasing is inherent in the

method. Rather than emitting and tracing a single ray from the source, multiple rays, bundled into a pyramid or beam can be emitted and traced instead. Such an approach was first introduced by Whitted in the field of computer graphics [194] and has since then inspired various other approaches, including cone tracing [3], whereby a single ray is replaced by a cone and the beam tracing approach which replaces a ray with a beam [64]. Greater details regarding the widely used beam tracing approach are provided in the following section.

Another problem associated with ray-based approaches is handling the large number of potential interactions between a propagating sound ray and any objects/surfaces it may encounter. A sound ray incident on a surface may be reflected both specularly and diffusely and be refracted and diffracted. Typical solutions to modeling such effects include the generation and emission of several "new" rays at each interaction point. Such approaches can lead to exponential running times and the algorithm can quickly become computationally intractable for all but very simple environments. Furthermore, a strategy is also needed to terminate a sound ray. One of the simplest solutions involves keeping track of the number of times a ray has been reflected (the *reflection count*) and terminating the ray once its reflection count has exceeded some pre-defined threshold value. Another approach is to terminate the ray based on its energy content. This second approach is more representative of the real world whereby the termination of a sound is determined by the amount of en-

ergy it has lost and not the number of times it has been reflected [200]. One measure of energy attenuation is the energy discontinuity percentage (EDP) [52]. The EDP represents the percentage of the original ray energy that must be lost before the ray is terminated. Regardless of whether a reflection count or EDP criteria is used to terminate an acoustical ray, assuming specular and diffuse reflections only (no diffraction and refraction) as commonly done, upon encountering a surface, three types of interactions may occur. A portion of the ray's energy may be absorbed by the surface, a portion reflected specularly and a portion reflected diffusely according to the following constraint [51]

$$\alpha + \delta(1-\alpha) + (1-\delta)(1-\alpha) = 1 \tag{3.1}$$

where, α is the incident surface absorption coefficient indicating the fraction of sound energy absorbed by the surface, δ is the incident surface diffuse reflection coefficient indicating the fraction of sound energy reflected diffusely and $(1-\delta)(1-\alpha)$ represents the amount of energy reflected specularly. Hence, at each point of incidence (provided $0 < \alpha < 1$ and $0 < \delta < 1$), two new rays are created; one that will be reflected specularly and the other diffusely. As a result, after M interactions, a total of 2^M rays can be generated. For example, considering a single ray and a reflection count of only 8, $2^8 = 256$ additional rays could be created and traced.

Rather than using such deterministic approaches to determine the type of

interaction between an acoustical ray and an incident surface, probabilistic approaches, such as a Russian roulette strategy [72] can be used instead. Russian roulette ensures that the path length of each acoustical ray is maintained at a manageable size yet due to its probabilistic nature, allows for paths of an arbitrary size to be explored. As described in Section 4.3.3, sonel mapping (introduced by Kapralos et al. [91]) employs a Russian roulette solution similar to the approach taken in photon mapping in order to determine the type of interaction between a sonel and a surface and to determine when the sonel is terminated. Simulations presented in Chapter 4.6 illustrate the capability of a Russian roulette strategy to provide an accurate, yet computationally reasonable solution to room acoustical modeling.

Despite the problems associated with ray tracing approaches, given their simplicity, they have been and are still widely used in a variety of room acoustical modeling applications. Examples include the RAYSCAT model that is used to predict noise levels in empty or fitted workshops [141] and the NORMAL model that is used to predict the sound field in long enclosures (and that was specifically designed to model speech intelligibility in underground stations [200]). Further applications include predicting noise in industrial workshops [82], predicting sound pressure levels and reverberation times in coupled spaces [183] and predicting the acoustics in classrooms [41].

Sonel mapping, the ray (particle)-based method developed by Kapralos et

al. [89, 90, 91, 92] and described in this dissertation, addresses some of the problems associated with existing ray-based approaches. Sonel mapping is capable of modeling specular and diffuse reflections and diffraction in a simple manner at a fraction of the computational time when compared to existing deterministic approaches. Given its stochastic nature (and in particular the use of a Russian roulette approach to determine some types of interaction), it allows for the possibility of exploring arbitrarily long paths that may not necessarily be explored using deterministic approaches. In addition, by employing a Russian roulette approach, the accuracy of the simulation can be improved by increasing the number of samples initially emitted from the sound source. Although this leads to an increase in computational time, an efficiency vs. accuracy trade-off can nevertheless be made. Greater details regarding the sonel mapping framework are provided throughout this dissertation and in particular, in Chapter 4.

As with sonel mapping, the acoustical simulation and visualization method known as *phonon tracing* presented by Bertram et al. [23], is also inspired by methods intended for photorealism and in particular, photon mapping. With phonon tracing, sound elements, termed *phonons*, are emitted from a sound source. After being emitted from the sound source following a particular source directional distribution function, phonons are traced through the scene while recording any interactions with any objects/surfaces they may encounter or un-

til they reach a receiver where they are collected. Upon encountering a surface, the energy carried by the phonon is reduced to account for surface and air absorption and is then reflected relative to the surface normal (e.g., specularly) provided its energy has not decreased below some pre-defined value or similarly, its reflection count (e.g., the number of times it has been reflected) has not increased beyond some pre-defined value. At each receiver, incident phonons are collected in the accumulating impulse response. Rather than modeling the receiver as a sphere with a given radius (as commonly done), the receiver is modeled using a Gaussian in order to provide "smoother" impulse response filters. A comparison of the phonon tracing method with the commercially available room acoustics program CARA [150] (that is based on image sources) has been performed. The response for a simple box-shaped room with a single sound source for various receiver positions was physically measured and then the room and each sound source/receiver configuration was modeled using both the phonon tracing method and CARA. The response calculated using phonon tracing was closer to the actual response when considering frequencies above 400Hz but differed significantly for lower frequencies due to diffraction effects and/or problems with the absorption coefficients used by both methods.

Although sonel mapping and phonon tracing are similar insofar that both are inspired by photon mapping, there are a number of critical differences. Phonon tracing is intended for mid and high frequencies in which reflections

are primarily specular. It thus ignores low frequencies and the diffraction effects associated with them (e.g., diffraction is not considered by phonon tracing). Given the mid and high frequency assumption, all reflections in the phonon tracing method are handled in the same manner - they are reflected with respect to the surface normal (e.g., they are treated as being purely specular). Finally, phonon tracing is a deterministic approach. In contrast to sonel mapping, tracing of a phonon is not terminated based on the value of a randomly generated number (e.g., using a Russian roulette approach), but rather, termination occurs after a phonon has been reflected a pre-defined number of times or after its energy level has decreased below some pre-defined threshold value.

3.2.2.3 Beam Tracing

Beam tracing was initially introduced in the area of computer graphics by Heckbert and Hanrahan [80] and later applied to acoustical modeling. Rather than tracing single linear rays from the source through the environment, beam tracing algorithms emit and trace pyramidal beams. An intersection test between the beam and the polygons comprising the environment is made. If an intersection is found, the original beam is clipped to remove any shadow region and a transmission beam, corresponding to the shadow region is constructed. The transmission beam is then mirrored over the plane of the intersecting poly-

gon. Beam tracing exploits geometric coherence since a single beam represents multiple rays and thus allows for large speedups when compared to conventional ray tracing. Beam tracing does not share the sampling artifacts associated with ray tracing (e.g., aliasing) or the overlap problems of cone tracing [64]. However, as opposed to conventional ray tracing, beam tracing is a much more time consuming and computationally intensive algorithm [83] given the complexity in performing the intersections between a beam and a polygon.

Funkhouser et al. [64, 180] developed an acoustical modeling approach based on beam tracing to model the early portion of the room impulse response and a statistical approximation to model the late portion of the room impulse response. They overcome the main limitation associated with beam tracing methods (e.g., the difficulty in tracing beams through three-dimensional space in order to generate propagation paths quickly and the ability to handle edge diffraction), through the use of a winged edge data structure [13]. The winged edge data structure allows sequences of surface and edge scattering to be precomputed and stored and later used to compute propagation paths to arbitrary receiver positions to allow for real-time auralization during interactive use by a user. Their method is composed of four phases, the first two of which are executed off-line prior to the start of the simulation. In the first phase (the *spatial subdivision* phase), spatial relationships between the polygons comprising the environment are computed and stored in a cell adjacency graph data struc-

ture. In the second phase (the *beam tracing* phase) separate beams to account for transmission, diffraction and reflection are emitted from each sound source and traced in order to construct a beam tree data structure. The beam tree data structure records the region of space reachable by each sequence of reflection and transmission paths from each sound source. In the third phase (the *path generation* phase), the beam tree data structure is used to compute the propagation paths from each sound source to the receiver. Finally, during the last phase (the *auralization* phase), the impulse response is used to develop a filter that, when convolved with an anechoic sound, results in a spatialized sound. Although the system is capable of providing interactive update rates and account for wedge diffraction effects, it is still a research prototype and does have its share of limitations [64]. In particular: i) the scene model must consist of planar polygons only, ii) diffuse reflections are not accounted for, iii) surface dimensions must be greater than the wavelength of audible sound, iv) valid only for three-dimensional models without highly faceted surfaces and v) the majority of occluding and reflecting surfaces of the environment must be static throughout the simulation.

3.2.2.4 Combined Methods

In addition to methods relying on a single approach to modeling the acoustics of an environment, many approaches combine multiple techniques, taking ad-

vantage of the strengths associated with each separate technique. Lewers [104] combined beam tracing with radiant exchange methods (e.g., radiosity). Beam tracing using a triangular beam tracing model (a triangular beam is defined by a central axis and three bounding planes, each forming a side of the beam) is used to find the image sources (specular reflections), of a particular room, in essence, computing the early portion of the impulse response. The latter (reverberant) portion of the room impulse response is determined using a radiant exchange model based on a network topology. Each surface is replaced by a node and sound paths between surfaces are replaced with lines (connections) between nodes. The connections indicate the movement of diffuse energy between nodes (surfaces) based on a form factor (that describes the fraction of energy leaving one node and arriving at another) between them.

Monks et al. [131] combined the image source and beam tracing approaches to model the early portion of the room impulse response and used a statistical model to approximate the latter portion of the room impulse response. Another example includes combining both image source and ray tracing approaches to predict the acoustics of ancient open-air theaters [118].

3.2.3 Diffraction Modeling

Since the dimensions of many of the objects/surfaces encountered in our daily lives are within an order of magnitude as the wavelength of audible sounds, dif-

fraction is an elementary means of sound propagation, especially when there is no direct path between the sound source and the receiver. Many auralization methods are based on geometric acoustics. They assume sound is a ray phenomena [48] and model all interactions between a sound ray and objects/surfaces as specular, thus ignoring important effects such as diffraction and diffusion. Although ray-based approaches are simple to model and implement, they are valid primarily for higher frequency sounds where reflections are indeed primarily specular. In addition, they typically ignore the wavelength of sound and any phenomena associated with it, including diffraction [36, 100, 178, 180]. However, failure to account for diffraction can lead to a non-realistic auditory simulation.

That being said, a limited number of research efforts have investigated acoustical diffraction modeling. The beam tracing approach of Funkhouser et al. [64, 180] described in the previous section includes an extension capable of approximating diffraction. Their frequency domain method is based on the *uniform theory of diffraction* (UTD) [93]. Validation of their approach by Tsingos et al. [179] involved a comparison between the actual measured impulse response in a simple enclosure (the "Bell Labs Box") and the impulse response obtained by simulating the enclosure. According to Tsingos et al., their combined technique was the first instance to use a physically-based diffraction model to produce interactive rate sounds in a complex virtual environment.

Tsingos and Gascuel [181] developed an occlusion and diffraction method that utilizes computer graphics hardware to perform fast sound visibility calculations that can account for specular reflections (diffuse reflections are not considered), absorption and diffraction caused by partial occluders. Specular reflections are handled using an image source approach. Diffraction is approximated by computing the fraction of sound that is blocked by obstacles between the path from the sound source to the receiver by considering the amount of volume of the first Fresnel ellipsoid that is blocked by the occluders. A visibility factor is computed using computer graphics hardware. A rendering of all occluders from the receiver's position is performed and a count of all pixels not in the background is taken (pixels that are "set" e.g., not in the background, correspond to occluders). Their approach handles a discrete set of frequency bands ranging from 31Hz to 8kHz and is primarily focused on sounds for animations. Although experimental results are not extensive, their approach is capable of computing a frequency dependent visibility factor that, unlike other ray-based approaches, takes advantage of graphics hardware to perform this in an efficient manner. Although their approach is not completely real-time, it is "capable of achieving interactive computation rates for fully dynamic complex environments" [181].

Tsingos and Gascuel later introduced another occlusion and diffraction method based on the Fresnel-Kirchoff optics-based approximation to diffrac-

tion [180, 182]. As with the Huygens-Fresnel approximation, the Fresnel-Kirchoff approximation is based on Huygens' principle. The total unoccluded sound pressure level at some point p in space is determined by calculating the sound pressure of a small differential area dS and integrating over the closed surface enclosing p (see Tsingos and Gascuel [182] for further details regarding this calculation in addition to an algorithm outlining the method). After determining the total unoccluded sound pressure arriving at point p from a sound source, diffraction and occlusion effects are accounted for by computing an *occlusion depth-map* of the environment between the sound source and the receiver (listener) using computer graphics hardware to permit real-time operation. Once the depth-map has been computed, the depth of any occluders between the sound source and the receiver can be obtained from the Z-buffer [83] whereby "lit" pixels correspond to occluded areas. The diffraction integral described by the Fresnel-Kirchoff approximation is then approximated as a discrete sum of differential terms for every occluded pixel in the Z-buffer. Given the use of graphics hardware, their method is fast and is well suited to the interactive auralization of diffracted energy maps [182]. Comparisons for several configurations with obstacles of infinite extent between their method and between boundary element methods (BEMs), gives "satisfactory quantitative results" [182].

Various other research efforts have examined non-geometric acoustics based

diffraction modeling. Torres et al. [178] describe a time-domain model based on the Biot-Tolstoy-Medwin technique [24], that computes edge diffraction components and combinations of specular and diffracted components. Lokki et al. [105], Calamia et al. [36] and Svensson et al. [174] have also investigated diffraction modeling based on the Biot-Tolstoy-Medwin technique [24]. The method of Calamia et al. provides an integrated approach to acoustical modeling whereby intermediate values typically used in diffraction calculations using the Biot-Tolstoy-Medwin technique are exploited to find specular reflections as well [36]. Such techniques are currently not applicable to interactive applications due to complexity issues and are therefore not considered further here.

Finally, the approaches presented above describe physical-based solutions. In contrast to highly detailed physical approaches, Martens et al. [113] describe a perceptually-based solution to the diffraction of sound by an occluder of "low computational cost that is capable of producing distinctive auditory spatial images associated with identifiable effects". Their approach consisted of three phases. In the first phase, acoustical measurements of an actual occluder were made in an anechoic chamber. Next, a digital filter model was designed to account for the observed effects to the impulse response depending on the occluder's size, position and orientation. Finally, based on the constructed filter model, experimental stimuli were prepared and presented to human listeners in order to determine the variation in human response due to variations of the

model parameters.

3.3 Summary

Current approaches to computational acoustical modeling can be broadly divided into two categories; *wave-based modeling* and *ray-based modeling* [160]. With wave-based methods, the objective is to solve the wave equation, also known as the *Helmholtz-Kirchoff* equation, to recreate a particular sound field. Unfortunately, an analytical solution to the *wave equation* is rarely feasible [160] and wave-based methods use numerical approximations, such as finite element methods (FEM), boundary element methods (BEM) and finite difference time domain methods (FDTD) instead [160]. These approximations are computationally very expensive, and therefore, such techniques are currently beyond our computational ability for all but very simple, trivial scenarios.

In a manner similar to geometric optics, ray-based acoustical modeling assumes sound acts as rays and light (visual) based rendering techniques are used to model the acoustics of an environment. Essentially, the propagation paths taken by the rays are found by tracing (following) these acoustical rays. Mathematical models are used to account for source emission patterns, atmospheric scattering, absorption of a sound ray by the medium itself and the interactions with any surfaces/objects a ray may encounter. Ray-based methods are fairly simple to model and are widely used in acoustical modeling appli-

cations. Unfortunately, they are only valid for higher frequency sounds where reflections are typically specular. Ray-based methods ignore the wavelength of sound and phenomena associated with wavelength. By ignoring such phenomena, they can only accurately model the early portion of the room impulse response (echogram). Ray-based methods also fail to properly account for diffuse reflections, leading to higher predicted reverberation times and generally, a decrease in the overall quality of the auralization.

In this and the previous sections a review of various existing acoustical modeling techniques and approaches have been presented. The focus on the following section switches to the global illumination method photon mapping. As sonel mapping is itself based on photon mapping, a detailed description is provided here of the photon mapping technique.

3.4 Photon Mapping

In parallel with the development of acoustical modeling algorithms, the computer graphics research community was developing sophisticated algorithms for modeling the energy transport associated with light. Early algorithms based on polygon coloring [83], ray tracing [6, 69] (also known as ray casting [83]), were being augmented with more global energy distribution algorithms such as radiosity [46, 70] and photon mapping [84]. Photon mapping is perhaps the most sophisticated of these algorithms, in that it can be viewed as a gen-

eral algorithm for modeling the transmission and interaction of energy. This dissertation adapts the photon mapping approach to the modeling of the transmission of acoustical energy (sound). This section reviews the basic theory behind photon mapping and identifies its limitations with respect to acoustical modeling.

Photon mapping is a two-pass method that is used to approximate the rendering equation initially proposed by Kajiya [86]. According to the rendering equation, the outgoing radiance (power per unit projected area perpendicular to the ray per unit solid angle in the direction of the ray [46]) $L_\circ(x,\vec{\omega})$, at a point x on a surface S in the direction $\vec{\omega}$ is equal to the radiance emitted at point x by the surface itself $L_e(x,\vec{\omega})$, in addition to any incoming radiance reflected at point x in the direction $\vec{\omega}$. This reflected radiance, denoted as $L_r(x,\vec{\omega})$, includes any radiance reaching point x from all incoming directions $\vec{\omega}'$ (and then reflected in direction $\vec{\omega}$) including any radiance coming directly from sources, or radiance which has been reflected off of other surfaces in the environment one or more times prior to reaching point x. The relationship between the radiance incoming from direction $\vec{\omega}'$ at point x and the radiance reflected in direction $\vec{\omega}$ once again at point x, is summarized by the bi-directional reflectance distribution function (BRDF) $f_r(x,\vec{\omega}',\vec{\omega})$. Given these considerations, mathematically, the reflected radiance term $L_r(x,\vec{\omega})$ is given by

$$L_r(x,\vec{\omega}) = \int_{\Omega} f_r(x,\vec{\omega}',\vec{\omega}) L_i(x,\vec{\omega}')(\vec{\omega}'\cdot\vec{n}) d\vec{\omega}'$$

where, $L_i(x,\vec{\omega}')$ is the incident radiance at point x, incoming from direction $\vec{\omega}'$, n is the surface normal vector at point x and the domain Ω is the unit hemisphere centered about point x. Given the different types of reflection phenomena associated with light waves (e.g., diffuse, specular and glossy), the BRDF can be further sub-divided into two parts, a specular/glossy component and a diffuse component, $f_{r,S}$ and $f_{r,D}$ respectively. The incoming radiance $L_i(x,\vec{\omega}')$ can be divided into a sum of the following three components

$$L_i(x,\vec{\omega}') = L_{i,l}(x,\vec{\omega}') + L_{i,c}(x,\vec{\omega}') + L_{i,d}(x,\vec{\omega}') \tag{3.2}$$

where,

$L_{i,l}(x,\vec{\omega}')$ is the incoming radiance coming directly from the light sources without being reflected.

$L_{i,c}(x,\vec{\omega}')$ is the incoming radiance resulting from caustics or in other words, indirect energy from light sources that has been reflected specularly or transmitted through a surface.

$L_{i,d}(x,\vec{\omega}')$ is the incoming radiance via diffuse reflections.

With this decomposition of the BRDF and the incoming radiance, the re-

flected radiance $L_r(x,\vec{\omega})$ can finally be divided into a sum of four terms, each term accounting for either the direct, diffuse, specular, or caustic illumination, representing the total reflected radiance at point x in direction $\vec{\omega}'$. Each term is computed by considering all incoming directions over a hemisphere centered around point x using integration. Hence, the reflected radiance can be decomposed into a sum of four integrals which describe the following components:

1. Direct illumination: $\int_\Omega f_r(x,\vec{\omega}',\vec{\omega})L_{i,l}(x,\vec{\omega}')(\vec{\omega}'\cdot\vec{n})d\vec{\omega}'$

2. Specular and glossy reflection:

   $\int_\Omega f_{r,S}((x,\vec{\omega}',\vec{\omega})(L_{i,c}(x,\vec{\omega}')+L_{i,d}(x,\vec{\omega}'))(\vec{\omega}'\cdot\vec{n})d\vec{\omega}'$

3. Caustics: $\int_\Omega f_D(x,\vec{\omega}',\vec{\omega})L_{i,c}(x,\vec{\omega}')(\vec{\omega}'\cdot\vec{n})d\vec{\omega}'$

4. Diffuse reflections: $\int_\Omega f_{r,D}(x,\vec{\omega}',\vec{\omega})L_{i,d}(x,\vec{\omega}')(\vec{\omega}'\cdot\vec{n})d\vec{\omega}'$

Finally, the outgoing radiance $L_o(x,\vec{\omega})$ is given as the sum of any radiance emitted by the surface at point x in direction $\vec{\omega}$ ($L_e(x,\vec{\omega})$) and any radiance reflected at point x in direction $\vec{\omega}$ ($L_r(x,\vec{\omega})$) [84]

$$
\begin{aligned}
L_o(x,\vec{\omega}) &= L_e(x,\vec{\omega}) + L_r(x,\vec{\omega}) \\
&= L_e(x,\vec{\omega}) + \int_\Omega f_r(x,\vec{\omega}',\vec{\omega}) L_i(x,\vec{\omega}')(\vec{\omega}'\cdot\vec{n}) d\vec{\omega}' \\
&= L_e(x,\vec{\omega}) + \int_\Omega f_r(x,\vec{\omega}',\vec{\omega}) L_{i,l}(x,\vec{\omega}')(\vec{\omega}'\cdot\vec{n}) d\vec{\omega}' + \\
&\quad \int_\Omega f_{r,S}((x,\vec{\omega}',\vec{\omega})(L_{i,c}(x,\vec{\omega}') + L_{i,d}(x,\vec{\omega}'))(\vec{\omega}'\cdot\vec{n}) d\vec{\omega}' + \\
&\quad \int_\Omega f_D(x,\vec{\omega}',\vec{\omega}) L_{i,c}(x,\vec{\omega}')(\vec{\omega}'\cdot\vec{n}) d\vec{\omega}' + \\
&\quad \int_\Omega f_{r,D}(x,\vec{\omega}',\vec{\omega}) L_{i,d}(x,\vec{\omega}')(\vec{\omega}'\cdot\vec{n}) d\vec{\omega}'
\end{aligned}
$$

Photon mapping is a two-pass process. In the first pass, *photons* are emitted from each light source and traced through the scene until they interact with a surface. Once they encounter a diffuse surface, they are stored in a structure called a *photon map*. In the second stage, the scene is rendered using the information provided by the previously collected photon map to provide a quick estimate of the reflected radiance. In addition, distribution ray tracing is employed to accurately model specular and caustic effects.

A brief description regarding the two stages of photon mapping and how these two stages are used to solve the computation of a very complicated integral (e.g., the integral of Equation 3.3) are provided in the following sections. A complete and detailed mathematical derivation of the rendering equation as well as a detailed description regarding the use of photon mapping to solve the decomposed rendering equation is given by Jensen [84].

3.4.1 Stage One: Photon Pass

During the first stage, photons are emitted from each light source in the scene in a manner dependent on the type of light (for example, when considering a point light source, photons are emitted uniformly in all directions). Each photon has an associated position (x,y,z coordinates corresponding to the point of intersection between the photon and the encountered surface), incoming direction (azimuth and elevation angles, θ and ϕ respectively) and power. Any type of light source can be modeled, including a point light source or a light which is square, spherical or directional. However, its total power is typically divided equally amongst all the photons it emits. Each emitted photon is traced through the scene using photon tracing, a technique similar to standard ray-tracing [84]. Upon encountering a surface, the photon can be either reflected, transmitted or absorbed by the surface. Which of these three events actually occurs is chosen probabilistically using a Russian roulette strategy [7] (see Appendix B).

When the encountered surface is diffuse or glossy (non-specular), the photon's information is updated (e.g., the position of the point of incidence and incident direction) and the photon (along with its updated information), is stored in the *global photon map*. The photon is then reflected in a randomly chosen direction in the hemisphere above the intersection point. When the encountered surface is purely specular (e.g., a mirror), the photon is not stored in the

global photon map but rather, reflected with a reflectance angle equal to the angle of incidence (e.g., pure specular reflection). Photons encountering specular surfaces are not stored in the global map as such an interaction would not provide any significant information given the small probability of matching an incoming photon from the specular direction [84]. To handle caustics, a separate photon map known as a *caustic map* is used to store caustic illumination or in other words, photons which have encountered a diffuse or glossy surface after previously being reflected specularly. Once such a photon encounters the diffuse surface, the photon, along with its relevant information, is stored in the caustic map and terminated (e.g., not reflected anymore). Stage one is independent of the camera (viewing) position.

3.4.2 Stage Two: Rendering

Once the global and caustic photon maps have been created, scene rendering begins. Rendering is accomplished using distribution ray tracing whereby the value (radiance) of each pixel is determined by averaging the value of the radiance encountered by several rays at their first points of intersection after tracing the rays from the "camera" position, through the pixel and into the scene. The value of the radiance at each point of intersection is calculated by solving the rendering equation (Equation 3.3), which contains the four separate integral terms used to describe the reflected radiance. For each of the

integrals, either an approximate or an accurate solution is obtained. Depending on which component of the expanded reflected radiance term $L_r(x, \vec{\omega})$ and whether an accurate or approximate solution is being sought, the photon maps may or may not be used. The radiance leaving a surface is estimated using the photons stored in the photon map and thus can be formulated as a *density estimation* problem where the goal is to estimate some probability density function (pdf), in this case representing the radiance leaving the surface, given a number of observed samples (the collected photons) [173]. Various density estimation techniques can be used to estimate the pdf, including the *histogram* method [176], *kernel* method [170] and finally, the method used in photon mapping, the *nearest neighbor* method [170]. With the nearest neighbor method, the radiance of a point is estimated by considering some pre-defined number of photons around the point of interest. Density estimation can be a complex and compute intensive task requiring a substantial amount of time to compute and a large amount of memory storage.

An accurate solution to the direct lighting component is computed using ray tracing methods. At the point of interest x, and for each light source, a ray, referred to as a "shadow ray", is emitted and traced towards the light source (depending on the type of light source, multiple shadow rays may be emitted). If the ray encounters any objects/surfaces, then there is no direct lighting contribution from that particular light source otherwise, its contribution is included

in the rendering equation. An approximate solution to the direct lighting component can be obtained by simply using the global photon map (e.g., the diffuse components only).

The specular and glossy components are computed without the use of the photon map as it would require the storage of a large amount of photons to obtain an accurate solution. Instead, this component is computed using Monte-Carlo ray tracing techniques whereby the integral is evaluated using Monte-Carlo sampling techniques. Caustics are computed accurately using the caustic photon map created in the first pass. Finally, an approximate solution to the term representing multiple diffuse reflections can be obtained using the global photon map directly. For an accurate evaluation, Monte-Carlo ray tracing with importance sampling is used instead.

3.4.3 Photon Mapping and Acoustical Modeling

Although sound and light are both wave phenomena and share many properties and characteristics (e.g., wavelength, frequency and amplitude), there are also many differences between them arising from the fact that sound waves are one type of a *mechanical* wave whereas light waves are a type of an *electromagnetic* wave. Mechanical waves propagate through disturbances of a medium (e.g., air or water). The particles of the medium do not move with the wave but rather oscillate about their equilibrium position while energy is being moved

[198]. Electromagnetic waves are formed when an electric field is coupled to a magnetic field (at right angles to each other) and together as a structure propagate through space [198]. In contrast to mechanical waves, electromagnetic waves do not require a medium to propagate and can therefore propagate in a vacuum or empty space. In addition, there are many properties associated with electromagnetic waves, such as polarization (which specifies the direction of the electric field), which do not apply to mechanical waves. Given these differences, one cannot simply apply photon mapping or any other computer graphics (image synthesis) based method directly to the task of acoustical modeling. In addition to the differences in wavelength between sound (0.02m to 17m) and light (about 400nm to 700nm), there are also several major differences associated with the propagation of sound and light. The fundamental differences between the propagation of light and sound waves that must be addressed by any acoustical modeling method are as follows [146, 163]:

1. Slower propagation speed of sound.
2. Greater attenuation (damping) of sound by the medium (air).

Although dependent on medium temperature, for most practical purposes, the speed of sound in air can be approximated by $343\text{m}\cdot\text{s}^{-1}$ [48], rather slow when compared to speed of light $299{,}792{,}458\text{m}\cdot\text{s}^{-1}$. When considering light propagation, given the high velocity of light, propagation delay effects are almost negligible and it is typically safe to assume propagation is instantaneous

in our natural surroundings (except perhaps when considering objects traveling at or near the speed of light). On the contrary, we are clearly capable of detecting (perceiving) the propagation delays of sound as it arrives from the sound source to the receiver both directly and indirectly via reflections. Similarly, the attenuation of light by the medium is so small that it is ignored by the conventional image rendering algorithms that assume light only interacts with surfaces/objects except in unusual situations such as when light passes through fog, clouds or smoke [163, 169]. The attenuation of sound by the medium (e.g., air) cannot be ignored since it can be significant and can certainly be perceived by humans. In fact, the absorption of the higher frequency components of sound varies as a function of distance, providing a cue to the distance between the sound source and human observer [88].

In addition to these fundamental differences, another difference between light and sound waves, is the coherence of sound waves as opposed to the incoherence of light waves (except for lasers which are also coherent) [64]. In other words, depending on the phase difference between the two sound waves, the waves may interfere with each other either constructively (no phase difference) or destructively (differing phase). As a result, phase must be considered by any simulation involving sound and furthermore, since the phase of a wave in such a simulation is determined by the distance traveled (e.g., the "path length"), this path length must be computed accurately [64]. Since light waves

are incoherent, the phase relationship between two light waves does not need to be considered by any simulation.

3.4.4 Photon Mapping Summary

Photon mapping is a two-pass "particle-based", probabilistic global illumination method developed by Jensen in 1995 in order to determine the illumination at any point in a scene. In the first pass, "photons" (the basic quantity of light) are emitted from each light source and traced through the scene until they interact with a surface. When photons encounter a diffuse surface, they are stored in a structure called a *photon map*. In the second stage, the scene is rendered using the information provided by the previously collected photon map to provide a quick estimate of the diffuse reflected illumination. Distribution ray tracing is employed to model specular effects. Photon mapping is independent of the scene geometry, thereby allowing for the illumination of arbitrary complex scenes to be computed. In addition, it can handle complex interactions between light and a surface, including pure specular, pure diffuse and glossy reflections and any combination of them. Before applying photon mapping to the task of acoustical modeling, several differences between the propagation of light and sound must be addressed. In particular, the slower propagation speed of sound and the attenuation (damping) of sound by the air must be accounted for. Recently, an acoustical modeling algorithm inspired by

photon mapping termed sonel mapping, has been developed. Sonel mapping, a stochastic particle-based method developed by Kapralos et al. [89], is capable of modeling specular and diffuse reflections and diffraction in a simple manner at a fraction of the computation time when compared to existing deterministic approaches. It provides the possibility of exploring arbitrarily long paths that may not necessarily be explored using other, deterministic approaches, allowing the accuracy of the simulation to be improved by increasing the number of samples initially emitted from a sound source.

Chapter 4

Sonel Mapping

This chapter details the sonel mapping method[1]. The chapter begins with an overview of sonel mapping followed by a detailed description regarding its probabilistic framework, outlining the modeling of the interaction between sound particles and surfaces they encounter, including specular and diffuse reflections, absorption and diffraction.

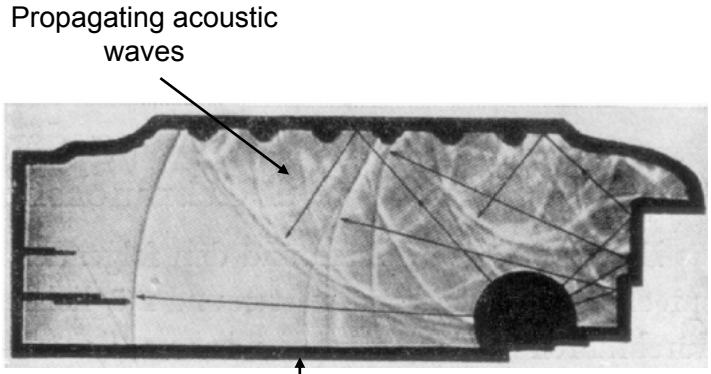
4.1 Defining the Problem and the Approach

Sound waves are a *mechanical wave* phenomenon that result from the rapid variations in air pressure (e.g., the back and forth oscillation of the molecules comprising the medium) caused from the vibrations of an object. Although in-

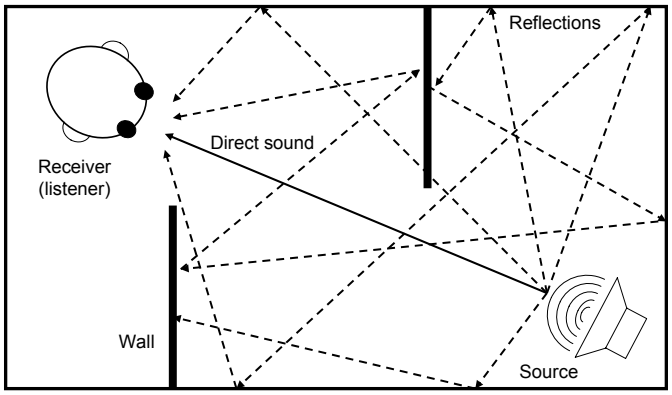
[1]A condensed description of the sonel mapping method appears in [92].

dividual molecules oscillate about their equilibrium positions, they do not move with the wave. What does move is energy. In other words, sound waves transport acoustical energy. After being emitted from sound sources, the waves propagate through the environment, interacting with objects/surfaces they may encounter and with the medium itself. Some of the waves may also reach a receiver (listener). An example illustrating the propagation of sound waves is provided in Figure 4.1(a) where reflections of ultrasound wave fronts in a cross-sectional model of the Gewandhaussaal hall in Leipzig [151] is shown.

As previously described, the goal of auralization is to recreate a particular listening environment, taking into account the acoustics of the environment and the characteristics of the listener. The pressure (e.g., force per unit area) at every point of wave propagation can be described by and the wave therefore satisfies, the *wave equation*, also known as the *Helmholtz Kirchoff* equation [30]. By solving the wave equation for every time instance and for every position within a particular environment, a particular soundfield can, in theory, be completely recreated. Unfortunately, an analytical solution to the wave equation is rarely feasible (perhaps only in some very simple and trivial scenarios) and beyond our current analytical and computational reach. As a result, the propagation of energy contained in a wave, be it a sound wave or any other type of wave, must be approximated using other techniques instead (e.g., image sources, acoustical ray tracing, acoustical radiosity - see Chapter 3 for



(a) Ultrasound wave fronts.



(b) Propagation of acoustical rays.

Figure 4.1: Sound propagation example. (a) Reflections of ultrasound wave fronts in a cross-sectional model of the Gewandhaussaal hall in Leipzig photographed using Schlierin photography [159] (reprinted from [151]). (b) Sound approximated as acoustical particles (or acoustical rays). Sound particles/rays emitted from the sound source propagate within the environment, interacting with any objects/surfaces they may encounter and may eventually reach the receiver (listener).

greater details regarding such approximations). The sonel mapping technique described here is one such approximation.

Sonel mapping is a two-stage, Monte-Carlo particle-based algorithm. At each sound source, acoustical modeling is approximated by emitting one or more sound particles (sound elements or "sonels") from sound sources and tracing these sonels through the environment while handling the interactions with any objects/surfaces they may encounter. Figure 4.1(b) provides a graphical illustration of the problem to be solved. Sonels are emitted from sound sources in a manner specified by the sound source distribution function that describes the directional and frequency distribution of each sound source. Once they are emitted, sonels are traced through the environment and interact with objects/surfaces they may encounter.

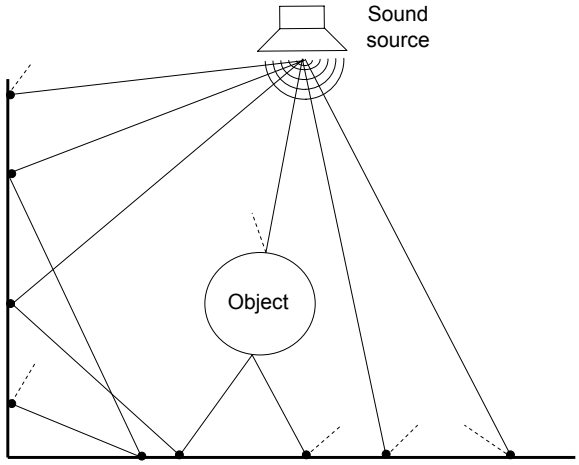
As sonels move through the environment they contribute to the *sonel map*. The sonel map is used in a second stage of the algorithm to provide an estimate of the sound energy incident at a particular point in the scene. An overview of the sonel mapping method in addition to greater details regarding the two stages comprising the algorithm, is presented below.

4.2 Sonel Mapping Overview

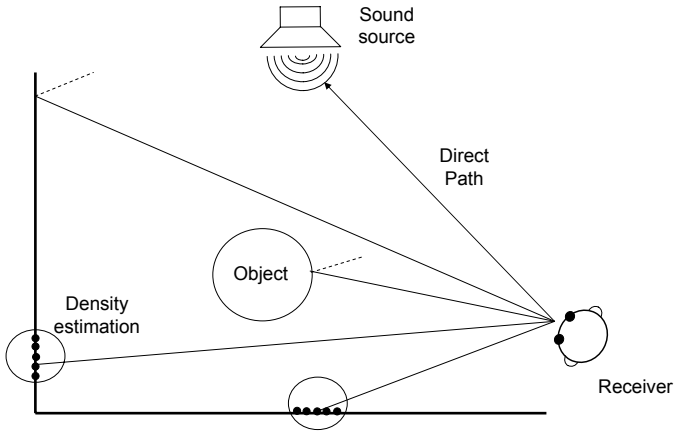
The main concept in sonel mapping (as in photon mapping) is to trace the energy propagation from the sound sources through the environment (scene)

while recording the interaction with any surfaces this energy may encounter in the sonel map (see Figure 4.2(a)). The sonel map is global data structure that stores acoustical energy as points. Once the sonel map has been constructed, using *density estimation* techniques, it allows for a very quick estimate of the sound energy at some point within the model to be made by considering the information within the sonel map only. In other words, the information can be re-used many times over without having to re-compute it using Monte-Carlo ray tracing. This can lead to tremendous computational time savings. Only sonels that are incident upon a diffuse surface are stored in the sonel map. Storing sonels incident on specular surfaces does not provide any useful information since the probability of having a matching sonel from the specular direction is small (zero for a perfect specular surface) and therefore, specular reflections are best handled using standard ray-tracing [84]. In addition, since the Huygens-Fresnel principle is formulated for the energy transfer between a sound source and a receiver, the diffracted energy reaching a receiver is determined in the acoustical rendering stage and diffracted sonels are also not stored in the sonel map. Once the sonel map has been constructed, the complete energy transmission process is computed by tracing out from the receiver using distribution (Monte-Carlo) ray-tracing coupled with the previously constructed sonel map (see Figure 4.2(b)).

Sonel mapping is a two-stage algorithm, a *sonel tracing stage* followed by



(a) Populating the sonel map.



(b) Computing the energy reaching the receiver (echogram).

Figure 4.2: Sonel mapping. (a) Sonels propagate from the sound sources and traced through the scene while recording the interaction with any objects/surfaces they may encounter. Depending on the type of interaction, sonels may be stored in the sonel map. (b) Once the sonel map has been constructed, the complete energy transmission process is computed by tracing out from the receiver using distribution (Monte-Carlo) ray-tracing coupled with the previously constructed sonel map.

an *acoustical rendering stage*:

Sonel tracing stage: Construct and populate the sonel map by emitting sonels from sound sources and tracing them through the environment while handling interactions with any objects/surfaces they may encounter.

Acoustical rendering stage: Estimation of the echogram using distribution ray tracing from the listener's position along with the information contained in the previously constructed sonel map.

Figure 4.3 illustrates the sonel mapping method. Prior to the simulation, a description of the scene and the model is required. The scene description provides information specific to the objects and surfaces comprising the environment including positional information (e.g., three-dimensional surface structure) along with surface characteristic properties (e.g., absorption, diffuse and specular reflection coefficient values). The sonel mapping algorithm can handle arbitrarily complex scenes and models. Sound source distribution functions that describe the frequency and directional emission patterns of each sound source are supplied in addition to sensitivity functions for each receiver.

Sonel mapping is an approximation to the actual acoustical energy transport. In reality, when a sound wave encounters a surface, multiple interactions may take place (e.g., a portion of the wave may be absorbed by the surface, reflected specularly, reflected diffusely, and diffracted). However, incorporating such multiple-interactions in any acoustical modeling algorithm can quickly

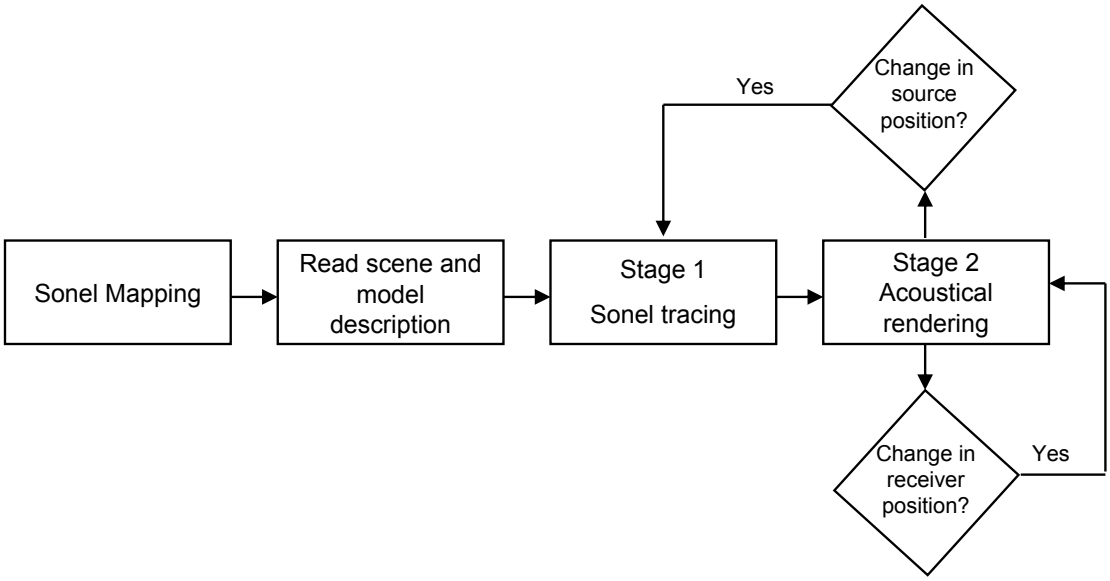


Figure 4.3: Overview of the sonel mapping acoustical modeling method. Sonel mapping is a two-pass Monte-Carlo particle-based acoustical modeling technique. Prior to beginning the simulation, a description of the scene and the model is obtained. The scene description describes the environment to be simulated. In particular, it includes positional information for all surfaces/objects along with their corresponding surface coefficient values (e.g., absorption, diffuse and specular reflection coefficient values). The model description provides information specific to the sound sources, receivers (e.g., position in the environment, energy distribution functions etc.) and for handling interactions between the propagating sonels/acoustical visibility rays and any surface/object they may encounter. Once the scene and model descriptions have been obtained, the simulation begins. The purpose of the first stage is to populate the sonel map. The sonel map is a global structure that, for each surface, records the amount of diffusely reflected energy incident on it. Once the sonel map has been populated, the second stage (acoustical rendering) can begin. The purpose of the acoustical rendering stage is to approximate the distribution of frequency dependent energy arriving at the receiver over a period of time (e.g., to compute the echogram) by using the previously constructed sonel map coupled with distribution (Monte-Carlo) ray tracing. Provided sound sources remain static, the information contained in the sonel map does not need to be updated and therefore, to account for the changing soundfield associated with a moving receiver, only the acoustical rendering stage needs to be performed.

lead to exponential running times making such an approach impractical for any real-time applications. In sonel mapping, at each point of interaction between a sonel and a surface rather than allowing multiple interactions, only one type of interaction occurs. The type of interaction that does occur is chosen probabilistically based upon characteristics of both the surface (e.g., reflection coefficient values) and of the sonel (e.g., frequency) using a Russian roulette approach. Russian roulette is an importance sampling technique used to increase the efficiency of an estimator [173]. The probability distribution function is used to eliminate portions of the domain that are not important [84]. As the number of sonels emitted from the sound source is increased (e.g., the number of samples is increased), the solution converges towards the correct result [84] and thus the approximation can model the combined sound energy propagation effects. As commonly done in various acoustical modeling applications, a sphere is used to represent a receiver [199].

In the first stage (the sonel tracing stage), sonels are emitted from each sound source and traced through the scene until they interact with a surface. Upon encountering a surface, a check is made to determine whether the sonel incident point is within the *diffraction zone* or *non-diffraction zone*. When the sonel is incident within the non-diffraction zone, the sonel will either be reflected specularly (assuming ideal specular reflection), reflected diffusely (assuming ideal or Lambertian reflection) or completely absorbed by the surface.

Diffusely reflected sonels are stored in the sonel map. When the sonel is incident within the diffraction zone, a new sonel is generated and emitted by choosing a random direction over the hemisphere centered about the diffraction point or the sonel will be transmitted through the diffraction zone unaltered. In the second stage (the acoustical rendering stage), the echogram is estimated through the use of the previously constructed sonel map coupled with distribution ray tracing. acoustical visibility rays are traced from the receiver into the scene where they may interact with any surfaces/objects they may encounter. When an acoustical visibility ray intersects a diffuse surface at point p, tracing of the ray terminates and the sonel map is used to provide an estimate of the acoustical energy leaving point p and arriving at the receiver using a *density estimation* algorithm. The energy is scaled to account for attenuation by the medium and added to the accumulating echogram. Specular reflections are handled using the same approach as in the sonel tracing stage whereby ideal specular reflections are assumed. When an acoustical visibility ray encounters a sound source, the fraction of energy leaving the sound source and arriving at the receiver is determined, scaled to account for attenuation by the medium and then added to the accumulating echogram. Diffraction effects that occur when an acoustical visibility ray encounters and edge are handled using a modified version of the Huygens-Fresnel principle. Provided the sound source remains static, the information contained in the sonel map does not

Bandwidth	f_l (Hz)	f_u (Hz)	f_c (Hz)
1	44	89	63
2	89	177	125
3	177	354	250
4	354	707	500
5	707	1400	1000
6	1400	2800	2000
7	2800	5600	4000
8	5600	11300	8000

Table 4.1: Lower (f_l), upper (f_u) and center (f_c) frequencies considered by the sonel mapping algorithm [122].

need to be updated and therefore, to account for the changing soundfield observed at the receiver when the receiver is moved through the environment, only the acoustical rendering stage needs to be re-computed.

4.3 Stage One: The Sonel Tracing Stage

4.3.1 Sound Sources and Emission

Following the common approach in architectural acoustics applications the distribution of sound frequency in a given sound source is approximated by a fixed number N_{freq} of frequency bands [122]. Table 4.1 provides a summary of the frequency bands commonly used in acoustical applications and which is used in this work. A sound source is specified by its energy distribution function over a number of fixed frequency bands (N_{freq}) and each frequency band is considered separately.

For each sound source, a pre-determined number of sonels (N_{sonel}) per fre-

quency band are emitted and traced through the environment. The minimum number of sonels N_{min} (e.g., $N_{sonel} \geq N_{min}$) that must be emitted given a receiver represented by sphere of radius r_k is [186]

$$N_{min} = \frac{4(v_s t_{max})^2}{r_k^2} \tag{4.1}$$

where, t_{max} is the duration of the echogram and $v_s = 343m \cdot s^{-1}$ is the speed of sound in air. Each emitted sonel propagates a portion of the sound source energy. Accuracy of the estimation increases as more sonels are emitted. However, increasing the number of emitted sonels leads to a corresponding increase in computation time. Given an omni-directional sound source with a power level of L_s dB (for a particular frequency band), the energy E_{sonel} ($\mathrm{W \cdot m^{-2}}$) of each sonel when emitted from the sound source is determined as [200]

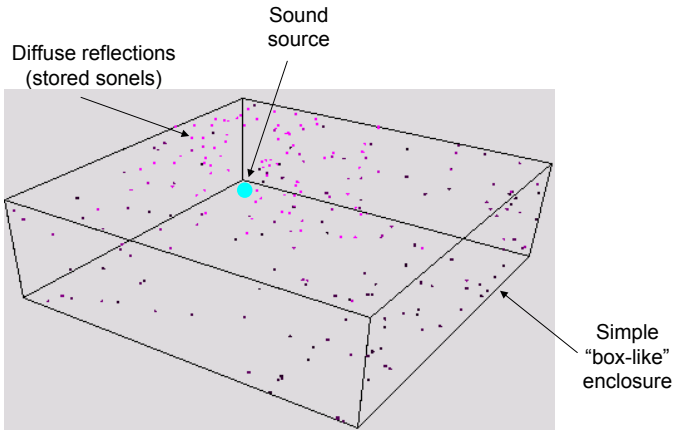
$$E_{sonel} = \frac{10^{L/10}}{N_{sonel}} \times 10^{-12}. \tag{4.2}$$

4.3.2 Details of the The Sonel Map Data Structure

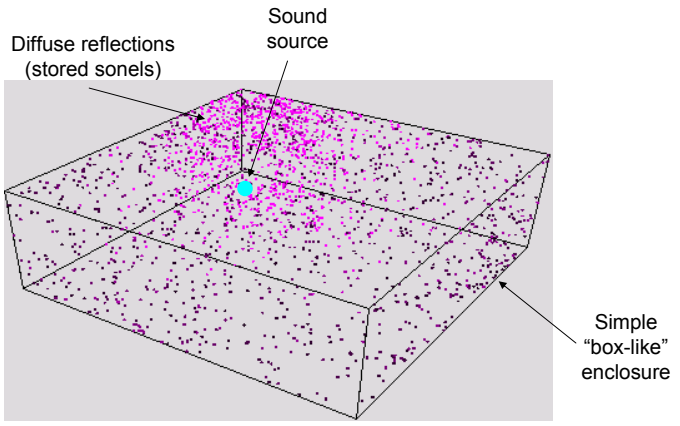
The sonel map is the acoustical analogue to the photon map used in photon mapping. The purpose of the sonel map is to store any sonels that have been reflected diffusely so that they can be used in the acoustical rendering stage to provide an estimate of the statistics of the sound energy contained at some point within the model. Figure 4.4 provides a graphical illustration of the

sonels contained in the model after the sonel tracing stage has completed. Reflection order is coded by color whereby darker colored points represent a higher reflection order. As with photon mapping, the energy estimate is determined by considering the nearest sonels at a point of interest and therefore, in order for the sonel map to be practical, the data structure used to implement the sonel map must be fast with respect to locating nearest neighbors in a three-dimensional point set [84]. The *kd-tree* is a multi-dimensional binary search tree whereby each node in the tree is used to partition one of the dimensions [20, 21] and as with the photon map, is used to implement the sonel map. Each node in the kd-tree contains one sonel in addition to pointers to its left and right subtrees. Each non-leaf node has one axis-orthogonal plane that stores the sonel and also cuts one of the dimensions (x,y or z) into two pieces. All sonels in the left subtree are below this plane and all sonels in the right subtree are above the plane. The kd-tree allows for the localization of sonels with a total of N sonels, in O(N) running time in the worst case and O(log N) time when the tree is balanced [84]. Furthermore, on average, it has been demonstrated that the time taken to locate k nearest sonels in the kd-tree is O(k + logN) [21, 84] making it an efficient and practical data structure to implement the sonel map.

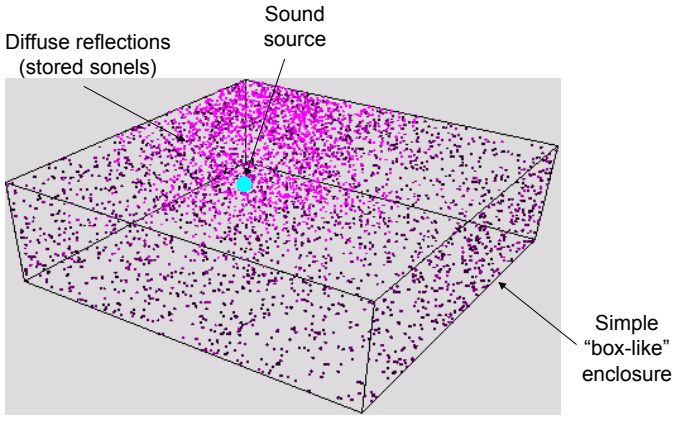
Once all the sonels in the model have been emitted from the sound source and the sonel map has been populated, the sonel map is balanced thus ensur-



(a) 100 sonels.



(b) 1000 sonels.



(c) 2000 sonels.

Figure 4.4: Stored sonels in the model example. Points in the model corresponding to a stored sonel due to a diffuse reflection ((a) 100 sonels, (b) 1000 sonels and (c) 2000 sonels). The surface absorption coefficient was set to 0.3 and the diffuse reflection coefficient to 0.7. Distance traveled (and therefore sonel energy) is coded by color whereby darker colors indicate a greater distance traveled by the sonel and therefore less energy.

ing that O(log N) searches occur. The balancing operation for a sonel map with N sonels has an O(NlogN) running time and in practise may take a few seconds to perform [84]. Provided the sound source remains static in the environment, this cost is incurred once only during the start of the simulation. Any movement of the receiver does not affect the sonel map in any manner and therefore, there is no need to update or re-compute the sonel map when only the receiver's position changes.

The purpose of the sonel tracing stage is to "populate" the *sonel map*. The sonel map records sonels which reflect off of surfaces diffusely. Upon encountering a surface for which it is determined that the sonel is reflected diffusely, the sonel will be stored in the sonel map while a "new" sonel is generated and reflected diffusely. Although specularly reflected and diffracted sonels are not stored, when the interaction between a sonel and a surface is either specular reflection or diffraction the sonel interaction still occurs (e.g., the sonel is reflected specularly or diffracted). Figure 4.5 provides a graphical overview of the processes defining the sonel tracing stage. This section provides greater details regarding each of these processes.

4.3.3 Sonel-Surface Interactions

The type of sonel-surface interaction differs depending on whether the sonel strikes the "surface" within the diffraction or non-diffraction zone (see Figure

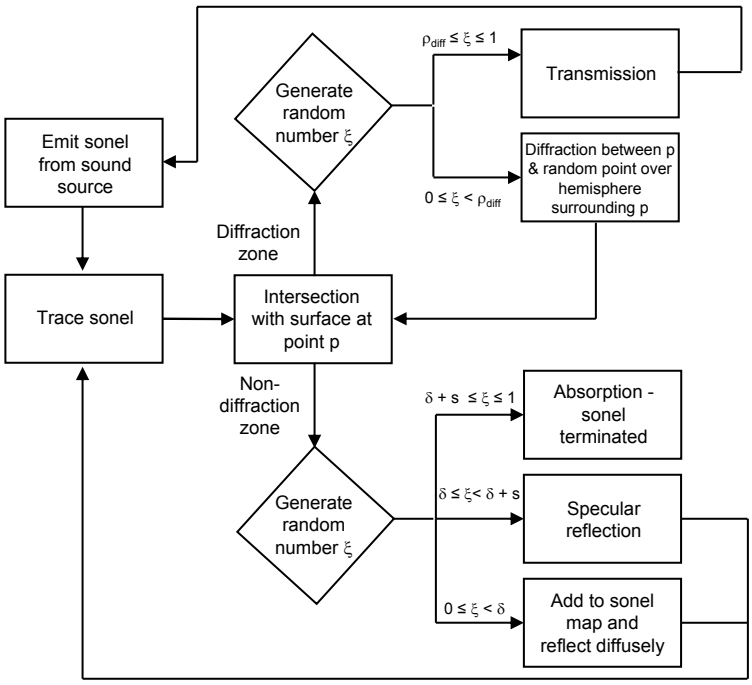


Figure 4.5: Stage one: the sonel tracing stage. The purpose of the sonel tracing stage is to populate the sonel map. Each sonel emitted from the sound source is traced through the environment. Upon encountering a surface at a position p, one of five interactions will occur. Initially, a check is made to determine whether the incident sonel lies within the diffraction or non-diffraction zone. When the sonel is incident within the non-diffraction zone, it will be either reflected specularly, reflected diffusely or completely absorbed by the surface based on the outcome of a randomly generated number ξ and the surface characteristics (e.g., diffuse coefficient δ, specular coefficient s and the absorption coefficient α). When the interaction is a diffuse reflection, the sonel is stored in the sonel map and is reflected by assuming ideal (or Lambertian) reflection. When the interaction is specular, the sonel is reflected assuming an ideal specular reflection. When the interaction is absorption, the sonel is terminated. Specularly reflected sonels are not stored in the sonel map. When the sonel incidence position is within the diffraction zone, the sonel is reflected by choosing a random direction over the hemisphere about interaction point p or is transmitted unaltered through the diffraction zone. Whether the sonel is diffracted or transmitted is determined based on the outcome of a randomly generated number ξ and the characteristics of both the surface (size or surface dimensions) and the sonel (frequency).

4.16 of Section 4.5). When the sonel falls within the diffraction zone it is either diffracted or it will experience transmission. When the sonel falls within the non-diffraction zone, it is either reflected specularly, reflected diffusely or completely absorbed by the surface. Which of these three interactions occurs is determined using a Russian roulette strategy.

The decision on which interaction actually occurs is collectively decided based upon the value of a uniformly distributed random number $\xi \varepsilon [0 \ldots 1]$ as follows

$$\begin{aligned} \xi \in [0 \ldots \delta] &\rightarrow \text{diffuse reflection} \\ \xi \in (\delta \ldots \delta + s)] &\rightarrow \text{specular reflection} \\ \xi \in (\delta + s \ldots 1] &\rightarrow \text{absorption} \end{aligned}$$

where, s and δ are the frequency dependent specular and diffuse surface coefficients respectively. In the event of a diffuse reflection, (e.g., $\xi \in [0 \ldots \delta]$) the sonel is stored in the sonel map and a new sonel is created and reflected diffusely from the sonel/surface intersection point p assuming ideal (or Lambertian) reflection. When the reflection is specular, (e.g., $\xi \in (\delta \ldots \delta + s)]$), a new sonel is created and reflected specularly assuming ideal specular reflection. If a deterministic termination criteria whereby the sonel's reflection count or energy dictates when the sonel is terminated, where being used each time

a sonel is reflected either specularly or diffusely, its energy would be attenuated to account for absorption by the surface. In contrast, with a Russian roulette termination criterion the sonel's energy is not attenuated to account for surface absorption when it is reflected. Absorption with a Russian roulette strategy is handled when the sonel is completely absorbed at the surface (e.g., $\xi \in (\delta + s \ldots 1]$). If the sonel is absorbed, tracing of the incident sonel is terminated.

4.3.3.1 Specular Reflection

When a sonel is reflected specularly, the sonel is reflected assuming ideal specular reflection whereby the angle of reflection is equal to the angle of incidence (see Figure 2.7 of Section 2.2.1). Given the surface normal $\vec{n}$ and incoming direction ω_i, the ideal specular reflection direction ω_s is given as

$$\omega_s = 2(\omega_i \cdot \vec{n})\vec{n} - \omega_i. \tag{4.3}$$

Prior to reflecting the sonel, its relevant parameters are updated to account for the intersection with the surface at point p. This includes adding the distance between the last intersection point and the sonel's current intersection point to the total distance traveled by the sonel and updating the current point of intersection (e.g., the incidence point). Example simulations where a number of sonels (one, 10 and 100) are emitted from the sound source and traced through

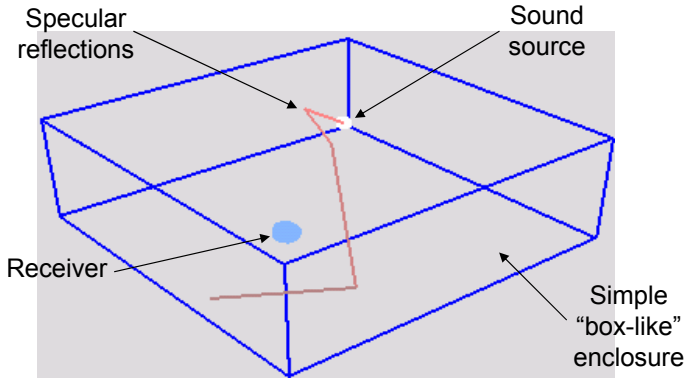
a simple box-like enclosure assuming specular reflections only are shown in Figure 4.6. In these examples, reflection order is coded by ray color whereby darker colors indicate higher reflection orders.

4.3.3.2 Diffuse Reflection

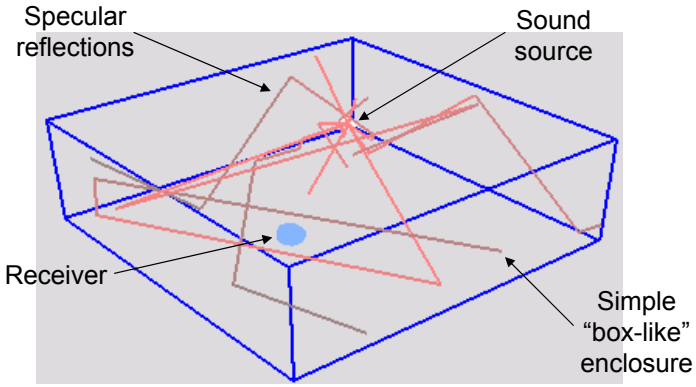
When a sonel is reflected diffusely, the information contained in the incident sonel is updated (e.g., the point of interaction on the surface p and distance traveled) and the sonel is stored in the sonel map (see Section 4.3.2 for further details regarding the sonel map). A new sonel is then generated and reflected assuming ideal (or *Lambertian*) diffuse reflection whereby the reflected direction is perfectly random [84] over the hemisphere surrounding p with a probability proportional to the cosine angle with the surface normal [175]. Given two uniformly distributed random numbers $\xi_1 \in [0 \ldots 1]$ and $\xi_2 \in [0 \ldots 1]$, the randomly reflected direction ω_d is given as [84]

$$\omega_d = (\theta,\ \phi) = cos^{-1}(2\pi\xi_2,\ \sqrt{\xi_1}) \tag{4.4}$$

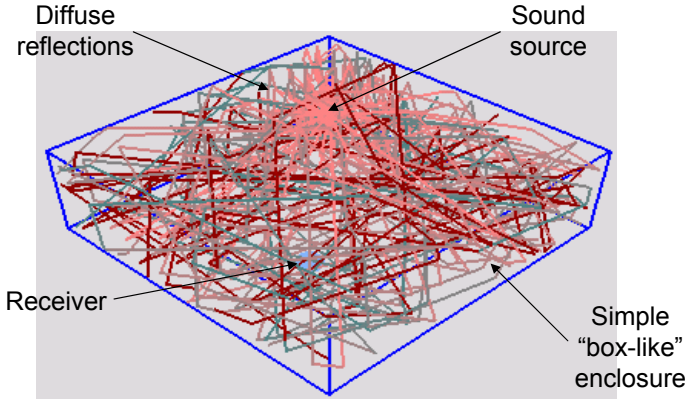
where, $(\theta,\ \phi)$ are spherical coordinates (ϕ is the angle with the surface normal and θ is the rotation around the normal). Example simulations where a number of sonels (one, 10 and 100) are emitted from the sound source and traced through a simple box-like enclosure assuming diffuse reflections only are shown in Figure 4.7. As with the specular reflection examples, darker ray



(a) One sonel.



(b) 10 sonels.



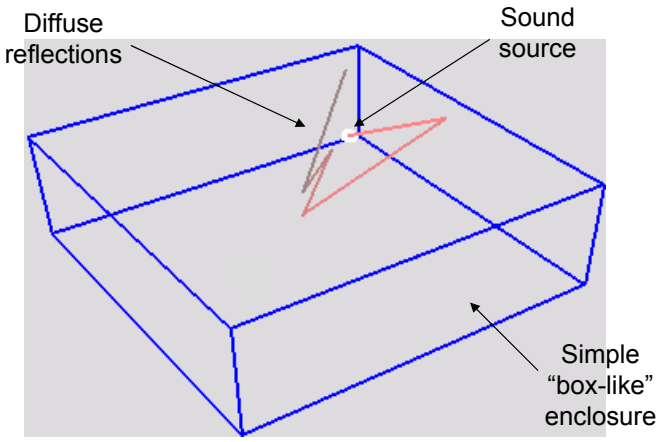
(c) 100 sonels.

Figure 4.6: Specular reflection example. Emitting sonels from the sound source ((a) one sonel, (b) 10 sonels and (c) 100 sonels) and tracing them through the environment assuming specular reflections only. Sonels are terminated using a Russian roulette termination criterion. Reflection order is coded by ray color whereby darker colors indicate higher reflection orders.

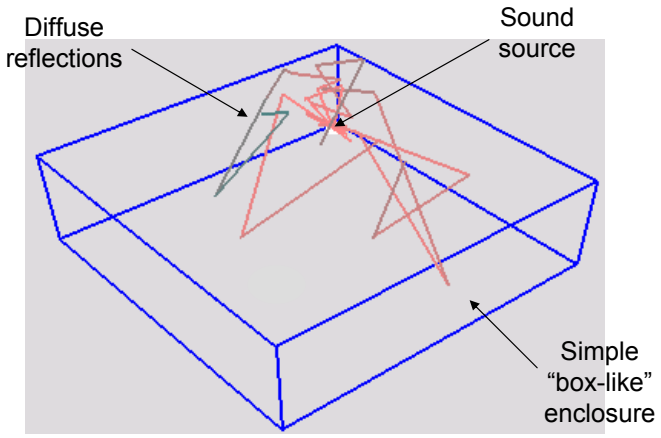
colors indicate higher reflection orders.

4.3.3.3 Diffraction

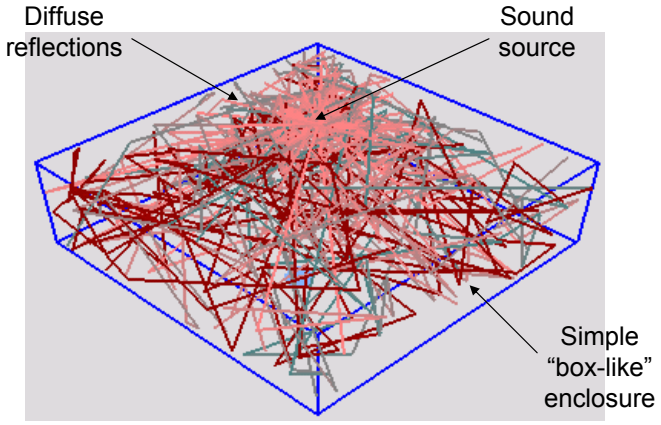
Being within the diffraction zone, the sonel incident position may not necessarily fall on the edge itself. However, the sonel is projected perpendicularly on to the edge. The projected position on the edge is denoted by $p_{edge} = (x_{edge}, y_{edge}, z_{edge})$. Diffraction is modeled using a modified version of the Huygens-Fresnel principle off of the edge to which it is closest. The Huygens-Fresnel principle provides an expression for the amount of energy leaving a sound source and arriving at a receiver in the unoccluded scenario (e.g., no obstacles or obstructions in the path between the sound source and receiver). The modifications made to the original Huygens-Fresnel principle account for the scenario where there is an edge between the sound source and receiver (e.g., the path between the sound source and receiver is occluded). The amount of energy reaching the receiver from the sound source is determined as in the original formulation but the energy is scaled by a visibility factor that describes the fraction of the path between them that is occluded. Diffracted sonels are not stored in the sonel map. When the sonel/surface interaction is diffraction, the sonel is reflected in a random direction (θ, ϕ) over the hemisphere surrounding point p_{edge} (see Figure 4.8). The receiver can be positioned anywhere relative to the sound source. Although no energy calculations between a sound source



(a) One sonel.



(b) 10 sonels.



(c) 100 sonels.

Figure 4.7: Diffuse reflection example. Emitting sonels from the sound source ((a) one sonel, (b) 10 sonels and (c) 100 sonels) and tracing them through the environment assuming diffuse reflections only. Sonels are terminated using a Russian roulette termination criterion. Reflection order is coded by ray color whereby darker colors indicate higher reflection orders.

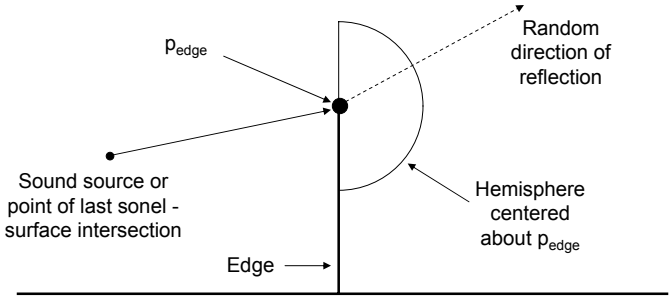


Figure 4.8: Handling diffraction in the sonel tracing stage. In the sonel tracing stage, when the sonel/surface interaction is diffraction, the sonel is reflected in a random direction (θ, ϕ) over the hemisphere surrounding point p_{edge}.

and receiver are made, reflecting the sonel in the hemisphere surrounding p_{edge} accounts for the fact that the receiver can be positioned anywhere relative to the sound source provided the edge separates them.

4.4 Stage Two: The Acoustical Rendering Stage

A graphical summary of the acoustical rendering stage is provided in Figure 4.9. The purpose of the acoustical rendering stage is to estimate the frequency dependent echogram (echogram for each of the frequency bands) through the use of the previously constructed sonel map coupled with distribution (Monte-Carlo) ray tracing. A number (N_{rays}) of frequency dependent acoustical visibility rays are traced from each receiver into the scene. In addition to estimating the diffusely reflected energy (using the sonel map), specularly reflected and diffracted energy are also estimated. As with the sonel tracing stage, any combination of specular reflections, diffuse reflections and diffraction are accounted

for in the acoustical rendering stage.

Each acoustical visibility ray contains information to describe the total distance it has traveled in addition to frequency. During each interaction (intersection) between an acoustical visibility ray and a position p on a surface, the distance between point p and the point p' from which the ray was last emitted/reflected/diffracted from is determined and this is added to the accumulating distance field of the ray (r_{ray}). When required by the simulation (e.g., when an acoustical visibility ray needs to be added to the accumulating echogram), the appropriate frequency dependent echogram "bin" (or location) is determined as

$$b_{f_i} = \left\lfloor (t_{total} \times \frac{1}{t_{rir}}) + 0.5 \right\rfloor \tag{4.5}$$

where, b_{f_i} represents bin i of the echogram corresponding to frequency f, t_{rir} is the temporal resolution of the echogram (e.g., spacing between "time steps" or "bins") and typically a resolution of about 5-10ms is sufficient [101]. t_{total} is the total time required for a particular sound following a particular propagation path between the sound source and receiver and is determined by dividing the total path length (distance) r_{total} by the speed of sound $v_s = 343m \cdot s^{-1}$

$$t_{total} = \frac{r_{total}}{v_s} = \frac{r_{total}}{343m \cdot s^{-1}}. \tag{4.6}$$

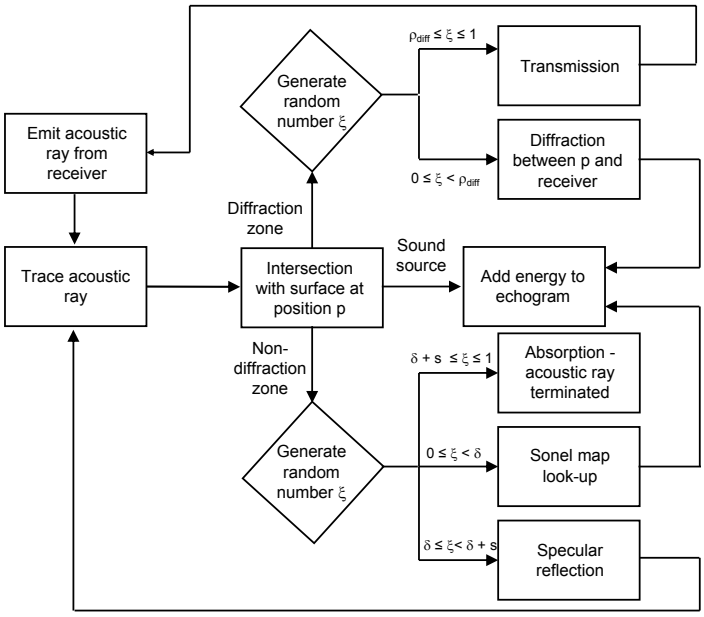


Figure 4.9: Stage two: the acoustical rendering stage. The purpose of the acoustical rendering stage is to estimate the frequency dependent room echogram through the use of the previously constructed sonel map coupled with distribution ray tracing to determine specular reflections and diffraction. Frequency dependent acoustical visibility rays are traced from the receiver through the scene while they interact with any surfaces/objects they may encounter. Upon encountering a sound source, a portion of the sound source's energy (the total sound energy divided by the total number of acoustical visibility rays traced from the receiver), is added to the accumulating echogram after being scaled to account for absorption by the medium. When a sound source is encountered and the acoustical visibility ray's order is zero, this represents direct sound energy. As with the sonel tracing stage, upon encountering a surface (provided the surface is not a sound source), a check is made to determine whether the incident acoustical visibility ray lies within the diffraction or non-diffraction zone. When the ray lies within the non-diffraction zone, it will be either reflected specularly, reflected diffusely or completely absorbed by the surface based on the outcome of a randomly generated number and the surface characteristics (e.g., diffuse coefficient δ, specular coefficient s and the absorption coefficient α) using a Russian roulette approach as done in the sonel tracing stage. When the interaction is diffuse reflection, the acoustical visibility ray is terminated and the sonel map is used to provide an estimate of the sound energy leaving the intersection point p and arriving at the receiver using a *density estimation* algorithm. Specular reflections and absorption are handled in a manner similar to the sonel tracing stage (e.g., reflected assuming ideal reflection and completely absorbed respectively). When the acoustical visibility ray incidence position is within the diffraction zone, the ray will either be diffracted using a modified version of the Huygens-Fresnel principle or it will be transmitted through the surface unaltered. Whether the acoustical visibility ray is diffracted or transmitted is determined based on the outcome of a randomly generated number and the characteristics of both the surface (size or surface dimensions) and the acoustical visibility ray (frequency).

Relative Humidity	63Hz	125Hz	250Hz	500Hz	1kHz	2kHz	4kHz
40%	0.013	0.05	0.09	0.6	1.07	2.58	5.05
50%	0.009	0.045	0.08	0.63	1.08	2.28	4.2
60%	0.008	0.025	0.06	0.64	1.11	2.14	3.72
70%	0.006	0.02	0.05	0.64	1.15	2.08	3.45

Table 4.2: Air attenuation coefficient (m) as a function of relative humidity and frequency for a temperature of 20°C and normal atmospheric pressure of 10^{-3}m^{-1}.

Prior to adding any energy to the echogram, the energy is scaled to account for attenuation of acoustical energy by the medium (air). Assuming planar sound waves, the attenuation of sound energy due to absorption by the air follows an exponential law [100]

$$E_r = E_\circ e^{-mr} \tag{4.7}$$

where, $E_\circ$ is the original sound energy, E_r is the energy after the sound has traveled a distance r and m is the air absorption coefficient that varies as a function of the conditions of the air itself (e.g., temperature, frequency, humidity and atmospheric pressure). Expressions for the evaluation of m are provided by Bass et al. [9] and a list of values for m assuming an air temperature of 20°C and normal atmospheric pressure of 10^{-3}m^{-1} for several relative humidity levels are provided in Table 4.2.

4.4.1 Acoustical Visibility Ray-Surface Interactions

Interactions between an acoustical visibility ray and any objects/surfaces it may encounter are handled in a manner similar to the sonel tracing stage.

Upon encountering a surface, a check is made to determine whether the ray/surface incidence point is within the frequency dependent diffraction or non-diffraction zone. When the ray is within the non-diffraction zone, as in the sonel tracing stage, a Russian roulette strategy is used to determine whether the acoustical visibility ray is reflected specularly, diffusely or completely absorbed. When the ray is incident within the diffraction zone it will be either diffracted off of the edge to which it is closest or transmitted through the dilated portion of the surface unaltered.

If an acoustical visibility ray encounters a sound source (represented by a sphere for the purposes of this stage), the energy propagating from the sound source to the receiver represented by this particular ray path is scaled to account for attenuation by the medium using Equation 4.7 with r_{total} assigned the total distance propagated by the acoustical visibility ray. Once the energy has been scaled to account for attenuation by the medium (air), it is added to the accumulating echogram using Equation 4.5 with t_{total} assigned the value equal to the time taken for the ray emitted by the receiver to reach the sound source.

4.4.1.1 Direct Sound

Direct sound is not explicitly accounted for (e.g., by explicitly checking whether the path between the sound source and the receiver is obstructed). Rather, direct sound reaching the receiver is measured probabilistically during the

acoustical rendering stage. In particular, upon encountering a surface, a check is made to determine whether the surface belongs to a sound source. If the ray encounters a sound source directly prior to encountering any other surface (e.g., reflection order of zero), then this represents direct sound and a portion of the sound source's energy (the total sound energy divided by the total number of acoustical visibility rays traced from the receiver), is scaled to account for attenuation by the air using Equation 4.7 with r_{total} assigned the value of the distance between the sound source and receiver)

$$r = \frac{\sqrt{(x_S - x_R)^2 + (y_S - y_R)^2 + (y_S - y_R)^2}}{343m \cdot s^{-1}} \tag{4.8}$$

where, (x_S, y_S, z_S) are the coordinates of the sound source and (x_R, y_R, z_R) are the coordinates of the receiver. The scaled energy is then added to the accumulating echogram using Equation 4.5 with t_{total} assigned the value of the time taken for the ray emitted by the receiver to reach the sound source

$$\begin{aligned} t_{total} &= \frac{\sqrt{(x_S - x_R)^2 + (y_S - y_R)^2 + (y_S - y_R)^2}}{v_s} \\ &= \frac{\sqrt{(x_S - x_R)^2 + (y_S - y_R)^2 + (y_S - y_R)^2}}{343m \cdot s^{-1}}. \end{aligned} \tag{4.9}$$

4.4.1.2 Specular Reflection

Specular reflections are handled similarly to the sonel tracing stage. When the interaction at the ray/surface interaction point is determined to be specular reflection, the acoustical visibility ray is reflected such that the angle of the reflected ray equals the angle of the incidence ray with respect to the surface normal (see Section 4.3.3.1).

4.4.1.3 Diffuse Reflection

When the interaction between the acoustical visibility ray and the surface at the intersection point p is diffuse reflection, the acoustical visibility ray is terminated and the sonel map is used to provide an estimate of the sound energy leaving point p. A *nearest neighbor density estimation* algorithm [170] is used to determine the diffuse energy component. An estimate of the energy at point p is made by averaging the energy of the n nearest sonels neighboring point p that are stored in the sonel map. This involves searching through the kd-tree that implements the sonel map for the n sonels that are located within a circle of radius r_s centered about the incidence point p on the surface. Although a search for n sonels is made, there may be less than n sonels that are within a distance of r to point p and therefore the estimate may be made with less than n sonels (denoted by n_{actual}).

In reality, not all sonels matching this criteria will necessarily reach the

receiver however, currently it is assumed all sonels do. The total path length (r_{total}) is equal to the total distance traveled by the acoustical visibility ray (denoted by r_{ray}) in addition to the total distance previously traveled by the sonel r_{sonel} (e.g., $r_{total} = r_{ray} + r_{sonel}$). The energy of each sonel is scaled by $1 \cdot n_{actual}$ and further scaled to account for attenuation by the air using Equation 4.7 (with $r = r_{total}$). The scaled energy of each of the n_{actual} sonels is then added to the appropriate "bin" b_i of the accumulating echogram using Equation 4.5 with t_{total} equal to the sum between the time for the acoustical visibility ray emitted at the receiver to reach point p and the total sonel propagation time

$$t_{total} = \frac{(r_{ray} + r_{sonel})}{v_s} = \frac{(r_{ray} + r_{sonel})}{343m \cdot s^{-1}}. \tag{4.10}$$

The sonel is then terminated.

4.4.1.4 Diffraction

Being within the diffraction zone, the acoustical visibility ray incident position may not necessarily fall on the edge itself. However, the incidence position is projected perpendicularly on to the edge. The projected position on the edge is denoted by $p_{edge} = (x_{edge}, y_{edge}, z_{edge})$. Diffraction of an acoustical visibility ray incident within the diffraction zone is modeled off of the edge to which it is closest.

A Summary of the Huygens-Fresnel Principle: The Huygens-Fresnel principle states that every point on the primary wavefront can be thought of as a continuous, direction dependent emitter of secondary wavelets (sources) that combine to produce a new wavefront in the direction of propagation [198]. These secondary wavelets are emitted in a direction dependent manner, essentially scaled by an *obliquity* or *inclination* factor $K(\theta)$ as follows [79]

$$K(\theta) = \frac{1}{2}(1+\cos(\theta)) \tag{4.11}$$

where, θ is the angle between the receiver and the direction of propagation of the primary wavefront. This expanding wavefront can be divided into a number of ring-like regions, collectively known as *Fresnel zones* [79]. The boundary of the i^{th} Fresnel zone (Z_i) corresponds to the intersection of the wavefront with a sphere of radius $r_\circ + i\lambda/2$ centered at the receiver where, $r_\circ$ is equal to the distance between the receiver and the expanding wavefront after it has traversed a distance of ρ from the sound source. In other words, the distance from the receiver to each adjacent zone differs by half a wavelength ($\lambda/2$). Each Fresnel zone contains a number of secondary sources that are assumed to emit their energy in phase with the primary wave. The secondary sources within each Fresnel zone i contain energy collectively equal to E_i and a portion of this energy may reach the receiver. By summing the contribution of energy reaching the receiver from each of the Fresnel zones, the total energy leaving the sound

source and reaching the receiver (E_{total}) can be calculated. Alternatively, it can easily be shown that the total energy is equal to half the energy of the first Fresnel zone reaching the receiver [79] or, mathematically

$$E_{total} \approx \frac{|E_1|}{2}. \tag{4.12}$$

Greater details regarding the Huygens-Fresnel principle, including a detailed mathematical derivation, are provided in Appendix C.

Acoustical Diffraction Using the Huygens-Fresnel Principle: Since acoustical visibility rays are propagating and interacting with objects/surfaces in the environment, an acoustical visibility ray may encounter an edge after being reflected off of a surface as opposed to coming directly from the receiver (acting as a sound source for the purpose of this stage). Hence, when referring to a receiver, this may be an actual receiver or the last interaction point between the acoustical visibility ray and the surface if the acoustical visibility ray has encountered another non-diffracting surface prior to encountering the edge. Once it is determined that an acoustical visibility ray is to be diffracted (e.g., when it falls incident within the diffraction zone of a surface), the acoustical diffraction technique to be described here utilizes the Huygens-Fresnel principle as described in the previous section to determine the acoustical energy reaching a receiver from a given sound source after being diffracted by an edge.

Essentially, given a sound source, receiver and edge, the energy reaching the receiver is determined by dividing the initial wavefront emitted by the sound source into a number of Fresnel zones and considering the energy arriving at the receiver from the first Fresnel zone as in the unoccluded scenario described in the previous section and in Appendix C. To account for diffraction effects, a visibility factor for the first Fresnel zone is introduced. The visibility factor (denoted by v_1) represents that fraction of the first zone that is visible relative to the receiver.

In order to determine the energy arriving at the receiver from the first Fresnel zone, the position of one of the secondary sources within the first Fresnel zone is required. Once the position of one of the secondary sources is known, the obliquity factor $K(\theta)$ (that describes the direction dependent secondary source propagation from the sound source) and as will be described, the visibility of the zone can then be determined. The radius ρ of the initial wavefront (e.g., the radius of the wavefront emitted from the sound source after it has propagated for some time t) is set to the distance between the position of the sound source (e.g., $S = (x_S, y_S, z_S)$) and the position on the edge ($p_{edge} = (x_{edge}, y_{edge}, z_{edge})$)

$$\rho = \sqrt{(x_{edge} - x_S)^2 - (y_{edge} - y_S)^2 - (z_{edge} - z_S)^2}.$$

Since ρ is the distance between the position of the sound source and p_{edge}, p_{edge}

must be located on the surface of the wavefront and within one of the Fresnel zones (the "initial" Fresnel zone) denoted by Z_{init} and calculated as follows

$$Z_{init} = \left\lfloor \frac{r_{init} - r_{\circ}}{\lambda} + 0.5 \right\rfloor \tag{4.13}$$

where, r_{init} is the distance between the receiver and p_{edge} and $r_{\circ}$ is the distance between the receiver and primary wavefront, given as

$$r_{\circ} = r_{SR} - \rho \tag{4.14}$$

where, r_{SR} is the distance between the sound source and the receiver. Although p_{edge} may lie anywhere within Z_{init} and not necessarily on its boundary, it is assumed that the obliquity factor is constant throughout the entire zone [79] and therefore, its position within the zone does not matter. Given the position of the secondary source in Z_{init}, the position of a secondary source within the first Fresnel zone (Z_1) can be determined. This is accomplished in two steps (see Figure 4.10):

1. Rotate p_{edge} such that it lies directly on the (imaginary) line between the sound source and receiver thus essentially moving p_{edge} to a new position denoted by $p_{\circ}$.
2. Move $p_{\circ}$ to yet another new position (p_1) within the first Fresnel zone.

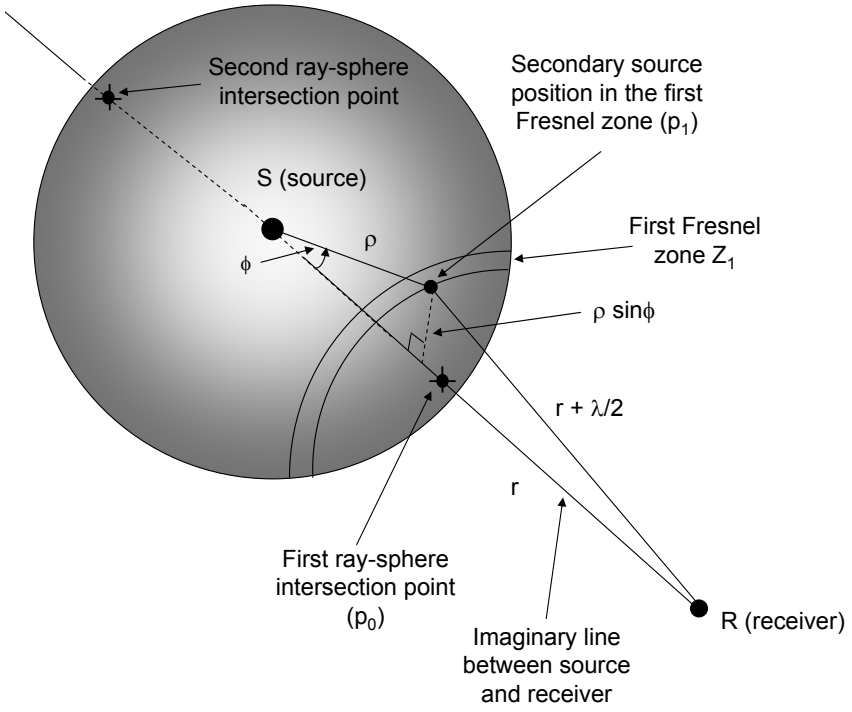


Figure 4.10: Determining the position of a secondary source within the first Fresnel zone.

Determining $p_\circ$: The original position p_{edge} that lies within Z_{init} is rotated such that it lies directly on the (imaginary) line between the sound source and receiver. Although this can be performed using a series of rotations about the central axes, this is actually accomplished by taking the (first) point of intersection (denoted by $p_\circ$) between the sphere representing the initial wavefront and a ray (normalized vector) whose origin is the receiver position and whose direction is towards the sound source. The procedure for determining p_1 is summarized below.

Determining p_1: Once the intersection point ($p_\circ$) has been determined, it is moved to the first Fresnel zone. Referring to Figures 4.10 and 4.11, angles θ (the horizontal angle of $p_\circ$ relative to the sound source) and ϕ (the vertical angle of $p_\circ$ relative to the sound source) are initially equal to zero and are calculated as follows

$$\theta = tan^{-1}\left(\frac{x_p - x_S}{z_p - z_S}\right), \quad \phi = cos^{-1}(\frac{y_p - y_S}{\rho}) \tag{4.15}$$

where, (x_S, y_S, z_S) and (x_p, y_p, z_p) are the spatial coordinates of the sound source and $p_\circ$ respectively. As with $p_\circ$, the position of the secondary source in the first Fresnel zone (p_1) also lies on the surface of the sphere corresponding to the initial propagating wavefront of radius ρ. The difference in distance between adjacent Fresnel zones and the receiver is $\lambda/2$. Therefore, the difference in

Algorithm 1 Procedure for determining p_1.

1: Set $\phi = 0$.
2: Initialize $d\phi$ to a small value.
3: Sample the "new" (updated) position $p_{new} = (x_{new}, y_{new}, z_{new})$ using the equations for the Cartesian coordinates of a sphere as given below:

$$\begin{aligned} x_{new} &= x_S + (\rho sin(\theta) sin(\phi)) \\ y_{new} &= y_S + (\rho sin(\theta) cos(\phi)) \\ z_{new} &= z_S + (\rho cos(\theta)) \end{aligned}$$

4: Compute the distance between the receiver and p_{new}.
5: Compute r_{diff} (the difference in distance between r_{new} and $r_\circ$).
6: **if** $(\lambda/2 - \varepsilon) \leq r_{diff} \leq (\lambda/2 + \in)$ **then**
7: Stop
8: **else**
9: Increment ϕ by a small amount $d\phi$ (e.g., $\phi = \phi + d\phi$) and go back to step 3.
10: **end if**

distance (r_{diff}) between the receiver and $p_\circ$ and the receiver and the secondary source in the first Fresnel zone p_1 must also be $\lambda/2$. Position p_1 is determined iteratively until r_{diff} is within ε of $\lambda/2$ or, mathematically,

$$(\lambda/2 - \varepsilon) \leq r_{diff} \leq (\lambda/2 + \varepsilon) \tag{4.16}$$

where, ε is an arbitrarily chosen "small" value. This procedure is summarized by Algorithm 1.

Once the position of a secondary source within the first Fresnel zone (Z_1) has been determined, the energy reaching the receiver from Z_1 can be calculated using Equation C.11 of Appendix C, which is reproduced below

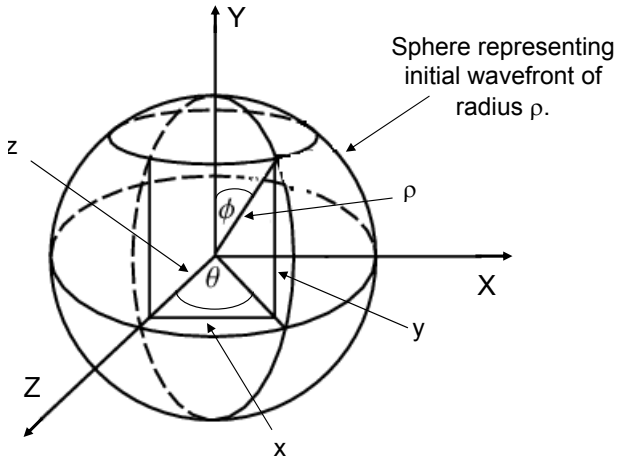


Figure 4.11: Spherical coordinates.

$$\begin{aligned} E_1 &= (-1)^{1+1}\frac{2K_1(\theta)E_A\rho\lambda}{(\rho+r_\circ)}\sin[\omega t - k(\rho+r_\circ)] \\ &= \frac{2K_1(\theta)E_A\rho\lambda}{(\rho+r_\circ)}\sin[\omega t - k(\rho+r_\circ)] \end{aligned} \tag{4.17}$$

where, $K_1(\theta)$ is the obliquity factor of Z_1, $r_\circ$ is the distance between the receiver and the expanding wavefront after it has traversed a distance of ρ from the sound source. t is the time taken for the secondary source in Z_1 to reach the receiver, $k = 2\pi\lambda$ is the wave-number and E_A is the energy per unit area of the secondary sources within a differential area of the Fresnel zone (see Appendix C). However, to account for absorption of sound energy by the medium (air), Equation 4.17 is replaced by

$$E_1 = K_1(\theta) \times E_\circ e^{-m(\rho+r_\circ)} \times \sin[\omega t - k(\rho+r_\circ)] \tag{4.18}$$

where, $E_\circ$ is the ray energy and m is the air absorption constant that varies as a function of the conditions of the air itself. As presented above, the energy reaching the receiver from the first Fresnel zone can be calculated assuming an obstruction-free path between each zone and the receiver (e.g., the first zone is completely visible to the receiver). Edge effects are accounted for by considering the *visibility weighting* v_1 for the first zone Z_1 relative to the receiver, using ray casting. As shown in Figure 4.12, a number of rays (denoted by n_{rays}) are emitted from the receiver to uniformly sampled positions within Z_1 (the mathematical details regarding this procedure are provided in Section C.3 of Appendix C), and a check is made to determine whether there is a clear path (e.g., no objects or obstructions) for each ray to propagate from the receiver to the sampled position (e.g., checking whether the particular uniformly sampled position on the first zone is visible from the receiver). A graphical illustration is provided in Figure 4.13 where an edge is placed between a sound source and receiver, partially occluding the direct path between them. Sampled positions visible to the receiver are illustrated in Figure 4.13(a) while non-visible samples are illustrated in Figure 4.13(b). The visibility weighting is determined by considering the number of visible (obstruction-free) rays (n_{vis}) relative to the total number of emitted rays (N_{vis})

$$v_1 = \frac{n_{vis}}{N_{vis}}. \tag{4.19}$$

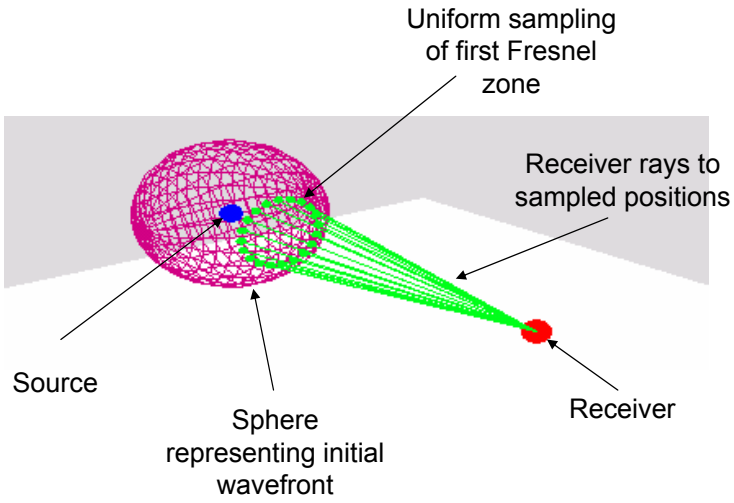


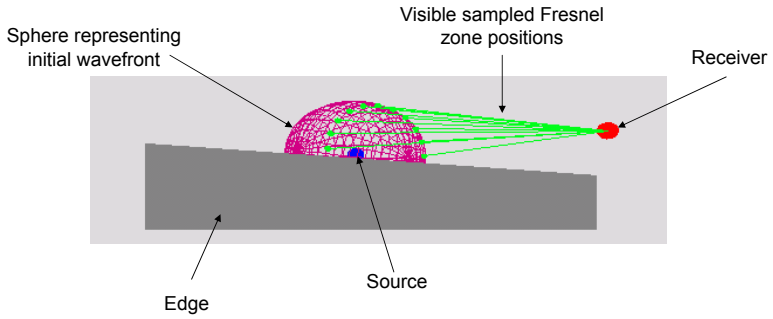
Figure 4.12: Rays sent from the receiver to uniformly sampled positions within the first Fresnel zone. In this example, sound source frequency is 63Hz. In this scenario, the path between each sampled position and the receiver is clear (e.g., unoccluded).

Taking edge effects into account, the total energy reaching the receiver from the first Fresnel zone Z_1 is given as

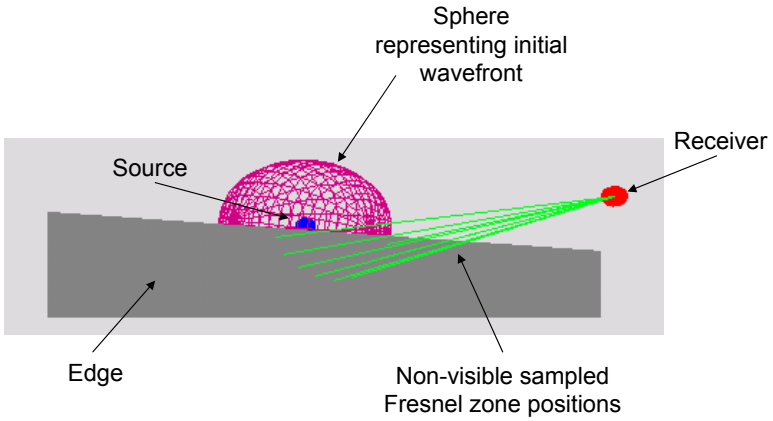
$$E_1 = v_1 \times K_1(\theta) \times E_\circ e^{-m(\rho + r_\circ)} \times \sin[\omega t - k(\rho + r_\circ)] \tag{4.20}$$

where, $t = (r_\circ + \lambda/2)/v_s$ is the time taken for the secondary sources within the first Fresnel zone to reach the receiver. Since all Fresnel zones are annular regions (rings) around the sphere representing the wavefront, the secondary sources within a particular Fresnel zone (including the first), will be equidistant to the receiver and will therefore reach the receiver at the same time t_1 given as

$$t_1 = \frac{r_1}{v_s} = \frac{r_1}{343m \cdot s^{-1}} \tag{4.21}$$



(a) Non-blocked acoustical visibility rays.



(b) Blocked acoustical visibility rays.

Figure 4.13: Sampling a Fresnel zone in the presence of an occluding edge. (a) The portion of the rays that are not blocked by the edge. (b) The portion of the rays that are blocked by the edge.

where, $v_s = 343\text{m}\cdot\text{s}^{-1}$ is the speed of sound in air and r_1 is the distance between the receiver and the secondary source in the first Fresnel zone.

Once E_1 has been determined, it is added to the appropriate "bin" b_i of the accumulating echogram using Equation 4.5 with t equal to the time taken for the energy to reach the receiver from the first Fresnel zone in addition to the time taken for the energy to reach the first Fresnel zone from the sound source.

Considering All Fresnel Zones: As described, acoustical diffraction modeling is accomplished using an approximation to the Huygens-Fresnel principle whereby the energy of first Fresnel zone is considered only. Rather than considering the first Fresnel zone only, the entire sphere representing the initial wavefront emitted from the sound source is divided into a number of Fresnel zones and the energy arriving at the receiver from each of these Fresnel zones is summed to determine the amount of energy reaching the receiver. The total number of zones (N_{zones}) is determined as follows

$$N_{zones} = \left\lfloor \frac{2\rho}{(\lambda/2)} \right\rfloor . \tag{4.22}$$

To account for diffraction effects, a visibility factor for each Fresnel zone is introduced. The visibility factor (denoted by v_i) represents the fraction of the i^{th} Fresnel zone visible from the receiver.

As with the first zone-only approximation previously described, the edge po-

sition p_{edge} is assumed to lie on the sphere representing the initial wavefront and within a particular Fresnel zone Z_{init}. Given the position of the secondary source in Z_{init}, the position of a secondary source within the first Fresnel zone (p_1) can be determined using the two step process previously described. Upon determining p_1, simple geometry allows for the position of a secondary source within zone Z_2 to be determined. The same reasoning can be applied to finding the position of a secondary source within the third Fresnel zone and subsequent zones until the position of a secondary source within all the Fresnel zones considered has been found (the mathematics describing this process are developed in Section C.2). Once the position of a secondary source within a zone Z_i has been determined, the energy reaching the receiver from Z_i can be calculated using Equation 4.18, which is reproduced below

$$E_i = (-1)^{i+1} \times K_i(\theta) \times E_\circ e^{-m(\rho + r_\circ)} \times \sin[\omega t - k(\rho + r_\circ)]. \tag{4.23}$$

Edge effects are accounted for by considering the visibility weighting v_i of each zone Z_i relative to the receiver, using ray casting. The visibility weighting of Fresnel zone i is determined by considering the number of visible (obstruction-free) rays n_{vis} relative to the total number of emitted rays N_{vis}

$$v_i = \frac{n_{vis}}{N_{vis}}. \tag{4.24}$$

Taking edge effects into account, the total energy reaching the receiver from zone Z_i is given as

$$E_i = v_i \times (-1)^{i+1} \times K_i(\theta) \times E_\circ e^{-m(\rho + r_\circ)} \times \sin[\omega t_i - k(\rho + r_\circ)] \tag{4.25}$$

where, $t_i = (r_\circ + i\lambda/2)/v_s$ is the time taken for the secondary sources within Fresnel zone i to reach the receiver. The total energy E_{total} reaching the receiver from the sound source taking edge effects into consideration is determined by summing the energy reaching the receiver from each of the N Fresnel zones

$$E_{total} = (v_1 \times E_1) + (v_2 \times E_2) + \cdots + (v_N \times E_N). \tag{4.26}$$

Using this approach as opposed to considering the first Fresnel zone only requires considerably more processing. In particular, the visibility factor v_i will have to be determined for each zone and this can become computationally expensive depending on the number of acoustical visibility rays emitted from the receiver. As previously mentioned, the sphere representing the initial wavefront is divided into a number of Fresnel zones such that the distance between adjacent zones differs by $\lambda/2$. Therefore, given the inverse relationship between frequency and wavelength, as frequency increases, the distance between adjacent Fresnel zones decreases thus resulting in a greater number of Fresnel zones. However, an increase in the number of Fresnel zones results

in a corresponding increase in the required computation time. Furthermore, as will be discussed in Section 5.1.4.5, this approach is more prone to numerical errors.

4.5 Implementation Details

For the current implementation and for all the simulations described in this dissertation, specular reflections are assumed to be ideal whereby the angle of reflection is equal to the angle of incidence (with respect to the surface normal). Diffuse reflections are assumed to be purely Lambertian and thus rather than have multiple reflections in arbitrary directions, a single reflection occurs and its direction is chosen uniformly over the hemisphere centered above the sonel/surface intersection point with a probability proportional to the cosine angle with respect to the surface normal. Diffraction is accounted for using a modified version of the Huygens-Fresnel principle. When absorption does occur, all of the sonel's energy is absorbed and not only a portion of it. Figure 4.14 illustrates the approximations made to handle the various interactions between a propagating sonel and an object/surface it encounters. Another simplification made is with respect to the distribution of sound frequency. In particular, frequency is handled by considering a fixed number of frequency bands. Each sonel is assigned one frequency of a particular frequency band and each frequency band is treated separately.

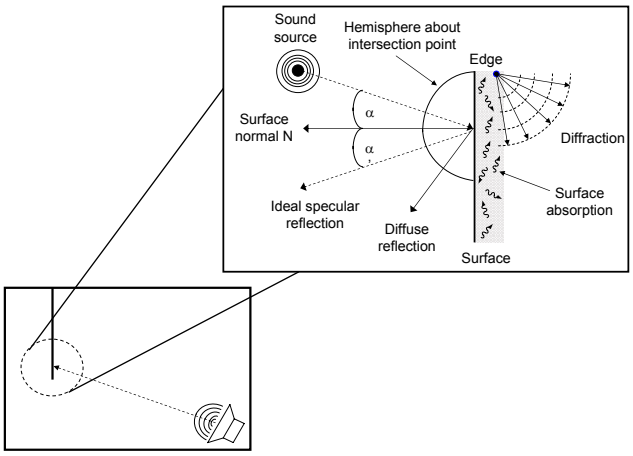


Figure 4.14: Approximating the propagation of sound waves in the environment along with their interaction with any objects/surfaces they may encounter.

The scene is modeled as a collection of planar, quadrilateral surfaces and each surface is represented by four vertices ($v_\circ$, v_1, v_2, v_3). Being a quadrilateral, each surface contains four edges (e.g., an edge between the vertices $v_\circ$ and v_1, v_1 and v_2, v_2 and v_3 and v_3 and $v_\circ$). A distinction is made between a *diffracting* and *non-diffracting* edge (specified in the scene description). A diffracting edge is an edge where a sonel can be diffracted off of it whereas a non-diffracting edge is an edge that is incident on to another surface and therefore, a sonel cannot be diffracted off of it (see Figure 4.15).

For the purpose of handling the modeling of acoustical diffraction, (Figure 4.16), each original surface is *dilated* by a sonel-frequency dependent amount equal to $\lambda/2$ (where, λ is the wavelength). The dilated surface is divided into two zones: i) the *diffraction zone* and ii) the *non-diffraction zone*. The region on the dilated surface within a distance of λ of the dilated surface edge comprises the diffraction zone and the remainder of the surface comprises the non-

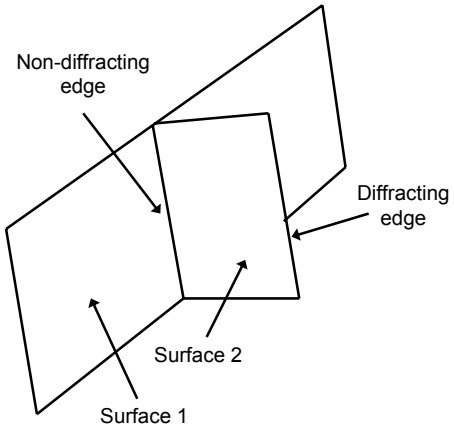


Figure 4.15: Diffracting and non-diffracting edges defined. A diffracting edge is an edge where a sonel can be diffracted off of. A non-diffracting edge is an edge that is completely incident on to another surface and therefore, a sonel cannot be diffracted off of it.

diffraction zone (Figure 4.16). The types of possible interaction experienced by the sonel will depend on which zone the sonel is incident upon. A sonel incident within the diffraction zone can either diffracted or transmitted though the surface. In the current implementation however, the probability of transmission is zero and therefore, when a sonel is incident within the diffraction zone it will always be diffracted. A sonel incident within the non-diffraction zone will be reflected either specularly or diffusely or will be absorbed by the surface.

The model description provides information specific to the sound sources, receivers (e.g., position in the environment, energy distribution functions etc.) and for handling interactions between the propagating energy particles (e.g., sonels) and any surface/object they may encounter.

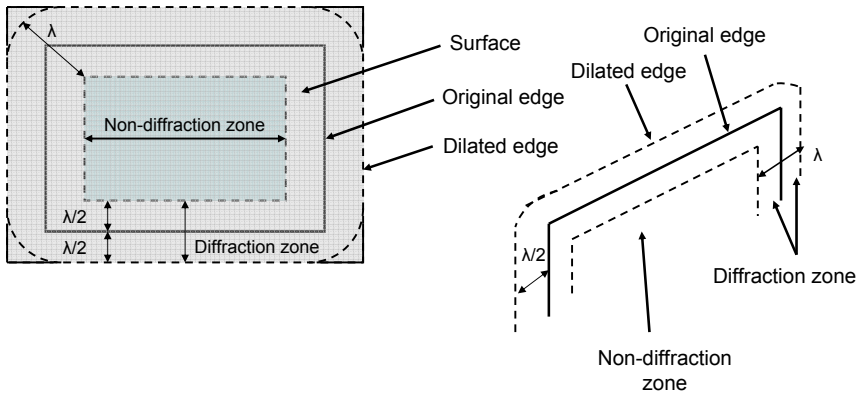


Figure 4.16: Defining the diffraction and non-diffraction zones. The diffraction zone is defined as the area within $\lambda/2$ of the original (non-dilated) surface edge while the remainder of the surface comprises the non-diffraction zone. The interaction between the sonel and surface is dependent upon which zone the sonel is incident upon.

4.6 Summary

Sonel mapping is a two-pass Monte-Carlo based approximation to recreating a sound field of a particular environment. It is a particle-based approach in which sound waves are approximated with particles that carry the relevant information necessary to simulate the propagation of sound waves emitted from a sound source propagating energy through the environment.

The first stage (the sonel tracing stage), constructs the sonel map. Sonels are emitted from each sound source and traced through the scene until they interact with a surface. Upon encountering a surface, a check is made to determine whether the sonel is within a diffraction or non-diffraction zone. When the sonel is incident within the non-diffraction zone, the sonel is either reflected specularly, reflected diffusely, or completely absorbed by the surface us-

ing a Russian roulette strategy. Diffusely reflected sonels are stored in the sonel map. When a sonel is incident within the diffraction zone, a new sonel is generated and emitted by choosing a random direction over the hemisphere centered about the diffraction point. The sonel map represents sound energy whose path contains a diffuse reflection.

In the second stage (the acoustical rendering stage), the frequency dependent echogram is estimated through the use of the previously constructed sonel map coupled with distribution ray tracing. The echogram at a receiver is determined using distribution ray tracing whereby acoustical visibility rays are traced from the receiver into the scene while recording their interaction with any surfaces/objects they may encounter. The direct sound reaching the receiver is determined when an acoustical visibility ray encounters a sound source directly. Indirect sound is approximated using stochastic (distribution) ray-tracing. When a ray intersects a diffuse surface at point p, tracing of the ray terminates and the sonel map is used to provide an estimate of the acoustical energy leaving point p and arriving at the receiver using a density estimation algorithm. The energy is scaled to account for attenuation by the medium and added to the accumulating echogram. Specular reflections are handled using the same approach as in the sonel tracing stage whereby ideal specular reflections are assumed. When an acoustical visibility ray encounters a sound source, the fraction of energy leaving the sound source and arriving at the

receiver is determined, scaled to account for attenuation by the medium and then added to the accumulating echogram. Diffraction effects that occur when an acoustical visibility ray encounters an edge are handled using a modified version of the Huygens-Fresnel principle.

Being a Monte-Carlo based method, sonel mapping allows for arbitrarily complex geometries, low memory consumption and the ability to provide an "efficiency vs. accuracy" trade-off whereby the solution can be made more accurate by increasing the number of emitted sonels. Its probabilistic nature also allows for paths of sound ray propagation of arbitrary length to be explored, paths that may not necessarily be explored using deterministic approaches. Finally, although the focus of this work is acoustical modeling for the purposes of auralization, sonel mapping is not specific to acoustical wave energy propagation. Rather, sonel mapping is a framework for energy propagation despite the fact that in this work, it is used to model acoustical energy propagation. Given the appropriate model parameters (for example, source emission functions, wave/surface interaction models etc.), it can be used to model the propagation of any type of wave energy, be it sound, light etc. including any type of interaction between the wave and the medium (if a medium is required for the wave to propagate) in addition to any surfaces/objects the wave may encounter.

Chapter 5

Comparison to Physical Acoustical Properties

Despite the availability of various acoustical modeling and auralization systems (see Chapter 3), there is a lack of detailed evaluation of the effectiveness and accuracy of these systems [179]. This lack of accuracy evaluation stems from the difficulties associated with making a quantitative comparison of measured and computed impulse responses given the complexity associated with real-world environments [179]. As a result, the majority of evaluations have involved making comparisons of various properties/statistics such as reverberation time (see [100]) and/or human listening tests to validate computer simulations [5, 179]. In this dissertation, various simulations were performed with various sound source, receiver and environmental configurations (e.g., presence or absence of occluders etc.) and the results of these simulations are compared to analytical/theoretical results in order to demonstrate that sonel mapping satisfies and conforms to real-world acoustical energy propagation. This in-

cludes comparisons between measured and theoretical reverberation times and measured vs. predicted sound energy attenuation. Although desirable, validation involving human user tests is beyond the scope of this dissertation. Validation that involves comparisons between measurements made in an actual (controlled) room and the results of the same measures made in a simulation of the room are also not included. Such tests require the use of a simple controlled room (environment) where various parameters (e.g., surface reflection and absorption coefficients) can be carefully controlled in order to allow meaningful comparisons to be made. Given the lack of such available data for a simple room whose parameters can be easily controlled, constructing such an environment is also beyond the scope of this dissertation. There have been attempts to address acoustical rendering algorithms using such an approach. For example, Tsingos et al. [179] developed the Bell Lab's Box for validating sound propagation simulations however, detailed data regarding sample measurements made in the room are not readily available as of the writing of this dissertation. Greater details regarding potential future user testing and physical measurement comparisons are provided in Chapter 7.

The simulations described in this chapter are divided into two categories i) individual component simulations and ii) simulations involving the complete algorithm (e.g., all components working together). The individual component simulations examine various details of the components comprising the sonel

Parameter	Value
Receiver Radius (r_k)	0.15m
Source Power (L_s)	90dB
Echogram bin spacing (t_{rir})	5ms
d_s	0.1
n_{max}	10
N_{vis}	30

Table 5.1: Sonel mapping algorithmic parameters for all simulations unless specified otherwise.

mapping algorithm (e.g., its ability to model diffraction, specular and diffuse reflections and absorption). The complete simulations examine the sonel mapping algorithm as an entire system whereby each of the possible interaction types (e.g., specular and diffuse reflections, diffraction and absorption) occur. All of the simulations described in this chapter were performed using a Linux-based PC with a Pentium III 500MHz processor and 512Mb RAM. Unless specified otherwise, the sonel mapping algorithmic parameters (as described in Chapter 4) used for the simulations are summarized in Table 5.1.

5.1 Individual Component Simulations

The simulations presented in this section consider the various components comprising the sonel mapping algorithm separately. In the first simulation, a graphical illustration of the various components comprising sonel mapping are presented. In the second simulation, only diffuse reflections are considered, in the third simulation all interactions are either diffuse reflections, specular reflections or absorption and finally, diffraction is considered in isolation.

5.1.1 Graphical Illustration

In this section, a graphical demonstration of the various components comprising the sonel mapping algorithm are presented in the form of color-filled contour plots. This is accomplished by illustrating the sound energy propagation as simulated for a particular environment (room) for various sound source, receiver and occluder (edge) configurations and providing comparisons to theoretical results. The simulations were conducted in a simulated enclosure (room) of 10.00m $\times$ 8.00m $\times$ 10.00m (Figure 5.1). The frequency of the sound source was 250Hz (λ = 1.37m). The sound source was positioned at location (0.69m, 4.00m, 4.83m) and remained stationary throughout all scenarios considered while the position of the receiver was varied across the x-z plane (e.g., y-axis remained constant at y = 4.00m) in equal increments equal to $\lambda/2$ or 0.685m (the x coordinate ranged from 1.37m to 8.93m while the z coordinate ranged from 0.35m to 8.97m). A flat surface (occluding plane) 3.50m $\times$ 5.00m was positioned such that it formed a plane along the "y-z" axis (e.g., constant x). The coordinates of the vertices comprising the edge were (3.45m, 0.00m, 3.45m), (3.45m, 5.00m, 3.45m), (3.45m, 0.00m, 6.45m) and (3.45m, 0.00m, 6.45m). In all scenarios considered, unless stated otherwise, the number of sonels emitted in the sonel tracing stage was $30,000$ and the number of acoustical visibility rays traced in the acoustical rendering stage was also $30,000$. Greater details regarding each of the scenarios considered are provided in the following sections.

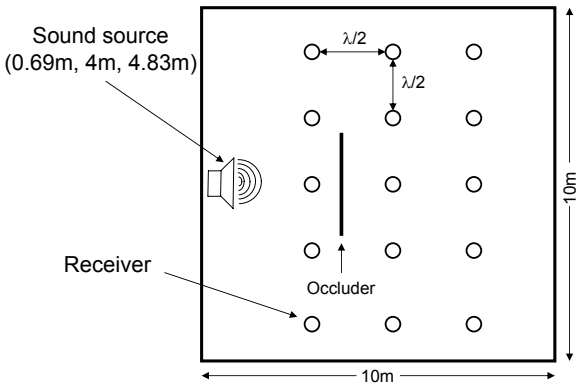


Figure 5.1: Room set-up for the individual component graphical simulations (not to scale). The height (y-coordinate) of the sound source and the receiver positions was constant at y = 4.00m. Receiver positions varied across the x-z plane in equal increments of $\lambda/2$ or 0.69m (the x coordinate ranged from 1.37m to 8.93m while the z coordinate ranged from 0.35m to 8.97m)

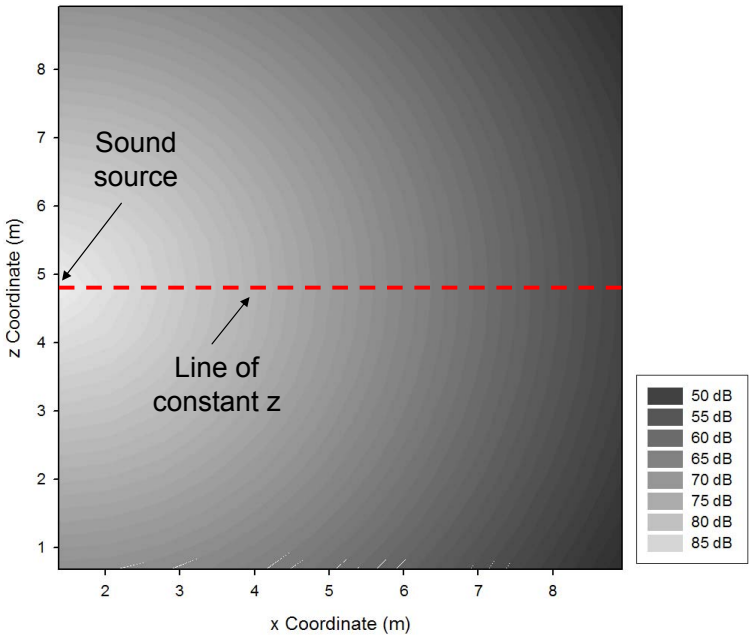
5.1.1.1 An Open Environment

In this scenario, the occluder was absent and the absorption coefficient (α) of each of the six surfaces comprising the room was assigned a value of one (e.g., $\alpha = 1$). In other words, upon encountering a surface, the sound was completely absorbed (e.g., no reflections). Since there was no obstructions present, there was a direct path between the sound source and each of the receiver positions. Furthermore, given that no sound was reflected, only direct sound could reach the receiver. Given the absence of diffuse reflections, there was no need for the sonel tracing stage and therefore, only the acoustical rendering stage was executed. The resulting contour plot is shown in Figure 5.2(a) where, receiver level (dB) is given as a function of position across the plane of constant y. Receiver level decreases with increasing distance from the sound source. In Figure 5.2(b), the energy ($W \cdot m^{-2}$) across a line of positions along the x-axis of

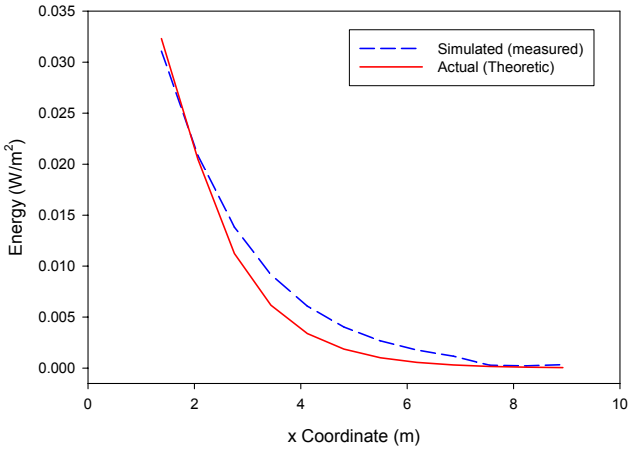
constant z (z = 4.80m) is shown (dashed blue line) illustrating the exponential decrease in energy with increasing sound source distance (energy as opposed to level is shown to illustrate the exponential decrease in energy with increasing sound source distance). Included in Figure 5.2(b) is the plot of the actual (theoretical) results (solid red line) of each corresponding measurements computed using Equation 4.7 that describes the attenuation of sound energy due to absorption by the air (see Section 4.3.3.3). The difference between the results of the simulated and actual measurements are not significantly different from each other (T value 0.28 P= 0.78 Degrees of Freedom 22).

5.1.1.2 Energy Propagation in the Presence of an Occluder Without Diffraction

This simulation is similar to the simulation presented above except for the presence of the occluder. As with the six surfaces comprising the room, the absorption coefficient of the surfaces of the occluder (both sides) was assigned a value of one (e.g., $\alpha = 1$) indicating that no sound was reflected off of it. The resulting plot of receiver level (dB) as a function of position across the plane of constant y is shown in Figure 5.3. As Figure 5.3(a) illustrates, given that the surfaces comprising the room and the occluder were perfect absorbers of sound energy, no sound energy could reach any receiver whose position was such that the occluder obstructed the direct path to the sound source. As a result, the



(a) Contour plot.



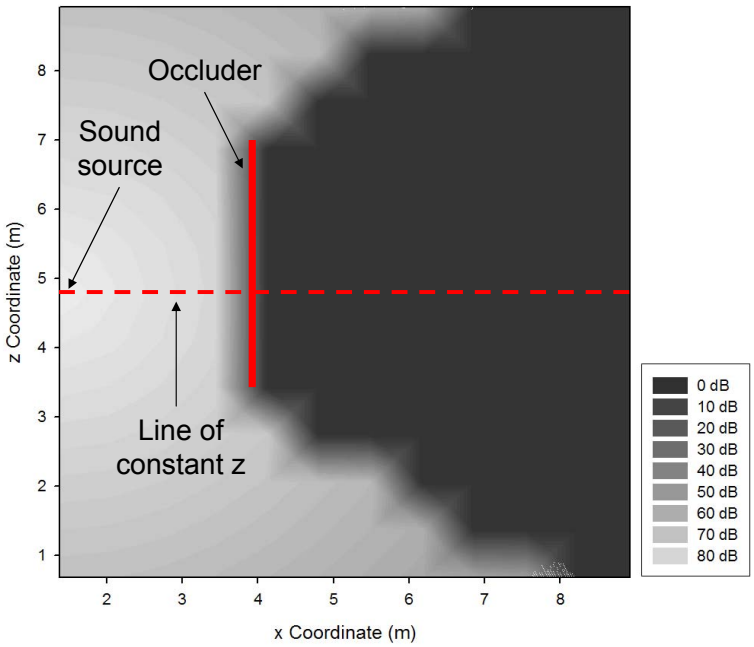
(b) Energy across a line of constant z.

Figure 5.2: (a) Contour plot (receiver level (dB) as a function of position across a plane in the x-z axis) for the simple energy propagation simulation. (b) Energy of a line of receiver positions across the x-axis of constant z ($z = 4.8$). vs. the actual (theoretical) results.

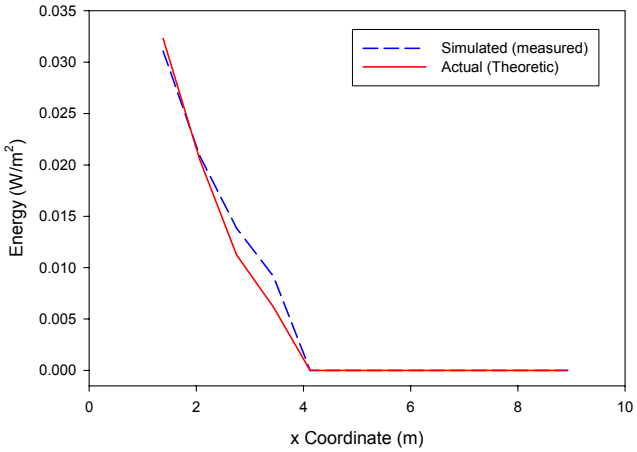
sound level at such receivers was zero or in other words, these receivers were within the "shadow region" (regions of black on the plot). In Figure 5.3(b), the energy ($W \cdot m^{-2}$) across a line of positions along the x-axis of constant z ($z = 4.80m$) is shown (dashed blue line) illustrating the decrease of energy in the shadow region due to the presence of the occluder and hence the occlusion of the direct sound. Included in Figure 5.3(b) is the plot of the actual (theoretical) results (solid red line) of each corresponding measurements computed using Equation 4.7 that describes the attenuation of sound energy due to absorption by the air (see Section 4.3.3.3). In this scenario, given the presence of the (ideal) occluder, any measurements whose x-coordinate was greater than $3.45m$ (e.g., greater than the x-coordinate of the occluder) resulted in an energy measurement of zero. The difference between the results of the simulated and actual measurements are not significantly different from each other (T value 0.90 P= 0.93 Degrees of Freedom 22).

5.1.1.3 Energy Propagation in the Presence of Specular Reflections

In this simulation the occluder was present and only specular reflections were permitted off of the surfaces comprising the occluder in addition to the six surfaces comprising the room. The absorption and specular reflection coefficients of the surfaces comprising the edge and the room were each set to 0.1 and 0.9 respectively. As in the previously described simulations, given the absence of



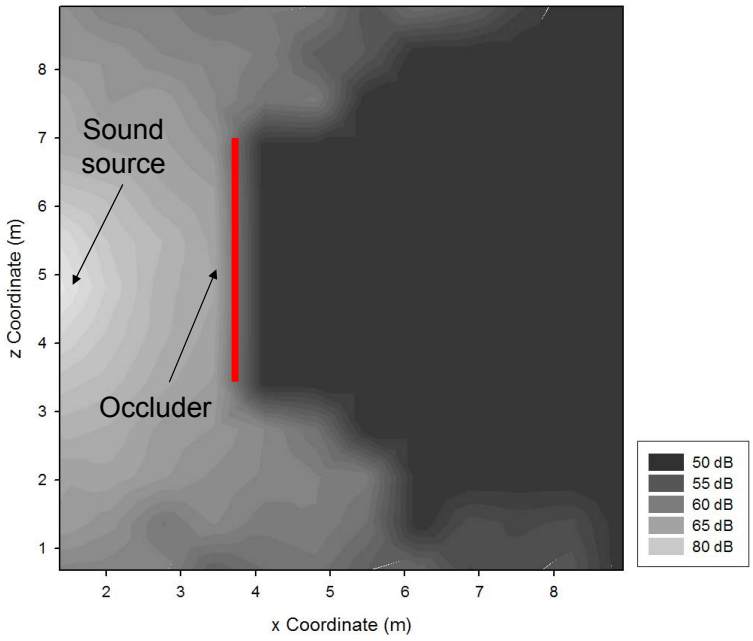
(a) Contour plot.



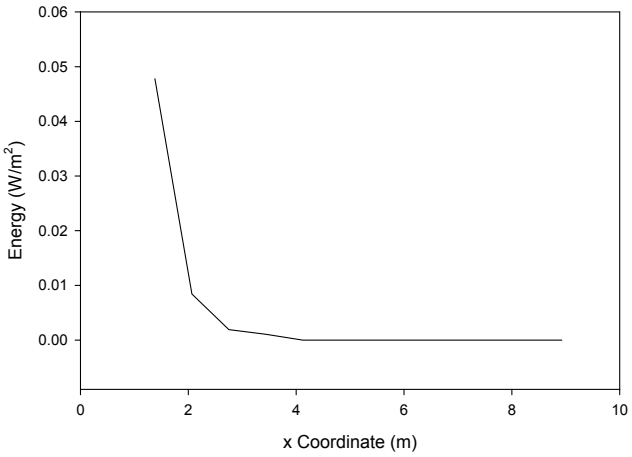
(b) Energy across a line of constant z.

Figure 5.3: (a) Contour plot (receiver level as a function of position across a plane in the x-z axis) for the energy propagation in the presence of an occluder without diffraction simulation. (b) Energy of a line of receiver positions across the x-axis of constant z (z = 4.8). vs. the actual (theoretical) results.

diffuse reflections, there was no need for the sonel tracing stage and therefore, only the acoustical rendering stage was executed. The resulting contour plot is shown in Figure 5.4(a). In Figure 5.4(b), the energy ($W \cdot m^{-2}$) across a line of positions along the x-axis of constant z (z = 4.8m). Given the presence of the occluder, direct sound could not reach the shadow region however, sound does reach portions of the shadow region via specular reflections off of the surfaces within the environment. A comparison between Figure 5.4(b) and Figure 5.3(b) illustrates the increase in energy with direct sound and in the presence of specular reflections as opposed to considering direct sound only (e.g., the maximum energy in Figure 5.4(b) is $0.0475 W \cdot m^{-2}$ while the maximum energy in Figure 5.3(b) is $0.0317 W \cdot m^{-2}$). Given the presence of the occluder, receiver positions where the occluder obstructed the direct path between the sound source and the receiver received no direct sound energy. However, although no direct energy reached such positions, indirect energy via specular reflection did reach receivers whose direct path to the sound source was occluded. Since no direct energy reached occluded receiver positions, the sound level at these receivers was less than the sound level at corresponding positions in the absence of the occluder.



(a) Contour plot.



(b) Energy across a line of constant z.

Figure 5.4: Specular reflections off of the six surfaces comprising the room in addition to the surfaces comprising the occluder in addition to direct energy. (a) Contour plot for the energy propagation in the presence of specular reflections simulation. (b) Energy of a line of receiver positions across the x-axis of constant z (z = 4.80m).

5.1.1.4 Energy Propagation in the Presence of Diffuse Reflections

In this scenario, diffuse reflections were considered only. However, in contrast to the specular reflection scenario described in the previous section, reflections were permitted only off the surface of the occluder. The absorption coefficient of the six surfaces comprising the room were each set to 1.0 (e.g., perfect absorbers), while the absorption and diffuse reflection coefficients of the occluder were set 0.1 and 0.9 respectively. In the scenarios previously described, since no diffuse reflections were considered, to determine the energy at each receiver position, only the acoustical rendering stage was executed (e.g., the purpose of the sonel tracing stage is to populate the sonel map that stores diffuse reflected energy only). In the current scenario, diffuse reflections were considered, and therefore, both the sonel tracing and acoustical rendering stages were executed. The resulting plot (receiver level (dB) as a function of position across the plane of constant y) is shown in Figure 5.5, where the level of the sound reaching the receivers that are in front of the occluder only (e.g., receiver positions whose x-coordinate is less than the x-coordinate of the edge), is illustrated. By considering receivers in front of the occluder only, emphasis is placed on the diffuse energy reflected off of the surface. The energy reflected off of the occluder is clearly evident in Figure 5.5.

To illustrate the effect of increasing the number of sonels and acoustical visibility rays in the sonel tracing and acoustical rendering stages respectively, in

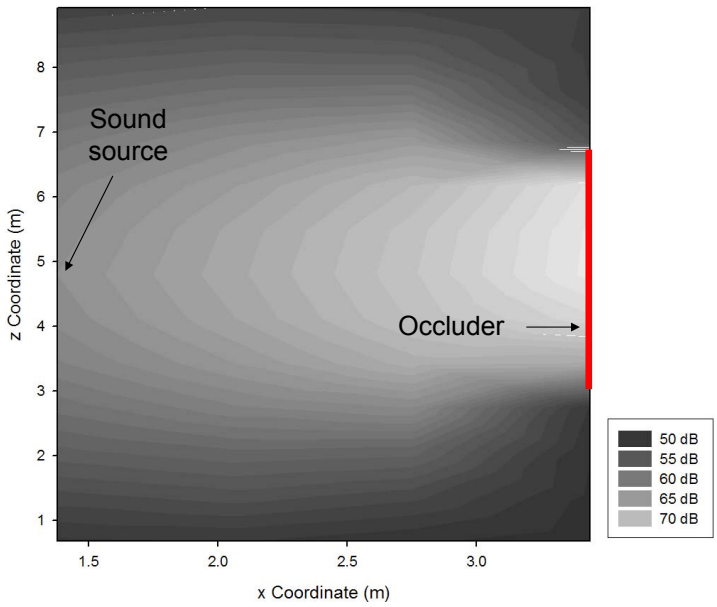
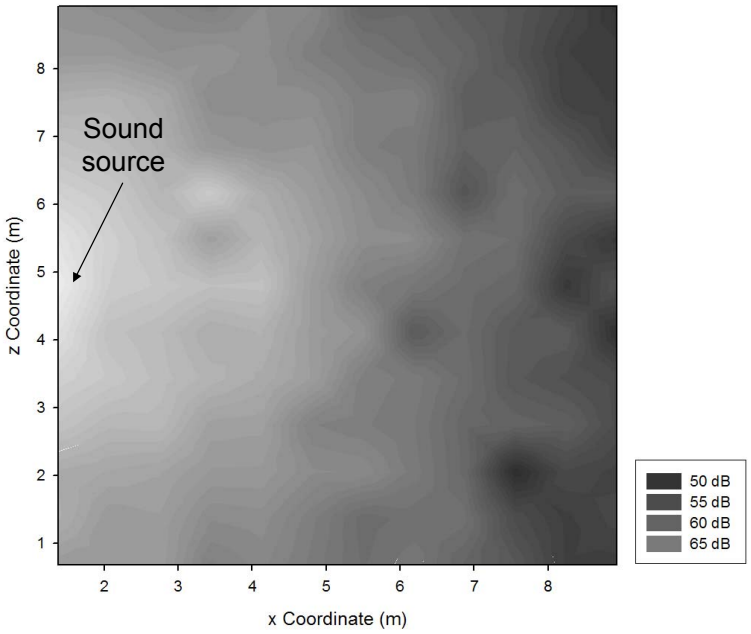
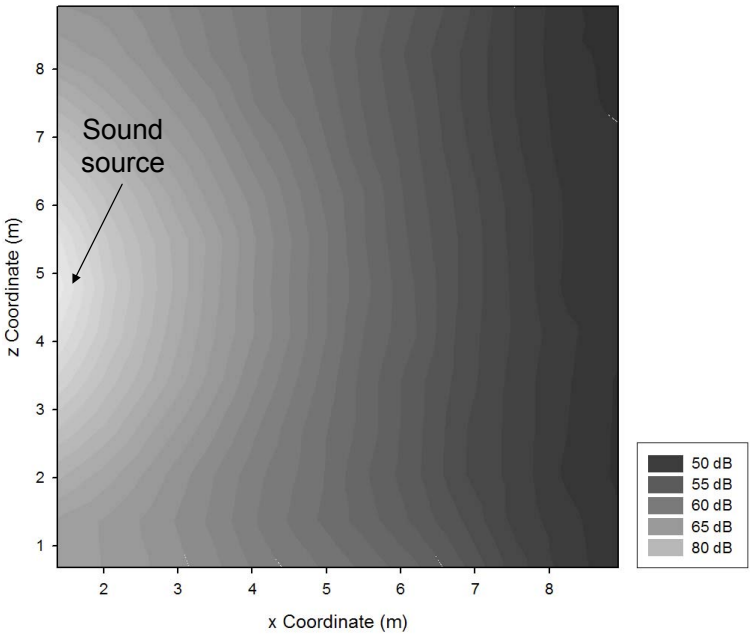


Figure 5.5: Contour plot (receiver level as a function of position across a plane in the x-z axis) for the energy propagation in the presence of diffuse reflections simulation. Diffuse reflections off of the surface comprising the occluder. The sound level for receiver positions in front of the edge (e.g., receiver positions whose x-coordinate are less than the x-coordinate of the edge) are shown only.

the following simulation, the occluder was absent and diffuse reflections were permitted off of the six surfaces comprising the room. The absorption and diffuse reflection coefficients of each surface were set to 0.1 and 0.9 respectively. Initially, 1000 sonels were emitted from the sound source in the sonel tracing stage and 1000 acoustical visibility rays were traced during the acoustical rendering stage and the sound level at each of the considered receiver positions was measured. The resulting plot is shown in Figure 5.6(a). This scenario was then repeated by increasing the number of sonels and acoustical visibility rays. In particular, $30,000$ sonels were emitted from the sound source during the sonel tracing stage and $30,000$ acoustical visibility rays were traced from the receiver during the acoustical rendering stage. The resulting contour plot is shown in Figure 5.6(b). The maximum sound level in the first case (Figure 5.6(a), 1000 sonels/acoustical visibility rays) is 65dB as opposed to 80dB in the second scenario (Figure 5.6(b), $30,000$ sonels/acoustical visibility rays). Given that there are less acoustical visibility rays, the probability of an acoustical visibility ray encountering a sound source during the acoustical rendering stage is lower thereby leading to a decrease in the sound level reaching the receiver. The plot in the $30,000$ sonels/acoustical visibility rays case is more "refined" than the 1000 sonels/acoustical visibility rays case illustrating that as the number of sonels/acoustical visibility rays increases, the distribution of sound energy does approach the actual energy distribution.



(a) 1000 sonels and acoustical visibility rays.



(b) 30,000 sonels and acoustical visibility rays.

Figure 5.6: Contour plot (receiver level as a function of position across a plane in the x-z axis) for the energy propagation in the presence of diffuse reflections simulation in the absence of the occluder. (a) 1000 sonels and 1000 acoustical visibility rays emitted during the sonel tracing and acoustical rendering stages respectively. (b) 30,000 sonels and 30,000 acoustical visibility rays emitted during the sonel tracing and acoustical rendering stages respectively.

5.1.1.5 Energy Propagation in the Presence of Diffraction

This simulation is similar to the simulation described in Section 5.1.1.2 (energy propagation in the presence of an occluder without diffraction) except that here diffraction was permitted off of the occluder. The absorption coefficient of each of the six surfaces comprising the room in addition to the surfaces comprising the occluder was assigned a value of one (e.g., $\alpha = 1$). As a result, no specular or diffuse reflections were permitted thus isolating any diffraction effects. The resulting plot (receiver level (dB) as a function of position across the plane of constant y) is shown in Figure 5.7. To illustrate the inverse relationship between diffraction and sound frequency, this demonstration was repeated for a sound source frequency of 500Hz. The theoretical diffraction model dictates an inverse relationship between frequency and diffraction. Given a greater sound source frequency, diffraction effects should be smaller or in other words, the shadow zone should greater. In addition to increasing the frequency of the sound source, the dimensions of the occluder (across the z-axis) were also decreased from 3.5m to 2.5m to further illustrate the inverse relationship between frequency and diffraction. The resulting plot (receiver level (dB) as a function of position across the plane of constant y) is shown in Figure 5.8. The "shadow region" (250Hz) found in Figure 5.7 although still present, is much smaller that the "shadow region" (500Hz) in 5.8. When diffraction occurs, sound can "bend" around the edge and therefore sound energy will still reach many receiver po-

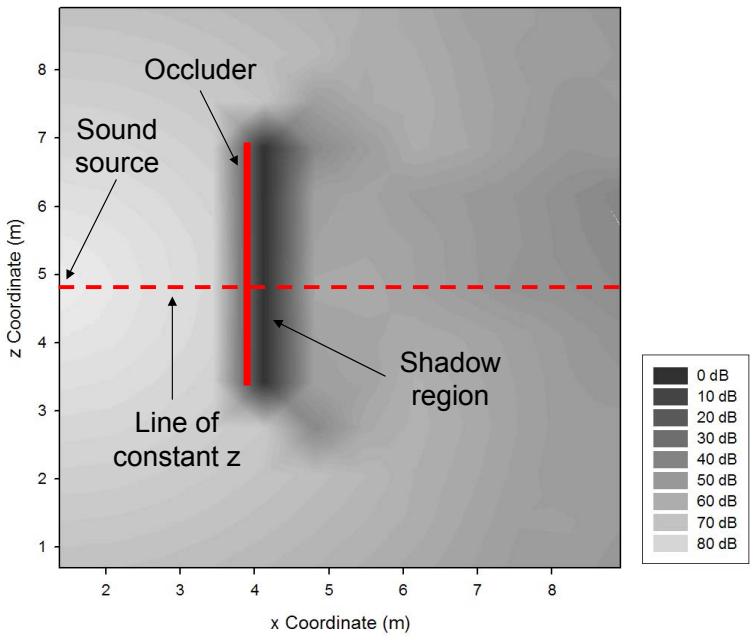


Figure 5.7: Contour plot (receiver level as a function of position across a plane in the x-z axis) for the energy propagation in the presence of an edge and in the presence of diffraction simulation (250Hz).

sitions where the direct path between them and the sound source is obstructed by the occluder.

In Figure 5.9, the energy ($W \cdot m^{-2}$) across a line of positions along the x-axis of constant z ($z = 4.8m$) of Figure 5.7 and Figure 5.8 is shown. A comparison between the two plots illustrates the effects of diffraction. The energy associated with the line of constant z for the frequency of 250Hz sound (dashed blue line) is generally greater than the energy of the 500Hz frequency sound (solid red line).

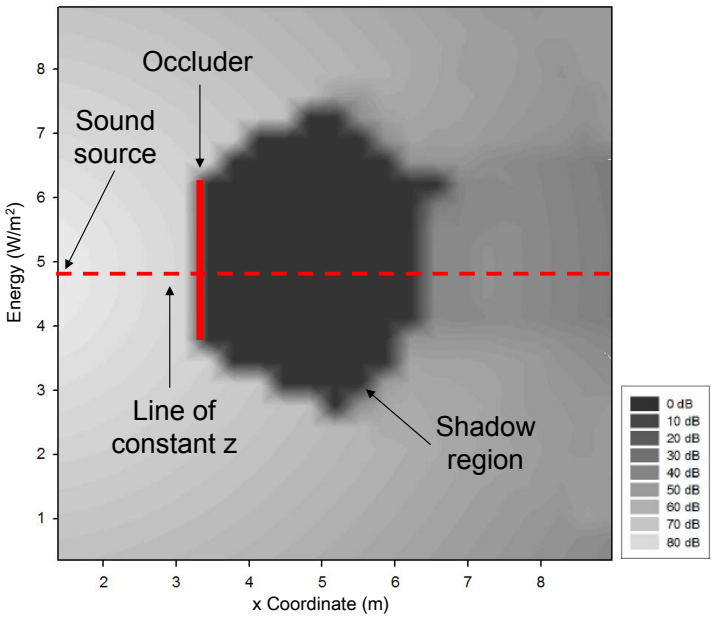


Figure 5.8: Contour plot (receiver level as a function of position across a plane in the x-z axis) for the energy propagation in the presence of an edge and in the presence of diffraction simulation (500Hz).

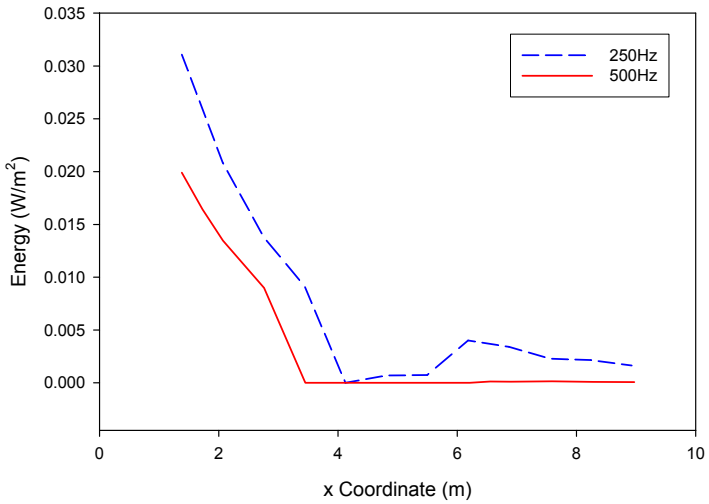


Figure 5.9: Energy of a line of receiver positions across the x-axis of constant z (z = 4.80m) for the energy propagation in the presence of an edge and in the presence of diffraction simulation for both the 250Hz (dashed blue line) and 500Hz (solid red line) frequencies.

5.1.2 Reverberation Time: Simulated vs. Theoretical Results

In this simulation[1] the reverberation time (the time required for the total energy emitted by a sound source to drop by a factor of one million (or 60dB) [100] and denoted by RT_{60}) for a simple, box-like enclosure was estimated by the system for several sound source/receiver configurations using a deterministic termination criterion (the energy discontinuity percentage (EDP) that represents the percentage of the original ray energy that must be lost before the ray is terminated). Two frequency bands, with center frequencies of 2kHz and 4kHz were considered in this simulation. The difference between the estimated and theoretical values predicted by Sabine's formula [122] provides a measure of system accuracy. The dimensions of the room were 10m $\times$ 9m $\times$ 8m. The position (x,y,z coordinates, in meters) of the single omni-directional sound source was $(9,8,7)$, while the positions of the four receivers were $(4,4,4)$, $(2,5,7)$, $(3,1,3)$ and $(6,6,4)$. The surfaces of the room (four walls, ceiling and floor) were assigned frequency dependent absorption coefficient values α corresponding to particular materials as given in [122] (e.g., a wooden floor on joists, smooth plaster ceiling and each of the remaining four walls $3/8$" thick plywood) and listed in Table 5.2.

As assumed in Sabine's formulation for reverberation time, in this test all reflections were assumed to be diffuse (e.g., there were no specular reflections

[1]These results originally appeared in Kapralos et al. [89].

Surface	Material	α (2kHz)	α (4kHz)
Floor	Wood (on joists)	0.15	0.11
Ceiling	Plaster (smooth finish)	0.14	0.10
Walls	3/8" Plywood panel	0.28	0.22

Table 5.2: Surface absorption coefficients for the reverberation time simulation.

Receiver Position	RT_{pre} (2kHz)	RT_{est} (2kHz)	%dif
$4,4,4$	2.42s	2.45s	1.24
$2,5,7$	2.42s	2.38s	1.65
$3,1,3$	2.42s	2.40s	0.83
$6,6,4$	2.42s	2.54s	5.00

Table 5.3: Results for the reverberation time simulation. Comparison between estimated (RT_{est}) and predicted (RT_{pre}) reverberation time for four receiver positions and a 2kHz sound source.

or diffraction). The diffuse reflection coefficient of surface i (δ_i) was obtained as $d_i = 1 - \alpha_i$ (where α_i is the absorption coefficient of surface i). The number of sonels emitted form the sound source was 100,000 and the EDP was set to a value of 100 - 10^{-6}. The number of acoustical visibility rays traced from the receiver was also $100,000$.

Reverberation times were estimated by computing a linear regression on the -5 to -35dB portion of the decay curve [103]. The decay curve itself was obtained from the echogram using Schroeder's backwards integration method [161]. A summary of the results is provided in Tables 5.3 and 5.4 where the absolute percent difference ($\%dif$) between the reverberation time as predicted by Sabine's formula (RT_{pre}) and the reverberation time as estimated by sonel mapping (RT_{est}) for the 2kHz and 4kHz sound sources are shown.

As shown in Tables 5.3 and 5.4, the maximum difference across both fre-

Receiver Position	RT_{pre} (4kHz)	RT_{est} (4kHz)	%dif
4, 4, 4	1.80s	1.81s	0.56
2, 5, 7	1.80s	1.71s	5.00
3, 1, 3	1.80s	1.72s	4.44
6, 6, 4	1.80s	1.84s	2.22

Table 5.4: Results for the reverberation time simulation. Comparison between estimated (RT_{est}) and predicted (RT_{pre}) reverberation time for four receiver positions and a 4kHz sound source.

quencies and all receiver positions is at or below 5.00%. The differences for the 2kHz frequency range from 0.83% to 5.00% with an average of 2.44% and a standard deviation of 0.07%. For the 4kHz frequency, differences range from 0.56% to 5.00% with an average of 1.77% and a standard deviation of 0.05%. Although the estimated values are very close to the predicted values, this was obtained by setting the energy discontinuity percentage (EDP) to a very high value of almost 100 when in reality, the EDP is typically set between 90 and 99 [200]. Setting the EDP to such a high value ensures the reverberation decay is linear and hence the linear regression used to estimate the reverberation time provides an accurate measure [200]. However, such a high EDP setting also leads to an increase in the required computation time since each ray/particle will propagate for a longer time.

5.1.3 Russian Roulette: Comparison to a Deterministic Approach

In this simulation[2], the applicability and effectiveness of a Russian roulette strategy to acoustical modeling applications is demonstrated. This demonstra-

[2]These results originally appeared in Kapralos et al. [91].

tion is accomplished by comparing the time required to compute reverberation time estimates using an energy discontinuity percentage (EDP) termination criterion and a Russian roulette termination criterion. The dimensions of the enclosure were 4m $\times$ 4m $\times$ 4m, the position (x, y, z coordinates, in meters) of the single omni-directional sound source was $(3.5, 3.5, 3.5)$ and the receiver was positioned at $(1.0, 1.0, 1.0)$. The absorption coefficient of each surface was set to 0.1 (a single frequency band was considered). The reverberation time as predicted by Sabine's formula (RT_{pre}), taking absorption by the medium into consideration was 1.03s. The difference between the time taken to compute the reverberation time estimate using an EDP termination criterion and then using a Russian roulette termination criterion is taken as the measure of performance in this simulation. As assumed in Sabine's formulation for reverberation time, in this test all reflections were assumed to be diffuse (e.g., "perfectly diffuse field" [51]). The diffuse reflection coefficient of surface i (δ_i) was obtained as $\delta_i = 1 - \alpha_i$ where, α_i is the absorption coefficient of surface i.

The time required to compute the reverberation times using an EDP termination criterion for various EDP values are shown in Table 5.5 (typical EDP values range from 90 to 99 [52]). For each EDP setting ranging from 90 to 99, the corresponding reflection count ("Ref. Count"), time taken to compute the estimate ("Time") and the estimated reverberation time ("RT_{est}") are provided. As in the previous simulation, reverberation times were estimated by comput-

EDP	Ref. Count (n)	Time (s)	RT_{60}
90.0	17	4.04	0.16
91.0	19	4.52	0.19
92.0	20	4.46	0.19
93.0	21	5.01	0.20
94.0	22	5.26	0.21
95.0	23	5.50	0.22
96.0	25	6.00	0.25
97.0	27	6.53	0.26
98.0	30	7.23	0.30
99.0	36	8.72	0.35

Table 5.5: Russian roulette simulation: reverberation time estimates using an energy discontinuity percentage (EDP) termination criterion.

ing a linear regression on the -5 to -35dB portion of the decay curve [103]. The decay curve itself was obtained using Schroeder's backwards integration method [161].

For each EDP-based reverberation time estimate, using a Russian roulette termination criterion, the number of sonels initially emitted from the sound source in the sonel tracing stage was adjusted such that the computed reverberation time was equal (within a small value) to the corresponding reverberation time computed with an EDP termination criterion. The number of sonels initially emitted from the sound source during the sonel tracing stage (stage one) was constant ($15,000$). Similarly, the number of acoustical visibility rays emitted during the acoustical rendering stage (stage two) was also constant at one tenth the number of sonels emitted in the sonel tracing stage (e.g., $15,000/10 = 1000$). A summary of the Russian roulette-based results are provided in Table 5.6 where, for each of the estimated reverberation times, the

number of sonels required to compute it (“Num. Sonels”), the maximum reflection count (“Max. Ref.”) encountered by any of the emitted sonels, the time taken to compute the solution (“Time”) and the percent difference (“%dif”) between the time taken to compute the reverberation time with an EDP termination criterion t_{edp} and the time to compute the reverberation time with a Russian roulette criterion t_{rus} are listed (the reverberation times for the 91.0 and 92.0 EDP measurements both computed to 0.19s and therefore, the corresponding reverberation time estimate using a Russian roulette approach was computed once only). A positive percent difference indicates $t_{edp} > t_{rus}$ and a negative difference indicates $t_{edp} < t_{rus}$. The percentage difference for each entry of Table 5.6 is positive indicating that employing a Russian roulette approach as opposed to the deterministic EDP approach results in reduced computation time. In other words, a deterministic approach takes much longer to compute. Percentage difference ranged from 3570 to 510, decreasing as the number of emitted sonels was increased.

5.1.4 Diffraction

5.1.4.1 Correctness of the Acoustical Diffraction Method

In this simulation the ability of the Huygens-Fresnel principle to model acoustical diffraction is examined. This is accomplished by considering the energy reaching the receiver from the sound source using the Huygens-Fresnel imple-

RT_{60}	Sonel Count	Max. Ref.	Time (s)	%dif
0.16	400	63	0.11	3570
0.19	500	81	0.14	3130
0.20	550	83	0.16	3030
0.21	600	66	0.17	2990
0.22	650	81	0.18	2960
0.25	850	73	0.26	2210
0.26	2000	73	0.57	1050
0.30	3000	85	0.85	750
0.35	5000	79	1.43	510

Table 5.6: Russian roulette simulation: using a Russian roulette termination criterion to obtain the corresponding reverberation times obtained using an EDP termination criterion.

mentation (with the visibility of the first Fresnel zone assumed to be one) and comparing these results with the results obtained using the harmonic spherical wave model [79]

$$E = \frac{E_\circ}{\rho} cos(\omega t' - k\rho). \tag{5.1}$$

Here, E is the energy arriving at the receiver, $E_\circ$ is the energy of the source at time $t = 0$, ρ is the radius of the sphere representing the initial wavefront and set to a value equal to the distance between the sound source and the position on the edge of the diffracting sonel (see Figure 5.10), and t' is the time it takes for the wave to propagate a distance ρ (e.g., $t' = \rho/343m \cdot s^{-1}$). The frequency dependent attenuation of sound energy by the air was ignored for this simulation to avoid any frequency absorption effects thus ensuring any reduction in sound energy was due to occlusion effects.

As shown in Figure 5.10, the sound source and edge position (p_{edge}) re-

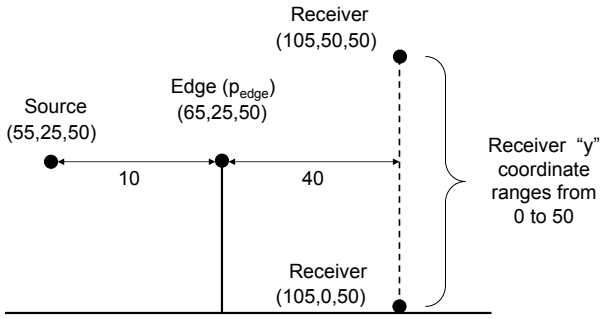


Figure 5.10: Room set-up used in the correctness of the acoustical diffraction method simulation.

mained stationary while the receiver's position (the "y" coordinate) was varied in unit increments from $y = 0$ to $y = 50$. The simulation was repeated for each of the frequency bands considered (e.g., 63Hz, 125Hz, 250Hz, 500Hz, 1kHz, 2kHz, 4kHz and 8kHz). The simulation was repeated using acoustical diffraction models where only the first Fresnel zone was considered and where all Fresnel zones were considered (see Section 4.4.1.4).

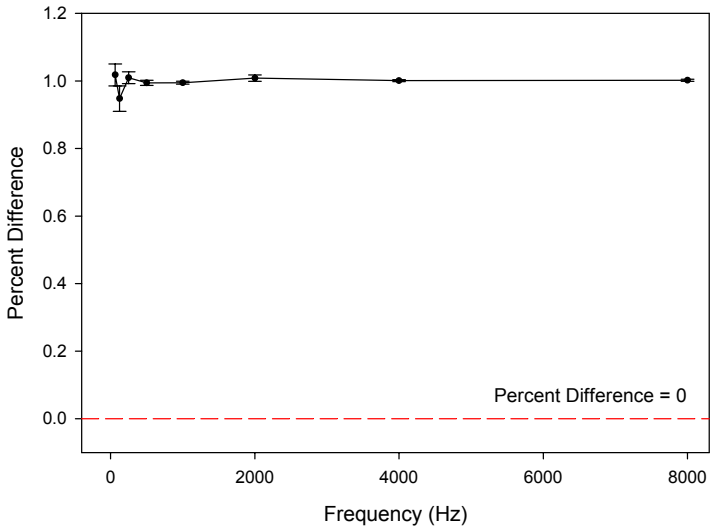
Table 5.7 lists the percentage difference ($\%dif$) between the predicted receiver sound level E_{pre} (Equation 5.1) and the receiver sound level computed using both acoustical diffraction implementations E_{est}. The average percent difference values for each frequency were obtained by averaging the energy values for each of the 51 receiver positions identified in Figure 5.10. Included in the table is the standard deviation (σ) for each average. A graphical summary of the results are presented in Figure 5.11 where the average percentage difference is plotted against frequency. Figure 5.11(a) illustrates the percentage difference as a function of frequency for the Huygens-Fresnel implementation

Frequency	One Zone		All Zones	
	Avg %dif	σ	**Avg %dif**	σ
63Hz	1.02	0.03	0.23	0.22
125Hz	0.95	0.04	0.23	0.21
250Hz	1.00	0.01	0.86	0.91
500Hz	0.99	0.01	0.61	0.26
1000Hz	0.99	0.01	0.79	0.51
2000Hz	1.00	0.01	1.47	1.32
4000Hz	1.00	0.01	5.88	2.58
8000Hz	1.00	0.01	2.85	2.38

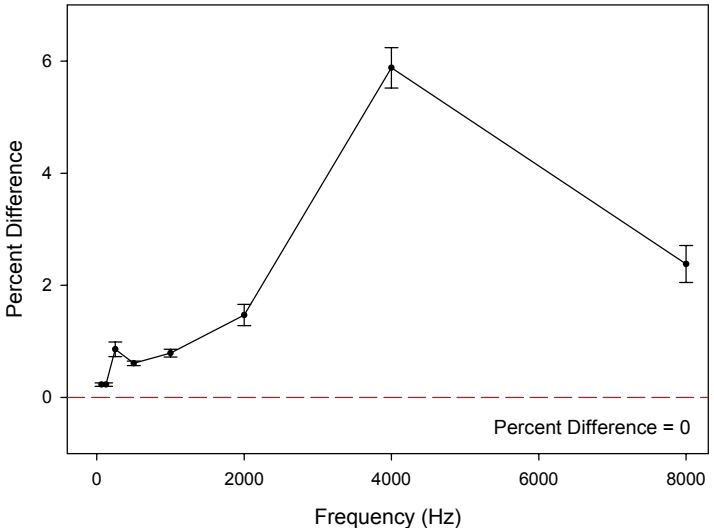
Table 5.7: Results for the correctness of the acoustical diffraction method simulation. Average percentage difference along with standard deviation as a function of frequency for both acoustical diffraction implementations.

where the first Fresnel zone was considered only while Figure 5.11(b) illustrates the percentage difference as a function of frequency for the Huygens-Fresnel implementation where all Fresnel zones were considered.

The smallest and largest average percentage difference for the diffraction implementation whereby only the first Fresnel zone was considered are small, 0.99 and 1.02 respectively. Despite ignoring the energy of all zones other than the first, this implementation provides a reasonable approximation. In contrast, the range of percentage differences for the diffraction implementation where all Fresnel zones were considered is larger, ranging from 0.23 to 5.88 and typically increase with increasing frequency. This increase in percent difference between the actual and computed receiver sound level may be due to numerical errors associated with locating a secondary source in each of the Fresnel zones. As frequency increases, the number of Fresnel zones also increases thus, any errors associated with locating a secondary source in a par-



(a) One Fresnel zone considered only.



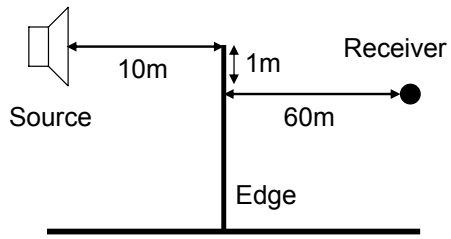
(b) All Fresnel zones considered.

Figure 5.11: Graphical summary of the results for the correctness of the acoustical diffraction method simulation. Average percentage difference along with error bars (standard deviation) as a function of frequency. (a) First Fresnel zone only considered and (b) all Fresnel zones considered.

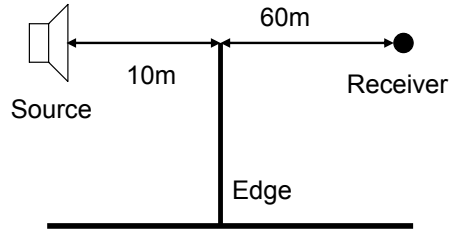
ticular Fresnel zone propagates through (e.g., locating a secondary source in Fresnel zone i requires the position of a secondary source in zone $i-1$ hence, an error in the position of the secondary source in zone $i-1$ may propagate and therefore, result in an incorrect secondary source position in zone i). In addition to providing more accurate results, using the approximation that considers the first Fresnel zone only is computationally more efficient.

5.1.4.2 First Fresnel Zone Visibility as a Function of Receiver Height

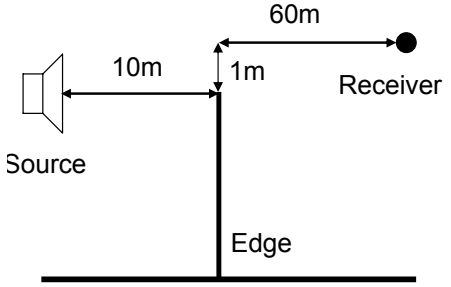
In this simulation, the visibility of the first Fresnel zone relative to the receiver and the sound level at the receiver was examined as a function of frequency. A stationary sound source and occluder were positioned at coordinates (40, 25, 50) and (50, 25, 50) respectively (at the same height with respect to the y-axis). The receiver was positioned at three locations: i) below the edge position at coordinates (110, 24, 50) (Figure 5.12(a)), ii) at the same height as the edge position at coordinates (110, 25, 50) (Figure 5.12(b)) and iii) above the edge position at coordinates (110, 26, 50) (Figure 5.12(c)). For each of the three scenarios, the energy reaching the receiver was calculated for each of the eight center frequencies (e.g., 63Hz, 125Hz, 250Hz, 500Hz, 1000Hz, 2000Hz, 4000Hz and 8000Hz). Frequency dependent attenuation of the sound by the air was ignored to allow for the frequency dependent diffraction effects to be examined. The purpose of this simulation is to compare the Huygens-Fresnel diffraction



(a) With respect to the y-axis, receiver is below the edge position.



(b) With respect to the y-axis, receiver is at same height as the edge position.



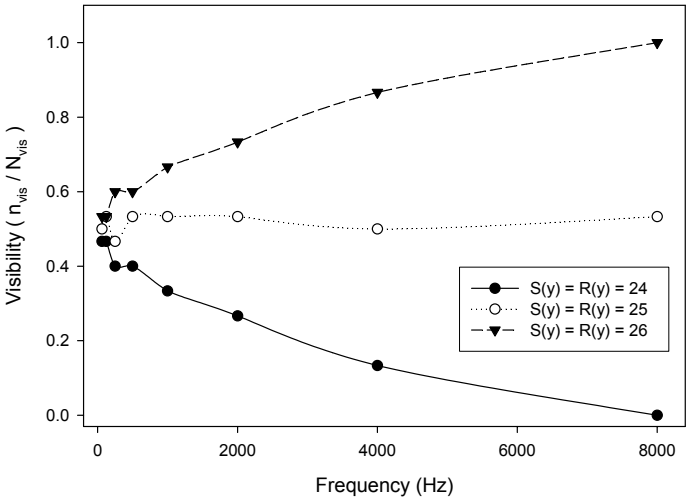
(b) With respect to the y-axis, receiver is above the edge position.

Figure 5.12: Set-up for the first Fresnel zone visibility as a function of receiver height simulation.

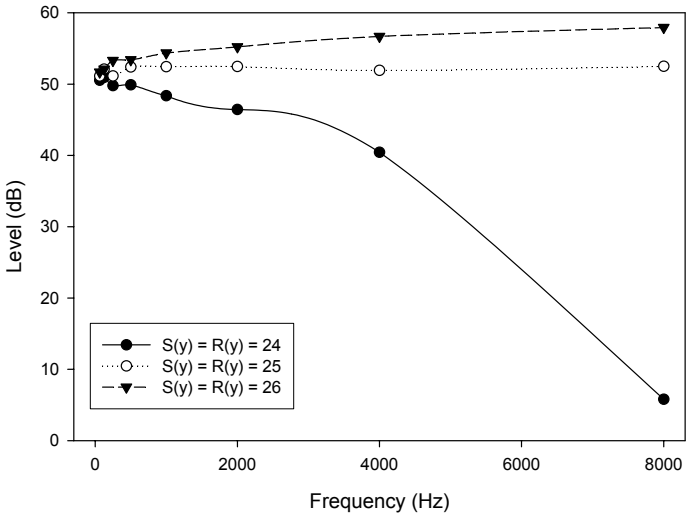
implementation with the theoretical diffraction model that states diffraction increases with increasing frequency (decreasing wavelength) for various sound source, edge and receiver configurations. Since the visibility of the first Fresnel zone is directly related to the amount of energy reaching the receiver via diffraction, frequency vs. visibility is used as a measure of performance.

A graphical summary of the results for each of the three scenarios is provided in Figures 5.13. In Figure 5.13(a), the visibility of the first Fresnel zone

relative to the receiver is plotted as a function of frequency. In Figure 5.13(b), receiver level is plotted as a function of frequency. In both plots, the filled circles and solid line represents the first scenario where the receiver is below the edge position. The open circle and short dashed line represents the scenario where the receiver is at the same height as the edge position (and above the sound source) and the triangle and long dashed line represents the scenario where the receiver is above the edge position. The results for the configuration considered in the first scenario (e.g., receiver below the edge position) are as expected. In particular, the visibility of the first Fresnel zone is inversely proportional to frequency whereby, as frequency increases, visibility decreases. The decrease in visibility is due to a decrease in the size of the first Fresnel zone and this results in a decrease in the sound energy reaching the receiver. As a result, as frequency increases, the sound energy reaching the receiver decreases, thus conforming to the theoretical model that predicts an increase in diffraction as frequency decreases [48]. The results of the second scenario where the receiver is positioned at the same height as the edge position are also as expected. The visibility is approximately 0.5 irrespective of frequency indicating that half of the zone is visible relative to the receiver. Finally, in the third scenario where the height of the receiver is greater than the height of the edge, visibility and frequency share a direct relationship whereby visibility increases with increasing frequency. This is due to the fact that as frequency



(a) Frequency vs. visibility.



(b) Frequency vs. level

Figure 5.13: Results for the first Fresnel zone visibility as a function of receiver height simulation: frequency vs. visibility and sound level for various sound source and receiver heights relative to the edge position. (a) Frequency vs. visibility and (b) frequency vs. receiver sound level.

increases, Fresnel zone size decreases and therefore, when the height of the receiver is greater than the height of the edge, less of the Fresnel zone will be

occluded relative to the receiver.

5.1.4.3 First Fresnel Zone Visibility Over a Plane of Receiver Positions

As with the previous simulation described above, the purpose of this simulation is to compare the Huygens-Fresnel diffraction implementation with the theoretical diffraction model that states diffraction increases with increasing frequency (decreasing wavelength) for various sound source, edge and receiver configurations. However, as opposed to the simulation considered above where only the height of the receiver was varied relative to the sound source and occluder, the receiver's position varied across a plane. The visibility of the first Fresnel zone was determined as a function of frequency while the sound source position remained stationary and the position of the receiver was varied across a plane (see Figure 5.14). The sound source was positioned at location (55, 23, 50) and the edge point (p_{edge}) was positioned at location (65, 25, 50). Receiver positions varied across the y-z plane whereby the x-coordinate remained stationary at 100 while the y coordinate varied from 0 to 50 and the z coordinate varied from 25 to 75, both increasing in unit increments. As with the previous simulation, frequency vs. visibility is used as a measure of performance.

The results of this simulation for the 63Hz, 125Hz, 250Hz, 500Hz, 1000Hz, 2000Hz, 4000Hz and 8000Hz frequencies are summarized in Figures 5.15 and 5.16 where the visibility of the first Fresnel zone relative to the receiver position is given as a function of frequency. For positions where the receiver height

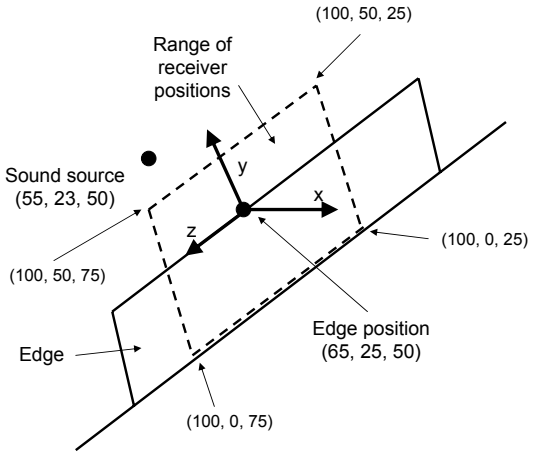
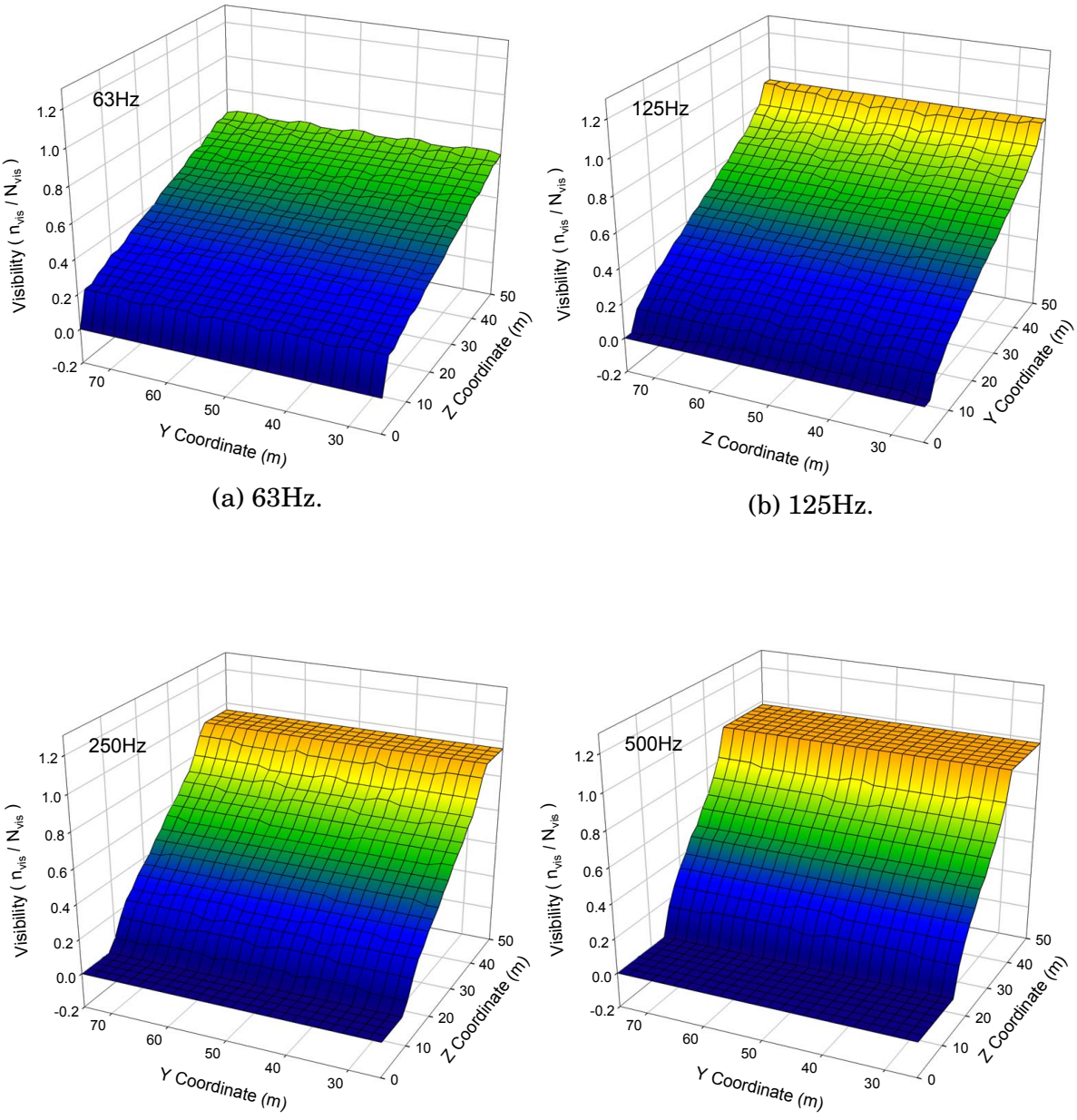


Figure 5.14: Set-up for the first Fresnel zone visibility over a plane of receiver positions simulation.

(y-coordinate) was less than the height of the edge, visibility decreased with increasing frequency as expected and as confirmed in the previous simulation. The decrease in visibility was due to a decrease in the visibility associated with increasing frequency. For positions where the receiver height was greater than the height of the edge position, visibility increased with increasing frequency, once again, as described in the previous simulation, due to the fact that as frequency increases, Fresnel zone size decreases. Therefore, when the height of the receiver is greater than the height of the edge, less of the Fresnel zone will be occluded relative to the receiver.

5.1.4.4 Diffraction by a Non-Infinite Edge

In this simulation, a "non-infinite" occluder with dimensions 2m $\times$ 2m was placed between the sound source and receiver. The configuration of the sound source, occluder and receiver is illustrated in Figure 5.17. The position of the



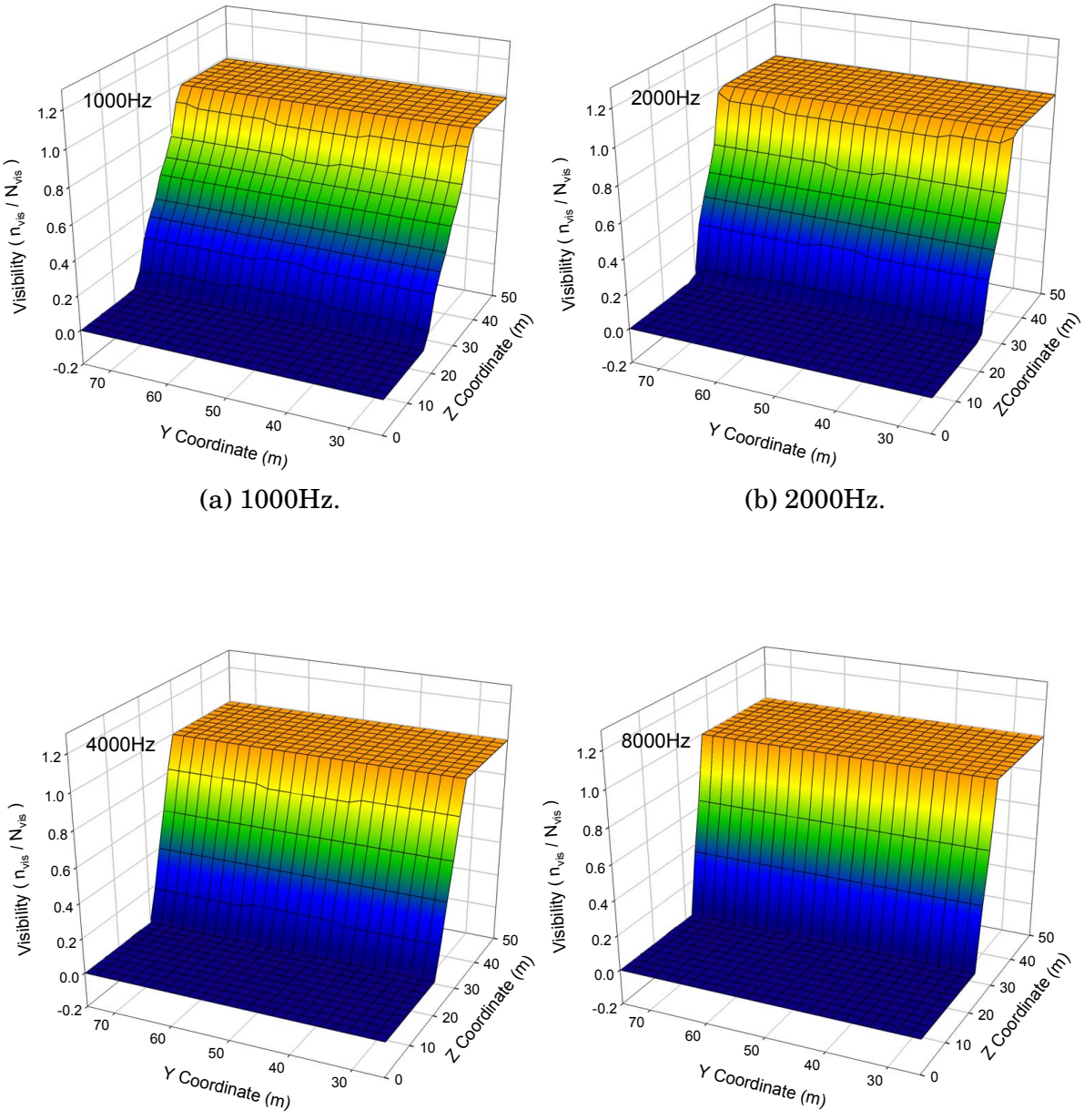
(a) 63Hz.

(b) 125Hz.

(c) 250Hz.

(d) 500Hz.

Figure 5.15: First Fresnel zone visibility over a plane of receiver positions simulation: frequency vs. visibility. (a) 63Hz, (b) 125Hz, (c) 250Hz and (d) 500Hz.



(a) 1000Hz.

(b) 2000Hz.

(c) 4000Hz.

(d) 8000Hz.

Figure 5.16: First Fresnel zone visibility over a plane of receiver positions simulation: frequency vs. visibility. (a) 1000Hz, (b) 2000Hz, (c) 4000Hz and (d) 8000Hz.

sound source remained stationary while the position of the receiver varied in one meter increments across the “y” and “z” coordinates, beginning at position (85, 75, 75) and ending at position (85, 85, 85). The sound source was positioned such that the y and z coordinates were centered with respect to the y and z coordinates of the edge. The purpose of this simulation is to once again examine diffraction effects by considering the visibility of the first Fresnel zone but over a non-infinite plane were sound can be diffracted via any of the four occluder’s edges. Since the corresponding wavelength of the frequencies considered are either greater than or less than the dimensions of the occluder, the inverse relationship between diffraction and frequency can be clearly demonstrated. Sound source energy was divided equally amongst four sonels (e.g., in this simulation, four sonels were emitted from the sound source only). It was assumed each emitted sonel fell incident on one of the four edges of the occluder and centered along the corresponding edge it was incident on (see Figure 5.17). This, along with the fact that only four sonels were emitted from the sound source ensures observations and conclusions can be made from the results. Assuming such a symmetrical configuration (e.g., each of the four emitted sonels is incident along one of the four edges and centered along the corresponding edge), allows for meaningful comparisons to be made without having to account for different sonel incident positions etc. In this simulation only edge effects were considered (e.g., no specular or diffuse reflections etc. were considered).

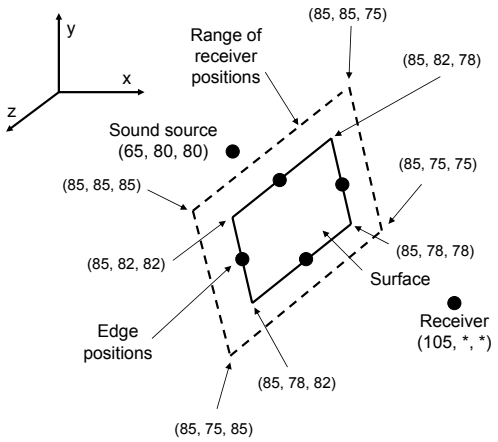
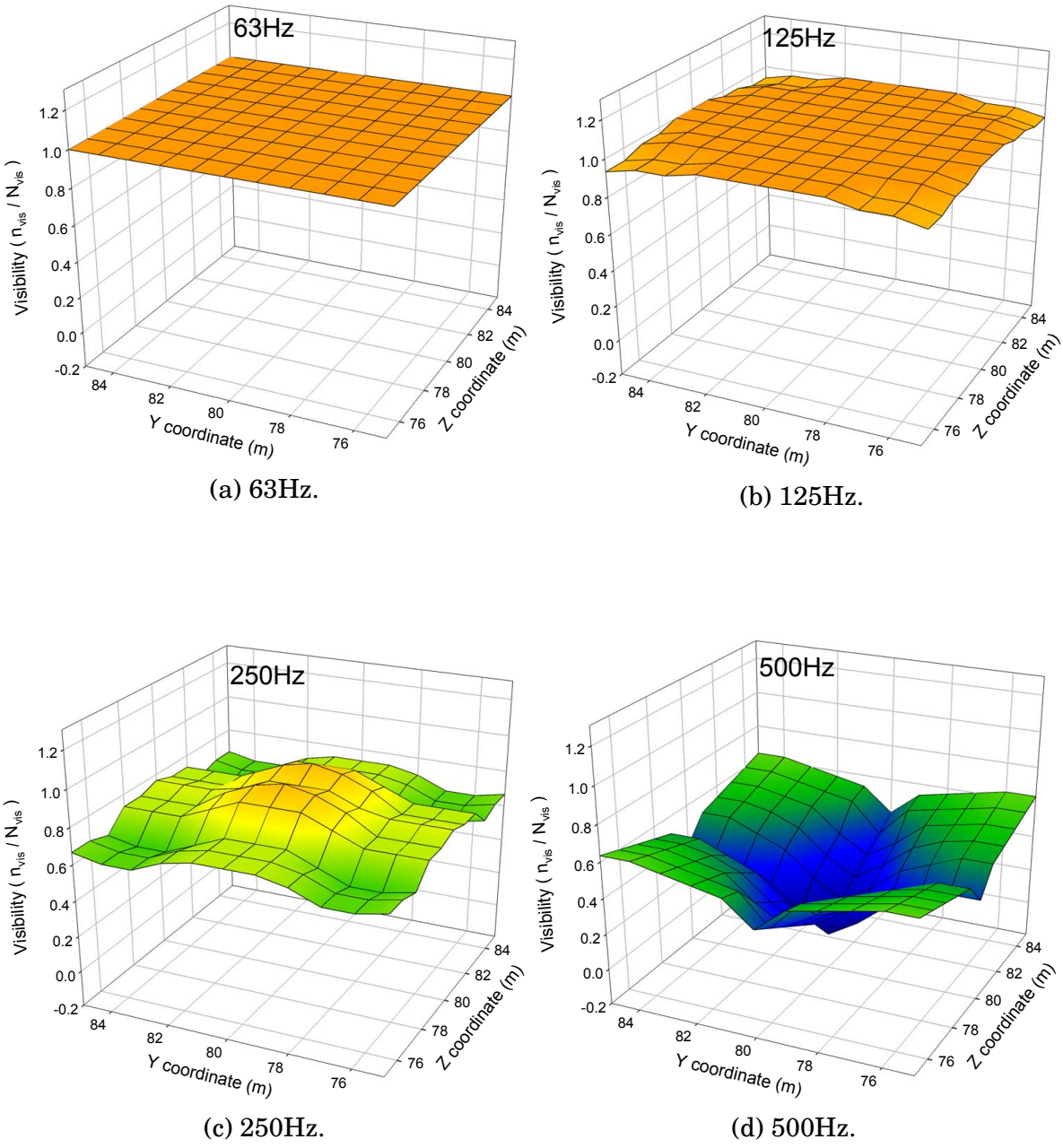


Figure 5.17: Set-up for the diffraction by a non-infinite edge simulation.

Visibility of the first Fresnel zone was calculated by averaging the visibility associated with each of the four edge positions.

The results of this simulation for the 63Hz, 125Hz, 250Hz, 500Hz, 1000Hz, 2000Hz, 4000Hz and 8000Hz frequencies are illustrated in Figures 5.18 and 5.19 where the visibility of the first Fresnel zone relative to the receiver is plotted as a function of receiver position. As shown in Figure 5.18(a), the visibility of the first Fresnel zone for the 63Hz frequency for each receiver position was equal to one indicating the first Fresnel zone was completely visible to the receiver for all receiver positions. As frequency is increased, visibility decreased until it became zero beyond 2000Hz since the first Fresnel zone was completely blocked by the occluder irrespective of the receiver's position (see Figures 5.19(a)-(d)).



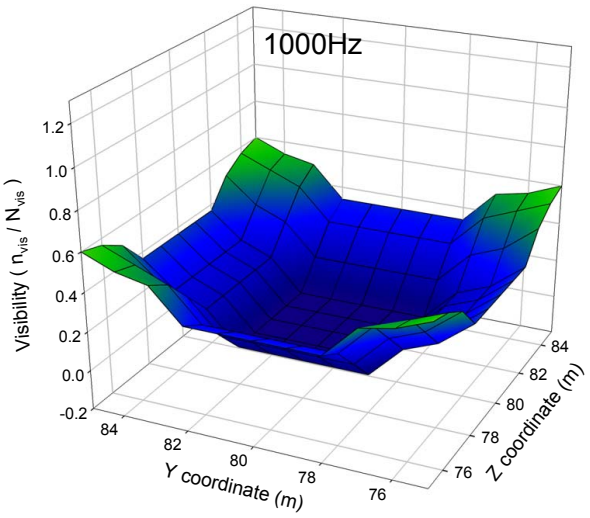
(a) 63Hz.

(b) 125Hz.

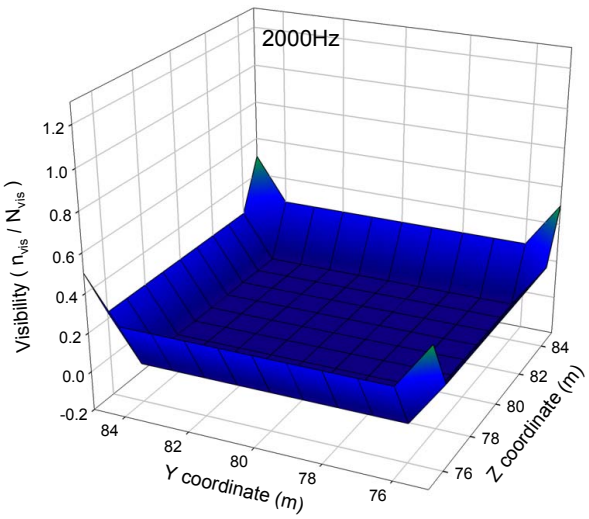
(c) 250Hz.

(d) 500Hz.

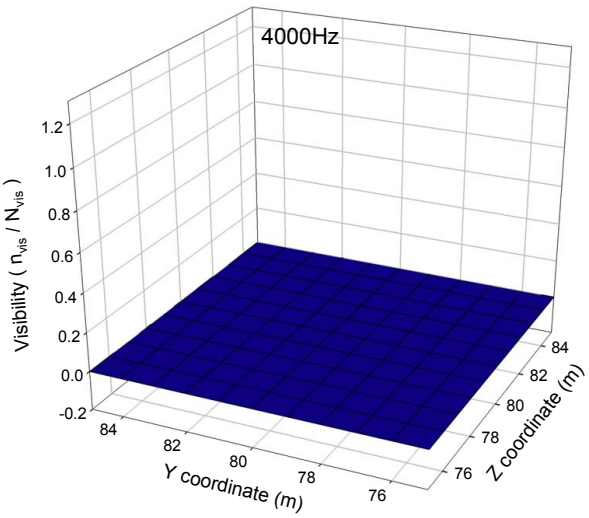
Figure 5.18: Diffraction by a non-infinite edge simulation: visibility as a function of frequency. (a) 63Hz, (b) 125Hz, (c) 250Hz and (d) 500Hz.



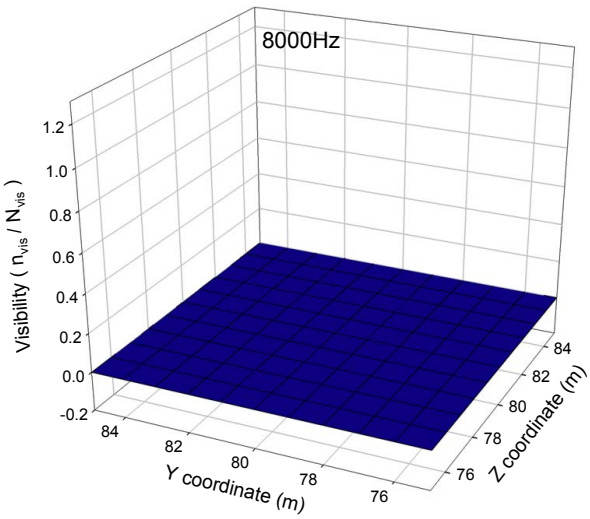
(a) 1000Hz.



(b) 2000Hz.



(c) 4000Hz.



(d) 8000Hz.

Figure 5.19: Diffraction by a non-infinite edge simulation: visibility as a function of frequency. (e) 1000Hz, (f) 2000Hz, (g) 4000Hz and (h) 8000Hz.

5.1.4.5 Diffraction Running Time Requirements

Running time requirements are an important aspect of any simulation method to be used in real-time applications. Therefore, this simulation examines the running time requirements the Huygens-Fresnel acoustical diffraction modeling approach. Both implementations (first Fresnel zone only and all Fresnel zones) were considered. The simulation was performed for the configuration shown in Figure 5.17 of the previous simulation. The sound source and edge position were constant at positions (65, 80, 80) and (85, 82, 85) respectively while the receiver position varied across the y and z coordinates (e.g., a plane of receiver positions with y and z beginning at position (85, 75, 75) and ending at position (85, 85, 85)). The results of this simulation are summarized in Table 5.8 and Figure 5.20 where the average running time and standard deviation for each frequency band (obtained over 225 measurements) to compute the diffraction modeling are given. Also included in the table are the running times for the diffraction approach where all of the Fresnel zones were considered.

When considering the first Fresnel zone only, the difference in running time from the smallest (11.42ms for the 200Hz center frequency) to the largest running time (12.27ms for the 125Hz center frequency) is 0.85ms and therefore, running time is approximately constant across frequency. In contrast, the running time when considering all Fresnel zones increases linearly with frequency, ranging from 16.90ms (63Hz) to 283.42ms (8000Hz). In addition to the first

Freq. (Hz)	One Zone		All Zones	
	Time (ms)	σ	Time (ms)	σ
63	12.04	2.55	16.09	2.76
125	12.27	2.49	16.36	3.91
250	11.87	2.83	19.73	3.34
500	11.96	2.88	28.04	3.49
1000	12.04	2.86	45.69	4.46
2000	11.42	2.45	80.18	6.89
4000	12.04	2.10	149.51	8.31
8000	11.69	1.98	283.42	13.22

Table 5.8: Results for the diffraction computation time requirements simulation: average diffraction running times along with standard deviation (σ) as a function of frequency for both diffraction implementations (e.g., one Fresnel zone and all Fresnel zones considered).

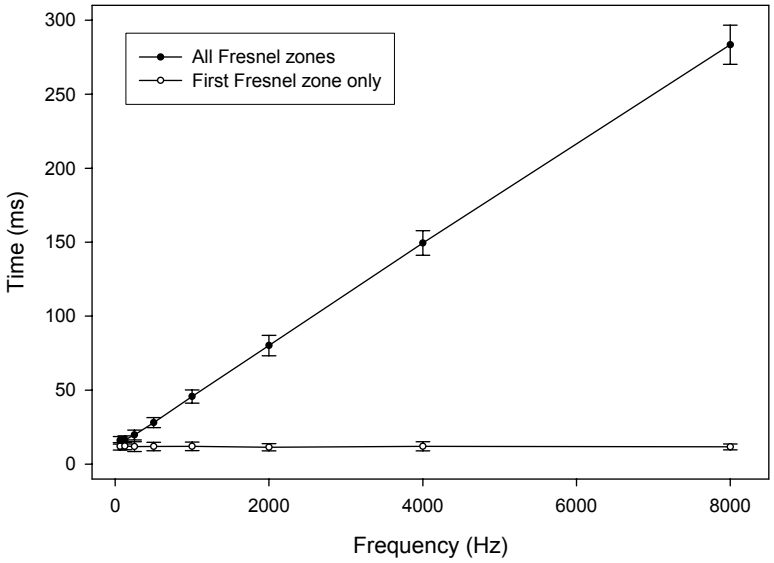


Figure 5.20: Results for the diffraction running time requirements simulation: average diffraction modeling running time vs. frequency with error bars (standard deviation).

Fresnel zone only implementation providing more accurate results (as demonstrated in the simulation described in Section 5.1.4) as opposed to the implementation where all zones are considered, its running time requirements are much less and constant across frequency.

5.2 Sonel Mapping as a "Whole"

The simulations presented in this section consider the various components comprising the sonel mapping algorithm working together as a complete system. In the first simulation, the sound propagation of a simple room with the presence of an occluder (edge) is simulated. The second simulation presents a graphical illustration of various aspects of the sonel mapping algorithm.

5.2.1 Simple Room Simulation

In this simulation, the sound propagation in the quasi-cubic enclosure illustrated in Figure 5.21 was simulated in order to examine the effect of altering the number of sonels emitted from the sound source on the recorded echogram[3]. All possible sonel-surface interactions were considered (e.g., specular and diffuse reflections, diffraction and absorption) in any combination. The simulation was performed for frequency bands with center frequency values of 125Hz, 250Hz, 500Hz, 1000Hz, 2000Hz and 4000Hz and for the $10,000$,

[3]These results originally appeared in Kapralos et al. [92]

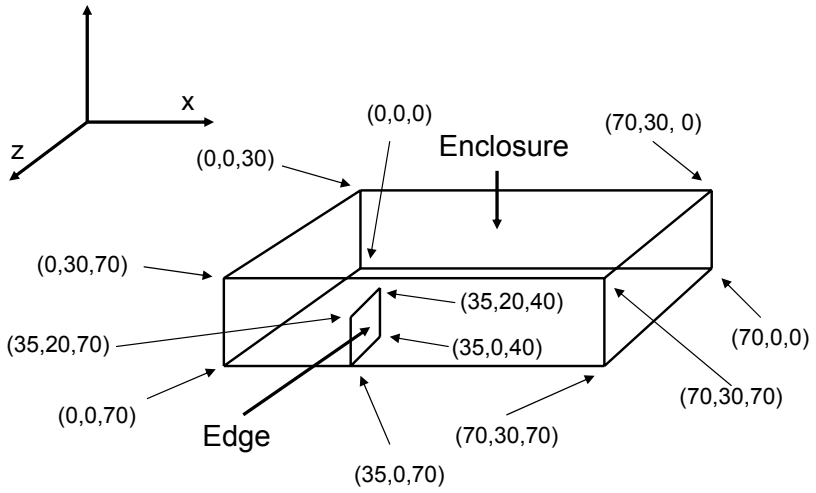


Figure 5.21: Set-up for the simple room simulation.

$100,000$, $500,000$, $1,000,000$ and $2,000,000$ sonels being emitted at the sound source and a corresponding number of acoustical visibility rays traced from the receiver. For each simulation, the total sound level arriving at the receiver over a brief interval of time (three seconds) was measured as was the time taken to compute the simulation. The dimensions of the box-like room were 70m $\times$ 15m $\times$ 70m, the position (x,y,z coordinates, in meters) of the single omni-directional sound source and single receiver were $(15,10,55)$ and $(60,9,60)$ respectively. For each frequency band considered, sound source energy was divided equally amongst all sonels emitted. The surfaces of the enclosure (four walls, ceiling and floor) were each assigned an absorption coefficient value of $\alpha = 0.15$. The diffuse and specular coefficients were set to a value equal to $(1-\alpha)/2$. A summary of the simulation results are displayed in Figures 5.22 and 5.23.

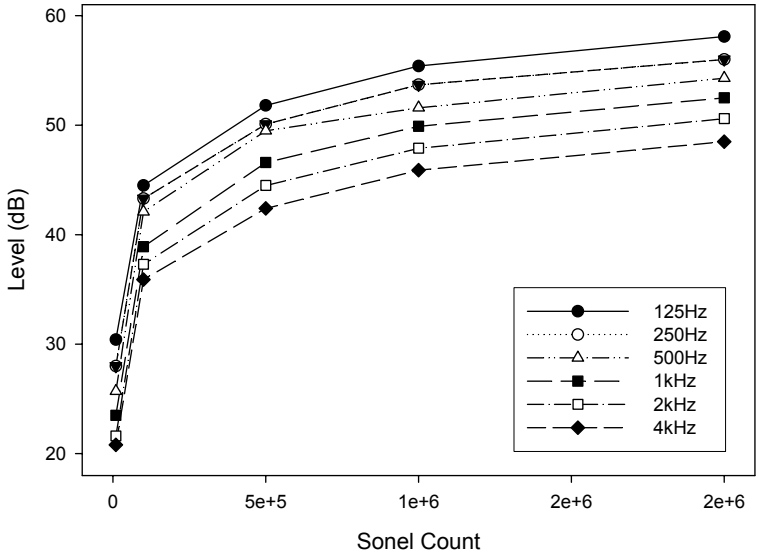


Figure 5.22: Results for the simple room simulation: receiver level as a function of sonel count for different frequencies.

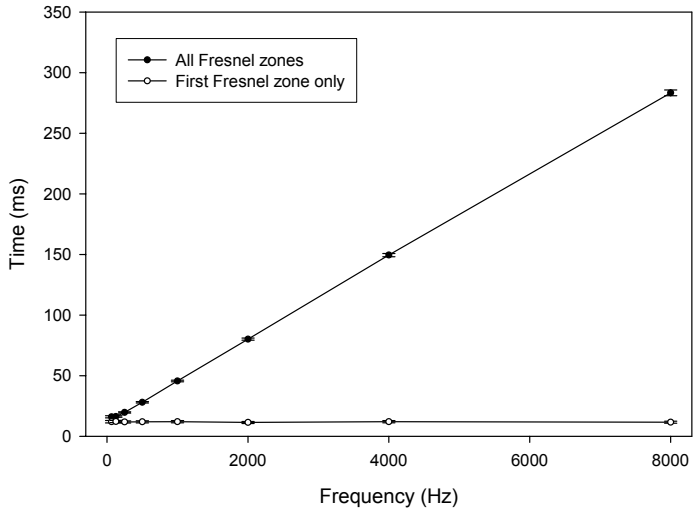


Figure 5.23: Average simulation time as a function of frequency band.

Sonel Count	**Time (s)**	σ
10,000	1.25	0.05
100,000	13.25	0.05
500,000	66.12	0.45
1,000,000	132.72	0.19
2,000,000	266.07	2.86

Table 5.9: Results for the simple room simulation: sonel count vs. average simulation time (averaged across each frequency band) along with standard deviation (σ).

5.2.2 Graphical Illustration

In this section, a graphical demonstration is provided illustrating the sound energy propagation in a particular environment (room) for various sound source, receiver and edge configurations. The dimensions of the room were 10m $\times$ 8m $\times$ 10m (see Figure 5.24). The sound source was positioned at location (1,1,1) and remained stationary throughout the simulation. The dimensions of the occluder were 6m $\times$ 6m, and when present, was positioned such that it formed a plane along the x-z axis. The coordinates of the vertices comprising the edge were (5,0,0), (5,0,6), (5,6,6) and (5,6,0). The absorption coefficient of each surface was arbitrarily set to 0.1 while the diffuse and specular reflection coefficients were each set to 0.45. Sound source energy (90dB) was equally divided amongst 500,000 sonels and a single frequency (63Hz) was considered. Three scenarios were considered: i) absence of the occluder and therefore no diffraction, ii) presence of the occluder but diffraction effects ignored and iii) presence of the occluder and diffraction effects accounted for. For each scenario, the sound source remained stationary while the position of the single receiver was varied across a plane along a portion of the x-z axis of constant height (e.g., y-coordinate remained constant at y = 5). Both x and z coordinates were varied in increments of 0.5m beginning at 5.5m and ending at 9.5m. For each of the three scenarios, the sonel tracing stage was performed at the start of the simulation and since the position of the sound source remained static, the sonel

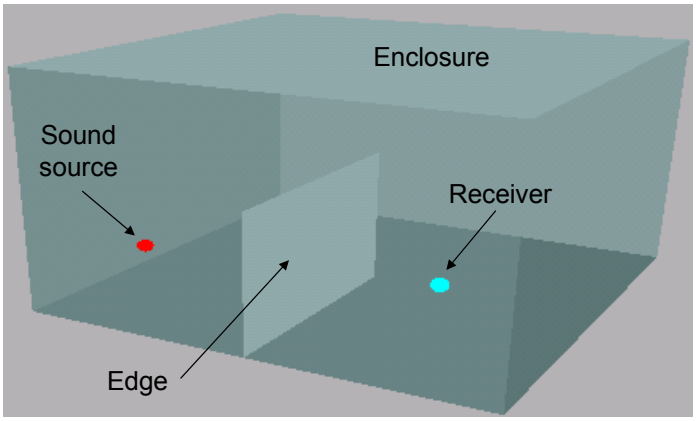


Figure 5.24: Graphical illustration of sonel mapping: set-up (sound source, sample receiver position and edge configuration). The sound source remained stationary while the position of the receiver was varied across a plane along a portion of the x-z axis of constant height.

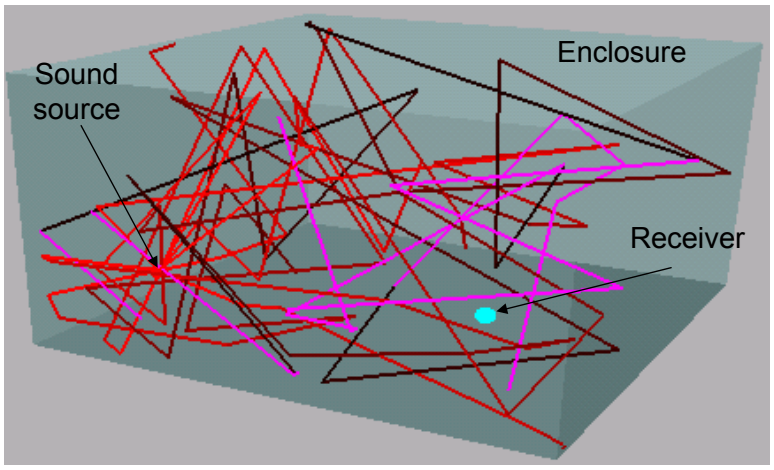
tracing stage was not performed again at any time during the simulation. For each receiver position, the acoustical rendering stage was performed (20,000 acoustical visibility rays were traced out from each receiver position) and the echogram obtained.

A graphical sample of the simulation for the first scenario (absence of an occluder) is provided in Figure 5.25(a) where the paths of the sonels initially emitted from the sound source are traced through the environment. Sonel reflection count is encoded by color whereby red indicates a reflection count of one and darker shades of red indicate greater reflection counts. For illustration purposes, this figure was obtained by emitting only 10 sonels from the sound source. A sample echogram estimated at one of the receiver positions is provided in Figure 5.26(a) while a plot illustrating the sound level at each receiver position as estimated over a three second time interval is shown in Fig-

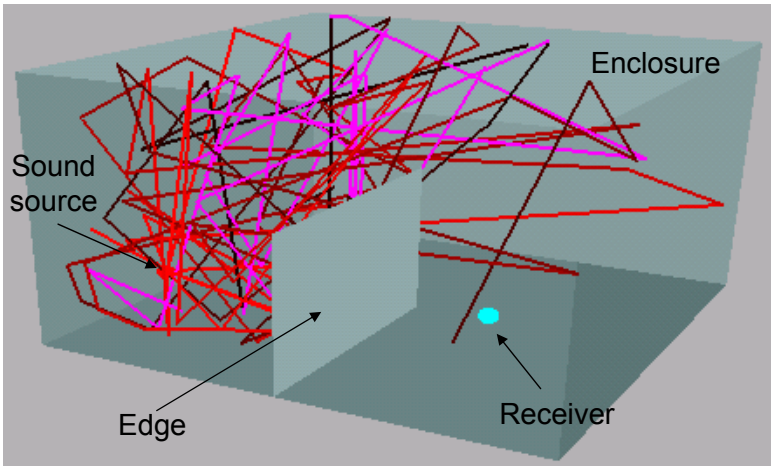
ure 5.26(b). The x-axis of the contour plot represents the receiver's x-coordinate while the y-axis represents the receiver's z-coordinate. Sound level is encoded by color whereby brighter colors denote higher energy levels.

In the second scenario, an occluder was placed between the sound source and receiver such that the direct path between them was occluded however, diffraction effects were ignored. A graphical sample of the simulation for this scenario is provided in Figure 5.25(b) where the paths of the sonels initially emitted from the sound source were traced through the environment. As with the first scenario, sonel reflection count (order) is encoded by color whereby red indicates a reflection count of one and darker shades of red indicate greater reflection counts. As before, this graphical simulation sample was obtained by emitting 10 sonels from the sound source. A sample echogram estimated at one of the receiver positions is provided in Figure 5.27(a) while a plot illustrating the sound level at each receiver position as estimated over a three second time interval is shown in Figure 5.27(b).

In the third scenario, the occluder was present as above but diffraction effects were considered. A graphical sample of the simulation for this scenario is provided in Figure 5.28. As in the previous sample simulations, to avoid clutter for the purposes of illustration, only 10 sonels were emitted and traced from the sound source. In Figure 5.28(a), only diffraction effects are included. Three sonels were diffracted and in each case, the edge position is indicated by

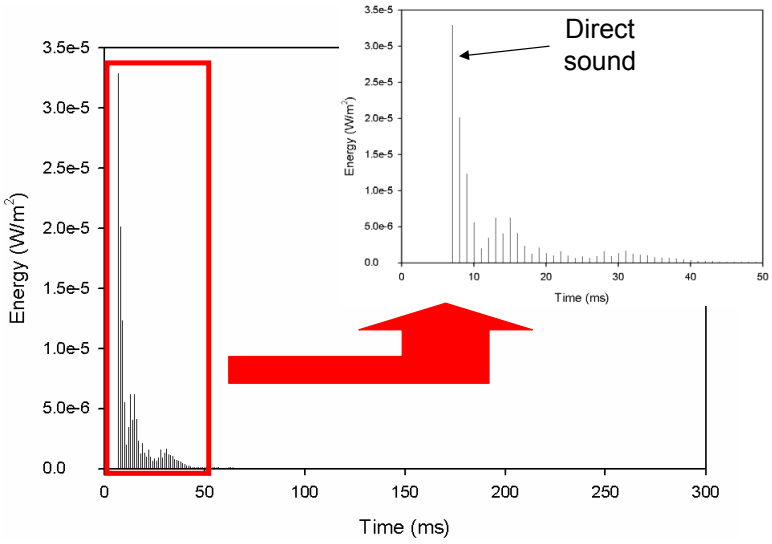


(a) Absence of the occluder.

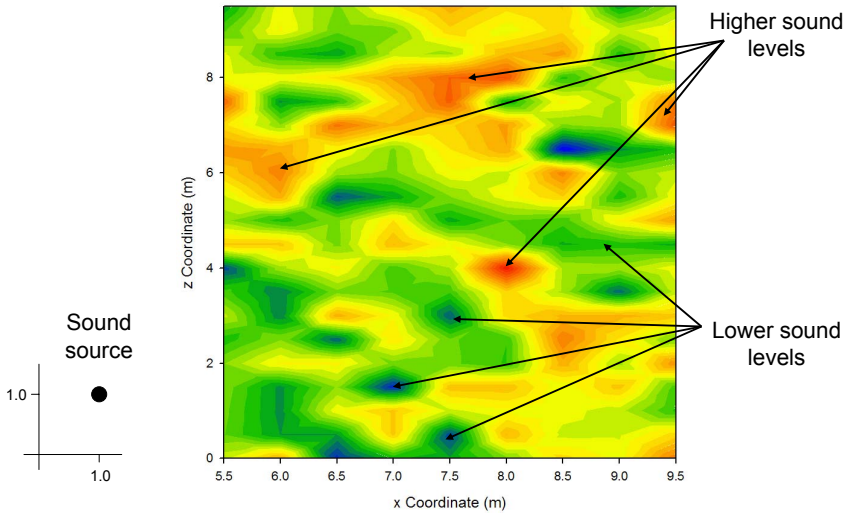


(b) Presence of occluder but diffraction effects ignored.

Figure 5.25: Graphical illustration of sonel mapping: sample simulations. (a) Absence of the occluder and (b) presence of the occluder but diffraction effects ignored. Reflection count is encoded by color whereby red indicates a reflection count of one and darker shades of red denote higher reflection counts.

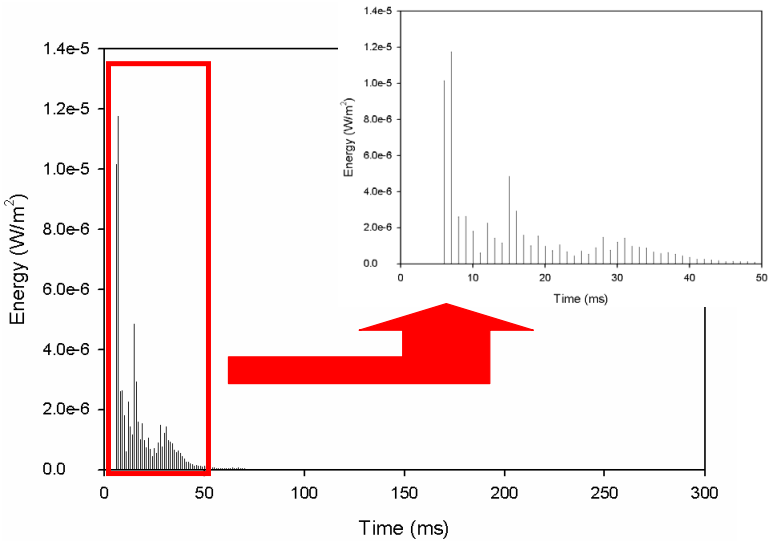


(a) Echogram: absence of the occluder.

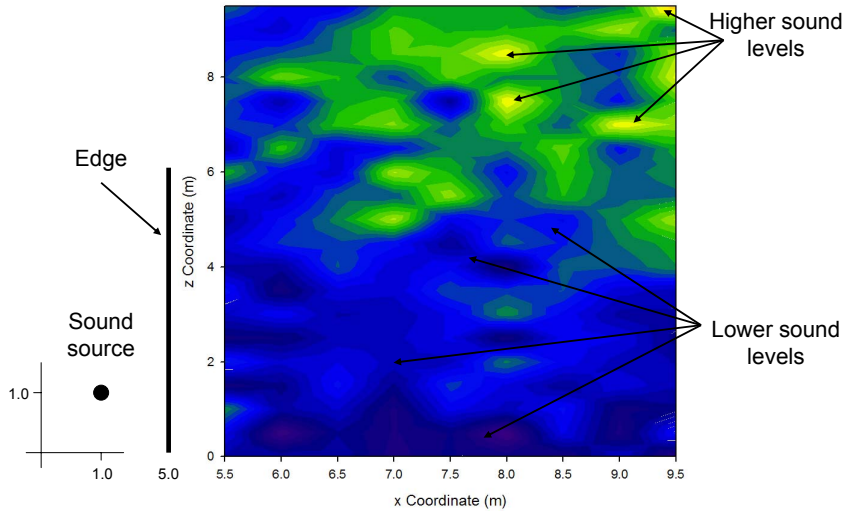


(b) Contour plot: absence of the occluder.

Figure 5.26: Graphical illustration of sonel mapping: echogram and contour plot (absence of the occluder). (a) Echogram, where the echogram spacing (e.g., spacing between echogram "bins") is 5ms and the total length is 3s. Included with the echogram is an inset illustrating the first 50ms of the echogram in greater detail. (b) Contour plot. Sound level over a plane of receiver positions at constant height (sound level is encoded by color whereby, brighter colors indicate greater levels).



(a) Echogram: presence of the occluder but no diffraction effects.

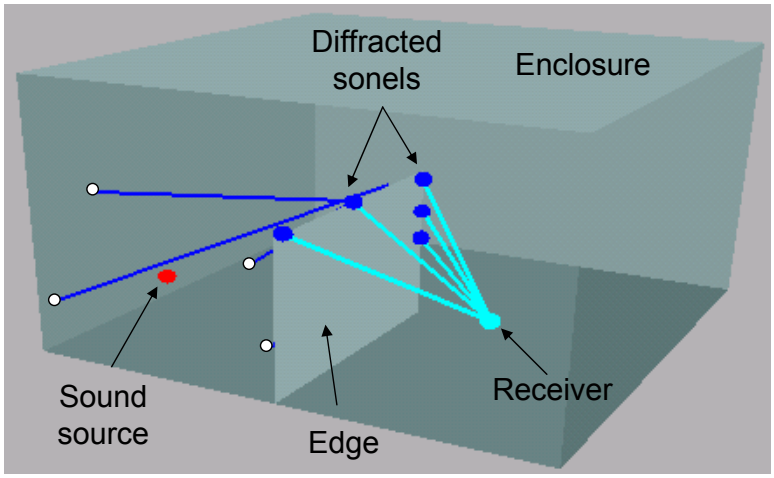


(b) Contour plot: presence of the occluder but no diffraction effects.

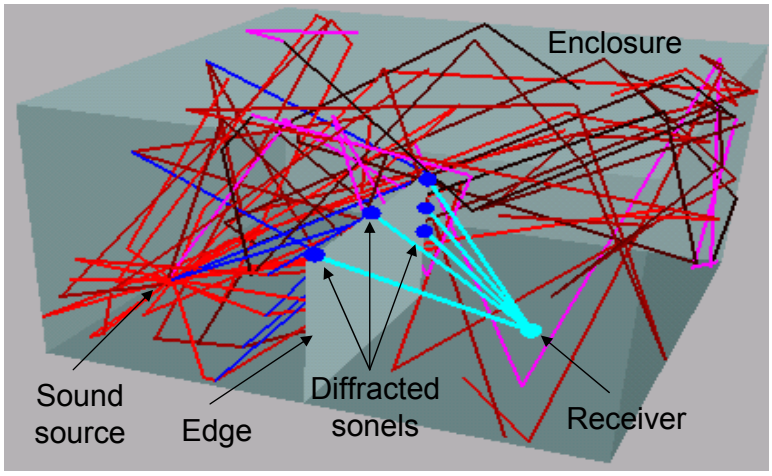
Figure 5.27: Graphical illustration of sonel mapping: echogram and contour plot (presence of the occluder with no diffraction effects). (a) Echogram, where the echogram spacing (e.g., spacing between echogram "bins") is 5ms and the total length is 3s. Included with the echogram is an inset illustrating the first 50ms of the echogram in greater detail. (b) Contour plot. Sound level over a plane of receiver positions at constant height (sound level is encoded by color whereby, brighter colors indicate greater levels).

a blue circle. Each path from the edge position to the receiver is colored light blue and the path from the sound source to the edge position is colored dark blue (in this example, there is no direct path from the sound source itself to the receiver but rather, the sonel arrives to the edge via one or more specular or diffuse reflections). In Figure 5.28(b), in addition to illustrating diffraction effects alone, specular and diffuse reflection paths are also included. A sample echogram estimated at one of the receiver positions is provided in Figure 5.29(a) while a plot illustrating the sound level at each receiver position as estimated over a three second time interval is shown in Figure 5.29(b).

Although the purpose of the simulation was to illustrate the operation of various aspects of sonel mapping, several observations confirming the correct operation of sonel mapping can be made. In particular, in the absence of an occluder, sound energy was distributed throughout the room. When the occluder was present in the absence of diffraction, a decrease in sound level behind the occluder occurred as expected given that the occluder will occlude many sonel paths from reaching the receiver beyond the occluder. Finally, when the occluder was present and diffraction was also present, sound level for receiver positions beyond the edge was greater than the sound level at similar locations when edge effects were ignored. This is also as expected given that diffracted sound reaches a receiver despite the fact that the direct path between the sound source and receiver was occluded.

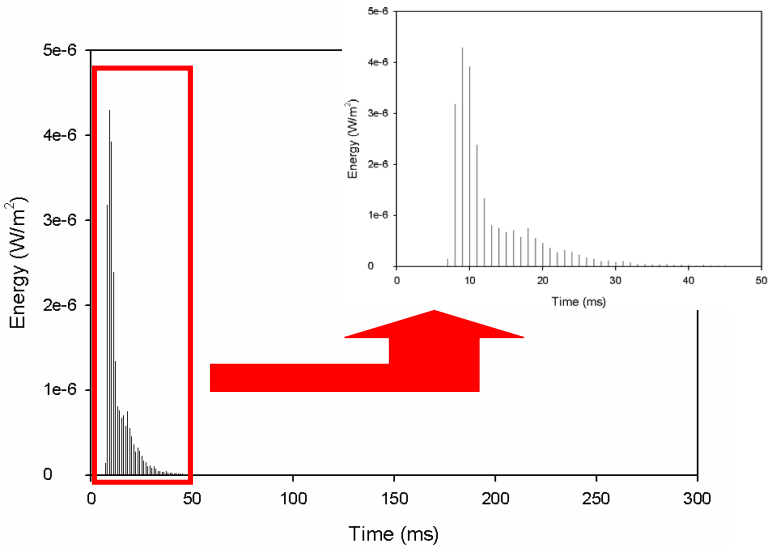


(a) Edge effects in isolation.

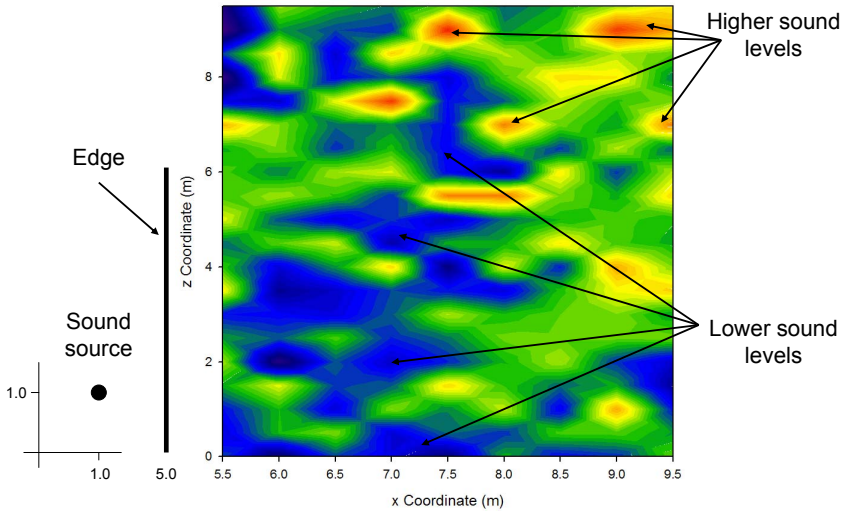


(b) Edge effects along with specular and diffuse reflections.

Figure 5.28: Graphical illustration of sonel mapping: sample diffraction simulation. In this illustration, 10 sonels were emitted and traced from the sound source allowing for specular and diffuse reflections, absorption and diffraction. Only diffracted sonel paths are illustrated. Three sonels were diffracted and in each case, the position of the occluder is indicated by a blue circle. Each path from the occluder (edge)point (p_{edge}) position to the receiver is colored light blue and the path from the sound source to the occluder position is colored blue (in this example, there is no direct path from the sound source itself to the edge but rather, the sonel arrives to the edge via one or more specular or diffuse reflections). (a) Diffraction effects are shown in isolation and (b) diffraction effects along with specular and diffuse reflections.



(a) Echogram: presence of the occluder and diffraction effects.



(b) Contour plot: presence of occluder and diffraction effects.

Figure 5.29: Graphical illustration of sonel mapping: echogram and contour plot (presence of edge with diffraction effects accounted for). (a) Echogram, where the echogram spacing (e.g., spacing between echogram "bins") is 5ms and the total length is 3s. Included with the echogram is an inset illustrating the first 50ms of the echogram in greater detail. (b) Contour plot. Sound level over a plane of receiver positions at constant height. (b) Contour plot. Sound level over a plane of receiver positions at constant height (sound level is encoded by color whereby, brighter colors indicate greater levels).

5.3 Summary

In this chapter, various simulations were presented that examined the operation and effectiveness of the sonel mapping algorithm. The simulations were divided into two categories i) individual component simulations and ii) simulations involving the complete algorithm. The component testing examined the various components comprising the sonel mapping algorithm in isolation thus allowing meaningful analysis on the correctness of their operation. The first simulation in this category presented a graphical illustration of the various components comprising sonel mapping and compared simulation results with theoretical predictions. For the second and third simulations in this category ("reverberation time: simulated vs. theoretical results" and "Russian roulette: a comparison to a deterministic approache"), differences in reverberation time, as predicted by Sabine's formula and calculated by the sonel mapping algorithm using either a Russian roulette or deterministic termination criterion, were used as the performance measure. The remainder of the simulations in the component testing category were concerned with the acoustical diffraction implementation. In particular, in the "correctness of the acoustical diffraction method" simulation, the performance measure was the difference between the energy arriving at a receiver from a sound source as calculated using the acoustical diffraction implementation and the energy as predicted using a stan-

dard harmonical spherical wave expression. The following three diffraction simulations considered the visibility of the first Fresnel zone and/or receiver sound level for various sound source, edge and receiver configurations (the "first Fresnel zone visibility as a function of receiver height", "first Fresnel zone visibility over a plane of receiver positions" and "diffraction by a non-infinite edge" simulations) while the last diffraction simulation ("diffraction running time requirements") examined the timing requirements of the acoustical diffraction implementation. Both versions of the algorithm were examined. In particular, a comparison was made between the timing requirements for the implementation where the first Fresnel zone was considered only and the implementation where all Fresnel zones were considered.

In the second category of simulations, sonel mapping was considered as a whole. For the first simulation in this category (the "single room" simulation), a rectangular environment with a single diffracting edge was simulated for a particular sound source and receiver configuration. In this simulation receiver level and simulation time were measured for various number of sonels emitted from the sound source. In the second simulation, a graphical illustration of the various components of sonel mapping operating in conjunction was presented. In particular, the propagation of sound energy of a particular environment was simulated for various sound source, receiver and occluder configurations.

Chapter 6

Discussion

This dissertation presented the sonel mapping acoustical modeling algorithm. Sonel mapping is the first application of the widely used and computationally efficient photon mapping technique to acoustical modeling. Sonel mapping uses the same basic approach as photon mapping but takes into account the special physical attributes of sound propagation and addresses the possible interactions when a propagating sound encounters a surface/object or obstruction in its path.

Sonel mapping is a framework for energy propagation despite the fact that in this dissertation, it is used to model the acoustics of a particular environment. Given the appropriate model parameters (e.g., source emission functions, wave/surface interaction models etc.), it can be used to model the propa-

gation of any type of wave energy, be it sound, light etc., including any type of interaction between the wave and the medium in addition to the geometry and properties of any surfaces/objects the wave may encounter. Given a particular model of wave (energy) propagation, sonel mapping can be used to simulate and recreate the particular wavefield. Sonel mapping takes advantage of the fact that in many situations, the environment remains static (e.g., no movement of the sound source, receiver or objects within the environment) for some time. Exploiting this static property by storing and reusing information regarding the sound energy in an environment within the sonel map, and relying on probabilistic methods to avoid potentially complex and computationally expensive operations required to model the propagation of sound energy, sonel mapping can be used at interactive rates.

6.1 Probabilistic-Based Sonel Mapping

As with photon mapping, sonel mapping is a probabilistic based approach employing Monte-Carlo ray tracing and a Russian roulette strategy throughout the various stages of its execution. With Monte-Carlo ray tracing, *point sampling* is used to provide an estimate of the sound energy in a model. With Russian roulette modeling, the determination of the type of interaction that occurs when a sonel encounters a surface is determined probabilistically based on the characteristics of the both the surface and the sonel. The advantages of

Monte-Carlo based approaches over deterministic ones include [84]:

- No tesselation of the environment required (arbitrary geometry and therefore no “meshing”).
- Low memory consumption.
- Support procedural geometry.
- Complex reflection models are supported.
- No need to precompute a representation for the solution.
- Solution is correct except for variance (noise).

At each sonel/surface interaction point, a check is made to determine whether the sonel is incident within the diffraction or non-diffraction zone. When the sonel is within the diffraction zone, the sonel is diffracted using a modified version of the Huygens-Fresnel principle. However, no additional sonels are created/emitted from the point of diffraction. When the sonel is incident within the non-diffraction zone, one of three types of interactions (absorption, specular reflection or diffuse reflection) is chosen probabilistically using a Russian roulette strategy. Using a Russian roulette strategy, a single interaction occurs at each sonel/surface interaction point as opposed to the multiple interactions inherent with many deterministic approaches. Russian roulette is widely used in image synthesis applications and is a key component of photon mapping, providing a computationally efficient alternative to deterministic approaches. With respect to auralization and acoustical modeling in particular, Russian roulette is a rarely sought approach.

6.1.1 Justification for the Use of Russian Roulette

Given the slow propagation speed of sound in air, time is an important component in any acoustical modeling system [146, 163]. Since the probability of tracing arbitrary long paths decreases as the number of sonel/surface interactions increase, the probability that a sonel is not terminated also decreases with time. This may lead to an inaccurate representation of the estimated room impulse response since the latter portion of the estimated room impulse response will contain very few samples. This in turn might lead to lower estimated reverberation times than predicted by Sabine's formula [200]. The use of Russian roulette in acoustical modeling must be evaluated on the basis of comparisons to other possible approaches.

Limitations with a Reflection Count Termination Criterion: A termination criterion based on a reflection count has its limitations as well. In particular, the reflection count must be set to a large value to ensure paths of arbitrary length are traced. Setting the reflection count to a very large value is clearly impractical due to memory and computational limitations (e.g., an increase in the reflection count leads to a direct increase in the amount of memory and computation time required by the simulation). As a result, when used as a termination criterion, the reflection count must be kept at a manageable level. Therefore there will be paths that are not traced at all, also leading to a non-linear decay of sound over time and therefore shorter than predicted rever-

beration times. With Russian roulette, although the probability of tracing a particular path of length n decreases as the number of times the sonel is reflected increases, paths of arbitrarily long lengths can nevertheless be traced given the probabilistic nature of the algorithm.

Limitations with an Energy Termination Criterion: A termination criterion based on a minimum energy content such as the energy discontinuity percentage or EDP (that represents the percentage of the original ray energy that must be lost before the ray is terminated), shares the same limitations as the reflection count termination criterion. Furthermore, when assuming diffuse reflections only, an EDP value can be directly converted to a corresponding reflection count (see [52]). The EDP must be set to a sufficiently large value to ensure paths of arbitrary length are traced. However, setting the EDP to a very large value is clearly impractical due to memory and computational limitations. Hence, as with the use of a reflection count termination criterion, there will be paths that are not explored. This was demonstrated in the reverberation time simulation described in Section 5.1.1.5. Referring to Tables 5.5 and 5.6 presented in Section 5.1.1.5, the maximum reflection count for each of the reverberation time estimates computed using Russian roulette were larger than the corresponding EDP based measures (the average reflection count with an EDP-based termination criterion over each of the EDP measurements was 24 whereas with the Russian roulette termination criterion the average reflec-

tion count was 76).

Increasing Accuracy by Increasing the Number of Sonels: Consider an environment where the absorption coefficient of each surface is α. Using a Russian roulette strategy, the probability of an incident sonel being reflected diffusely or specularly (denoted by p_{ref}) is $p_{ref} = 1 - \alpha$. Hence, at each surface/sonel interaction point, the incident sonel will either be reflected (specularly or diffusely) or absorbed. Which of these interactions does occur can be described by a *Bernoulli trial* [142]. The probability that a particular sonel will be consecutively reflected n times (P_{ref_n}) can be described mathematically as

$$P_{ref_n} = \prod_{i=1}^{n} p_{ref}.$$

This corresponds to generating a sequence of n consecutive random numbers ξ_i, with each $\xi_i \leq p_{ref}$. Assuming $p_{ref} < 1$, the probability of generating this sequence decreases as n increases. With Russian roulette, the probability of tracing a path of length n therefore decreases as n increases. However, by increasing the number of sonels initially emitted by a sound source, the probability of generating such a path of length n can be increased. Essentially, the decreasing probability of generating a path of length n as n increases (when considering a single sonel), can be "counter-balanced" by increasing the number of sonels emitted by a sound source. Mathematically, replacing each sonel originally emitted at the sound source with M sonels, the probability of gener-

ating a single path of length n becomes

$$P_{ref_n} = M \times \prod_{i=1}^{n} p_{ref}.$$

For a given probability level P_{ref_n}, it is possible to choose a value of M such that the probability of generating a path of length n reaches P_{ref_n}. Furthermore, as previously described, as n is increased, the probability of tracing a path of length n decreases, becoming zero as n approaches infinity

$$\lim_{n\to\infty} \prod_{i=1}^{n} p_{ref} = 0.$$

In reality, the path length n of a propagating sound will be finite. In other words, a sound will not propagate indefinitely but will eventually lose all of its energy after a portion of it is absorbed at each reflection point and by the medium. Hence, for any practical application, an appropriate M can always be found to ensure the probability of tracing a sonel until its energy is negligible exceeds a pre-defined threshold value.

Reduced Computation Time: The use of a Russian roulette approach leads to a reduced computation time while still allowing arbitrarily long paths to be traced. An exponential increase in the number of sonels to be traced (assuming two new sonels are reflected at each interaction between a sonel and a surface) is clearly impractical for any real time applications except perhaps in certain

simple, trivial environments. The results of the Russian roulette simulation presented in Section 5.1.1.5 confirm this. In this simulation, reverberation time estimates for a particular environment (sound source and receiver configuration) were computed initially using an energy discontinuity percentage (EDP) termination criterion. Then, for each EDP-based reverberation time estimate, using a Russian roulette termination criterion, the number of sonels initially emitted from the sound source in the sonel tracing stage was adjusted such that the computed reverberation time was equal to the corresponding reverberation time computed with an EDP termination criterion. The signed percent differences between the time taken using the EDP and Russian roulette approach was taken as the measure of performance in this simulation. For all measurements, the percent difference was positive indicating the EDP based method took more time to compute. Differences ranged from 510% to 3570%.

6.2 Diffraction Modeling Using a Modified Version of the Huygens-Fresnel Principle

Since the dimensions of many of the objects/surfaces encountered in our daily life are within the same order of magnitude as the wavelength of audible sounds, diffraction is an elementary means of sound propagation, especially when there is no direct path between the sound source and the receiver [180]. Sonel mapping addresses the modeling of diffraction effects. Acoustical diffrac-

tion is accomplished using a modified version of the Huygens-Fresnel principle [79]. Diffraction effects in the sonel mapping approach are handled by utilizing the Huygens-Fresnel principle to model the acoustical energy reaching a receiver from a given sound source after being diffracted by an edge. Essentially, given a sound source, receiver and edge, the energy reaching the receiver is determined by considering the energy arriving at the receiver from the first Fresnel zone as in the unoccluded scenario. To account for diffraction effects, a visibility factor for the first Fresnel zone is introduced. The visibility factor represents the fraction of the first Fresnel zone visible from the receiver. Essentially, positions on the first Fresnel zone are uniformly sampled and ray casting is used to determine the fraction of the zone visible relative to the receiver. The total visibility of the zone is equal to the fraction of sampled positions where a clear path between the sampled position and the receiver exists versus the total number of positions sampled.

Although various acoustical diffraction modeling techniques have been introduced in the literature (see Section 3.2.3), the majority of these techniques rely on finite element or boundary element methods and are currently not applicable for interactive applications due to complexity issues. Such techniques are therefore not considered further here. That being said, a limited number of research efforts have investigated acoustical diffraction modeling for virtual environment applications. The acoustical diffraction modeling method pre-

sented in this dissertation is similar to the approach previously introduced by Tsingos and Gascuel [181]. In their approach, diffraction is approximated by computing the fraction of sound that is blocked by obstacles between the path from the sound source to the receiver by considering the amount of volume of the first Fresnel ellipsoid as opposed to a sphere, that is blocked by any occluders in the path between the sound source and receiver. A visibility factor is computed using computer graphics hardware. A rendering of all occluders from the receiver's position is performed and a count of all pixels not in the background is taken (pixels that are "set" e.g., not in the background, correspond to occluders). Although experimental results are not extensive, their approach is capable of computing a frequency dependent visibility factor that, unlike other ray-based approaches, takes advantage of graphics hardware to perform this in an efficient manner.

6.2.1 Diffraction Modeling Results in Greater Detail

Simulations were presented that demonstrated the ability of sonel mapping to model diffraction effects in a very simple and efficient manner allowing it to be computed at interactive rates. In the first diffraction simulation (see Section 5.1.4) the correctness of the Huygens-Fresnel principle as implemented in this work was shown. This was performed by calculating the energy reaching the receiver from the sound source for various sound source, receiver and edge configurations using the Huygens-Fresnel implementation (with the visibility of

the first Fresnel zone assumed to be one) and comparing the computed results with the results as calculated using a mathematical expression describing a harmonic spherical wave. The smallest and largest average percentage differences were found to be 0.95 (standard deviation of 0.04) and 1.02 (standard deviation of 0.03) respectively.

This simulation was repeated by considering all Fresnel zones instead of only the first Fresnel zone as done in sonel mapping (see Section 4.4.1.4). The energy reaching the receiver was determined by dividing the entire sphere representing the initial wavefront after being emitted from the sound source and traveling some distance, into a number of Fresnel zones and summing the energy reaching the receiver from each of these Fresnel zones. In this scenario, the range of percent differences for the diffraction implementation were all Fresnel zones were considered was larger, ranging from 0.23 to 5.88 and typically increased with increasing frequency. This increase in percent difference between the actual and computed receiver sound level may be due to numerical errors associated with locating a secondary source in each of the Fresnel zones. As frequency increases, the number of Fresnel zones also increases thus, any errors associated with locating a secondary source in a particular Fresnel zone propagates through (e.g., locating a secondary source in Fresnel zone i requires the position of a secondary source in zone $i-1$ hence, an error in the position of the secondary source in zone $i-1$ may propagate and therefore, result in an

incorrect secondary source position in zone *i*). In addition to providing more accurate results, using the approximation that considers only the first Fresnel zone is computationally more efficient. Running time requirements when considering the first Fresnel zone only are less than the timing requirements for the implementation that considers all Fresnel zones. This is of course directly related to the additional time required to determine the position of a secondary source in each additional Fresnel zone after the position of a secondary zone within the first Fresnel zone is determined as well as calculating the visibility weighting of each additional Fresnel zone relative to the receiver. Furthermore, running time when considering all Fresnel zones is directly proportional to frequency given that the number of Fresnel zones increases with increasing frequency whereas, in the approximation where one Fresnel zone is considered only, running time is constant with respect to frequency. The increase in running time can be considerable, especially for higher frequency sounds. As reported in Section 5.1.4, running times for the implementation whereby the first Fresnel zone was considered only ranged from 11.42ms to 12.27ms while running times ranging from 16.90ms to 283.42ms were observed when all Fresnel zones were considered.

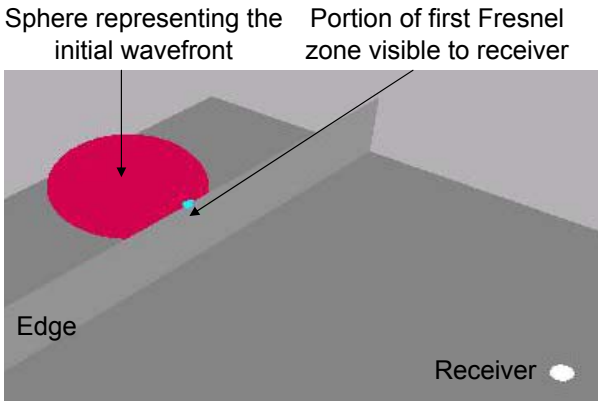
Theoretical diffraction models predict that diffraction increases as the ratio between object size and frequency increases due either to a decrease in frequency or a decrease in object/surface size or both [48]. In other words, lower

frequency sounds are diffracted more. The acoustical diffraction approach presented in this dissertation is in agreement with theoretical diffraction models as demonstrated in several simulations. In the "first Fresnel zone visibility as a function of receiver height" simulation (see Section 5.1.4), the visibility of the first Fresnel zone was determined as a function of frequency. The computed visibility (whose value is between zero and one) was used to scale the unoccluded energy reaching the receiver after being emitted from the sound source and therefore, there is a direct relationship between visibility and energy reaching the receiver (e.g., as visibility increases, the energy reaching the receiver increases as well).

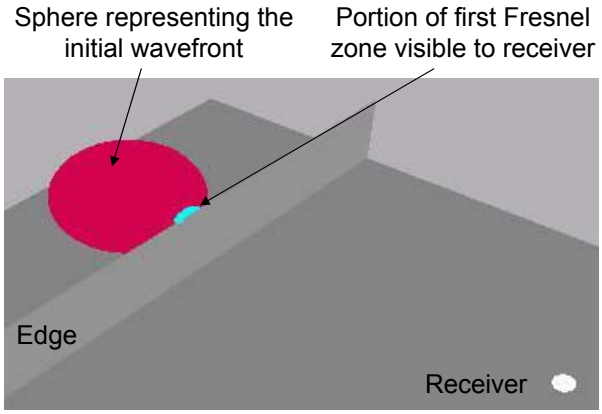
A stationary sound source and occluder (edge) were positioned at the same height (with respect to the y-axis), while the receiver was positioned at three locations (with respect to the y-axis): i) below the edge position, ii) at the same height as the edge position and iii) above the edge position. The results for the configuration considered in the first scenario (e.g., receiver below the edge position) are as expected. In particular, the visibility of the first Fresnel zone is inversely proportional to frequency whereby, as frequency increases, visibility decreases. The decrease in visibility is due to a decrease in the size of the first Fresnel zone. A decrease in visibility also leads to a decrease in the sound energy reaching the receiver. As a result, as frequency increases, the sound energy reaching the receiver decreases. This conforms to the theoretical

model that predicts lower frequency sounds (and therefore longer associated wavelengths) are diffracted more than higher frequency sounds [48]. The results of the second scenario where the receiver is positioned at the same height as the edge position are also as expected. The visibility is approximately 0.5 irrespective of frequency indicating that half of the zone is visible relative to the receiver. Finally, in the third scenario where the height of the receiver was greater than the height of the occluder, visibility and frequency share a direct relationship whereby visibility increases with increasing frequency. This is due to the fact that as frequency increases, Fresnel zone size decreases and therefore, when the height of the receiver is greater than the height of the edge, less of the Fresnel zone will be occluded relative to the receiver. This is illustrated in Figures 6.1 where the visibility of the first Fresnel zone is illustrated for receiver below the edge position (Figure 6.1(a)), at the same height as the edge position (Figure 6.1(b)) and above the edge position (Figure 6.1(c)), for a 4000Hz sound. The increase in visibility associated when the receiver is positioned above the edge is clearly evident.

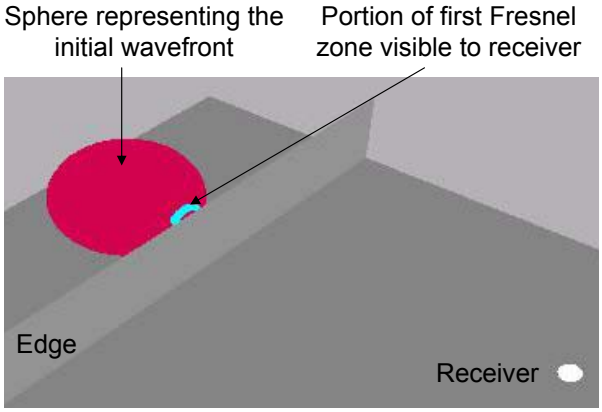
The simulations described above were generalized by keeping a similar configuration but considering a plane of receiver positions (see the “first Fresnel zone visibility over a plane of receiver positions” simulation of Section 5.1.4). For positions where the receiver height (y-coordinate) was less than the height of the occluder, visibility decreased with increasing frequency as expected and



(a) Receiver height less than edge height.



(b) Receiver height equal to edge height.



(c) Receiver height greater than edge height.

Figure 6.1: Fresnel zone visibility for a 4000Hz sound for a stationary sound source and occluder position and varying receiver height (y-axis coordinate). (a) Receiver height (y-axis) less than the occluder height (y-axis), (b) receiver height (y-axis) equal to occluder height (y-axis) and (c) receiver height (y-axis) greater than occluder height (y-axis).

as confirmed in the previous simulation. The decrease in visibility was due to a decrease in the visibility associated with increasing frequency. For positions where the receiver height was greater than the height of the occluder, visibility increased with increasing frequency, once again, as described in the previous simulation, due to the fact that as frequency increases, Fresnel zone size decreases and therefore, when the height of the receiver is greater than the height of the occluder, less of the Fresnel zone will be occluded relative to the receiver.

In the final visibility based simulation ("diffraction by a non-infinite edge" of Section 5.1.4.4), a "non-infinite" plane surface (occluder) with dimensions 2m $\times$ 2m was placed between the sound source and receiver. The position of the sound source remained stationary while the position of the receiver varied in one meter increments across the "y-z" plane.

As shown in Figure 5.18, the visibility of the first Fresnel zone for the 63Hz frequency for each receiver position was equal to one indicating the first Fresnel zone was completely visible to the receiver for all receiver positions. This is expected given that the wavelength of a 63Hz sound wave is 5.4m, much larger than the 2m $\times$ 2m edge plane in the path between the sound source and receiver (see Table 6.1 for the corresponding wavelength of each frequency considered). For the 125Hz frequency, the first Fresnel zone was completely visible when the receiver was positioned at locations corresponding to the cen-

ter of the plane but visibility decreased slightly as the receiver moved away from the center of the edge plane with respect to the y-z axis. As frequency was increased beyond 125Hz, visibility of the first Fresnel zone increased slightly as the receiver was moved away from the center of the edge (with respect to the y-z axis). Although the wavelength corresponding to a 250Hz frequency sound is 1.37m (less than the dimensions of the edge plane), the first Fresnel zone was still partially visible when the receiver was located at or near the center of the edge plane (with respect to the y-z axis e.g., when the receiver was located at y=80 and z=80, the visibility is 0.93). This is due to the fact that the first Fresnel zone is a ring with a radius greater than zero extending beyond the 2m $\times$ 2m occluder. For the frequencies of 1000Hz and 2000Hz, visibility was lowest when the receiver's y and z coordinates equaled the y and z coordinates of the sound source (e.g., 80 and 80 respectively corresponding to the center of the occluder) and increased slightly as the receiver moved away from the center of the occluder. This indicates that a small portion of the first Fresnel zone was visible to the receiver as the receiver moved away from the occluder. However, beyond 2000Hz visibility was zero for all receiver positions since the first Fresnel zone was completely blocked by the occluder irrespective of the receiver's position (see Figures 5.18).

Frequency (Hz)	Wavelength (λ)
63	5.44
125	2.74
250	1.37
500	0.69
1000	0.34
2000	0.17
4000	0.09
8000	0.04

Table 6.1: Frequency and corresponding wavelength. With a velocity of $v_c = 343\text{m}\cdot\text{s}^{-1}$ [48] for sound in air and a frequency of fHz, wavelength $\lambda = v_c/f$.

6.3 Sonel Mapping as a "Whole"

Verifying the correct operation and effectiveness of the individual components comprising the sonel mapping algorithm is a much easier task then verifying the operation of the algorithm as a "whole" with all components working collectively to provide an estimate of the room impulse response. Proving the effectiveness of any acoustical modeling system is a difficult task and currently, no standard evaluation method exists. In fact, although various auralization systems do exist, a detailed evaluation of their accuracy is lacking (see [179]). Various approaches have been considered in the past to address this issue including experiments using human subjects and comparisons between various measures made in an actual room and the corresponding measures estimated by simulating the same room by the system. Evaluation of the sonel mapping algorithm functioning as a complete system was performed in two simulations. In the first simulation (the "simple room" simulation), a single sound source

and receiver were placed within a rectangular room with a single occluder that contained two diffracting edges such that the direct path between them was occluded. The sound reaching the receiver over a short duration was then estimated several times, each time varying the number of sonels emitted from the sound source and for various frequencies.

Generally, a decrease in sound level was observed as frequency was increased (and hence wavelength decreased). This is to be expected given the inverse relationship between wavelength and diffraction (in the sonel mapping method, as wavelength is decreased, the surface diffraction zone is decreased and therefore, the likelihood of diffraction also decreases). An increase in receiver sound level was also observed with increasing sonel count. This too is also expected given that the likelihood of a sonel interacting with a receiver as the number of propagating sonels is increased also increases. However, increasing the sonel count leads to a direct increase in the computation time. This is illustrated in Figure 5.23 where sonel count is plotted against simulation time. Simulation time for each sonel count was averaged across each frequency channel. Although error bars (standard deviation) are included in Figure 5.23, the resulting error (standard deviation) associated with averaging across each frequency band is small and therefore, error bars may not be clearly visible. The results are provided in Table 5.9 and as shown, there is a direct, linear relationship between sonel count and simulation time. With the

computer used for this simulation, the average time to emit and trace one sonel is approximately 13ms.

The second simulation was provided to illustrate the operation of sonel mapping. The propagation of sound energy in a particular environment under several different sound source, receiver and occluder configurations was modeled: i) absence of the occluder, ii) presence of the occluder but diffraction effects ignored and iii) presence of the occluder and diffraction effects accounted for. For each of the three scenarios, the position of the single receiver was varied across a plane along a portion of the "x-z" axis of constant height (e.g., y-coordinate remained constant at y = 5). Both x and z coordinates were varied in increments of 0.5m beginning at 5.5 and ending at 9.5.

Although the purpose of the simulation was to illustrate the operation of various aspects of sonel mapping, several observation confirming the correct operation of sonel mapping can be made. In the first scenario (absence of the occluder), as shown in the contour plot (sound level vs. receiver position) of Figure 5.26(b), sound level was distributed throughout the portion of the room that was sampled (in the contour plot, sound level is encoded by color whereby, brighter colors indicate greater levels). This is expected given that there were no obstructions within the room and there was a clear path between the sound source and receiver for all receiver positions considered (in the example echogram of Figure 5.26(a), the direct sound is clearly evident).

In the second scenario, an occluder was placed between the sound source and receiver thus obstructing the direct path between them. As illustrated in the contour plot of Figure 5.27(b), the decrease of sound level behind the occluder is clearly evident. This is also expected given that the occluder blocked many of the direct paths between the sound source and receiver. However, the height of the room was 8m and the height of the edge was 6m and therefore, there will still be sonels that do reach beyond the occluder despite the fact that there is no direct path between the sound source and receiver. Finally, in the third scenario where the occluder was present but diffraction effects were accounted for, sound level for receiver positions beyond the occluder was greater than the sound level at similar locations when edge effects were ignored (see Figure 5.29(b)). This increase in sound level arriving at the receiver is also expected given that diffracted sound reaches a receiver despite the fact that the direct path between the sound source and receiver is occluded.

6.4 Limitations of the Sonel Mapping Algorithm

Although sonel mapping does overcome many of the problems inherent with many of the currently available acoustical modeling systems, there are several limitations associated with the algorithm as currently implemented. Many of these limitations result from the several assumptions that are presently in place and not necessarily related to limitations with the algorithm itself. For

example, for the purposes of modeling diffraction effects, it is currently assumed that the scene is comprised of planar occluders (edges) only (e.g., no curved surfaces) and therefore, any non-planar objects must be approximated with planar surfaces. This planar assumption is not required when considering specular and diffuse reflections only but is required to model diffraction effects and particular, to dilate a surface by frequency dependent amount and thus define the diffraction and non-diffraction zones. Furthermore, edges in the scene where a sonel can be diffracted must be explicitly specified by the user. Other limitations associated with the current implementation are related to the simplified sound source distribution functions and the fact that refraction is ignored. Ignoring refraction limits sonel mapping to indoor environments whereby refraction can in fact be typically ignored [48].

6.5 Summary

This chapter examined the effectiveness and correct operation of the sonel mapping method by examining the results of the simulations presented in Chapter 5 in greater detail. In many situations, the environment remains static (e.g., no movement of the sound source or objects within the environment) for some time. By exploiting this static property and therefore storing and reusing information regarding the sound energy within the environment as done with the sonel map, efficiencies can be gained. Results of various simu-

lations indicate the correct operations of the components comprising the sonel mapping algorithm as well as the collective operation of the individual components working as a complete system to provide an efficient estimation of the sound propagation through an environment.

Although the focus of this work is acoustical modeling, sonel mapping is itself not specific to acoustical wave energy propagation. Rather, sonel mapping is a framework for energy propagation despite the fact that in this work, it is used to model the acoustics of a particular environment. Given the appropriate model parameters (for example, source emission functions, wave/surface interaction models etc.), it can be used to model the propagation of any type of wave energy, be it sound, light etc. including any type of interaction between the wave and the medium (if a medium is required for the wave to propagate) in addition to any surfaces/objects the wave may encounter. Given a particular model of wave (energy) propagation, sonel mapping can be used to simulate and recreate the particular wavefield. Finally, there are various limitations associated with sonel mapping,many of them stemming from the various assumptions currently in place. As described in Chapter 7, many of these limitations will be addressed in future work.

Chapter 7

Conclusions and Future Work

Given the importance of spatial hearing to humans, incorporating spatialized sound cues in immersive displays seems obvious. In fact, doing so can be beneficial for a variety of reasons. Although the inclusion of spatial sounds can lead to greater realism and quality, in contrast to the sounds present in our natural surroundings sounds present in many acoustical displays typically lack spatial information. Spatial auditory cues are overlooked by the majority of immersive displays where historically, emphasis has been placed on the visual sense instead. Furthermore, when present, the spatial sound cues do not necessarily reflect natural cues. The majority of systems that do convey auditory information do so poorly, typically assuming that all interactions (reflections) between a sound wave and objects/surfaces in the environment are specular,

despite that in our natural settings, acoustical reflections may be diffuse and there may also be diffractive and refracted components as well. Failure to accurately model reflection phenomena leads to a decrease in the spatialization capabilities of the system, ultimately leading to a decrease in performance and a decrease in presence or immersion. Over the last few decades the field of virtual audio has progressed considerably due in part to the realization of the benefit spatial sound can offer. Spatial sound technology is now being exploited by a wide variety of applications. It has become an integral part of many multi-modal virtual environments and it is being incorporated into real world applications where it provides human operators further information without compromising the information presented via other modalities.

Although much progress has been made in the field of spatial audio, the majority of accurate acoustical modeling techniques are far too computationally expensive to meet the dynamic and interactive requirements of virtual environments. This dissertation presented the sonel mapping algorithm. Given a model of wave (energy) propagation, sonel mapping can be used to simulate and recreate the particular wavefield regardless the complexity of the model itself. This chapter provides a summary of the sonel mapping algorithm and highlights its major contributions. The chapter ends with a suggestion of potential areas for future research.

7.1 Summary

Sonel mapping is the application of the photon mapping technique to acoustical modeling. Sonel mapping uses the same basic approach as photon mapping but takes into account the physical attributes of sound propagation while addressing the possible interactions when a propagating sound encounters a surface/object or obstruction in its path (e.g., specular or diffuse reflection, diffraction or absorption).

As with photon mapping, sonel mapping is a two-pass Monte-Carlo particle-based technique. In the first pass (the *sonel tracing* stage), sonels are emitted from each sound source and traced through the scene until they interact with a surface. The distribution of sound frequency in a given source is approximated by considering the center frequency of a fixed number of frequency bands (channels). Each sonel represents the energy contained in one frequency band. When a sonel encounters a diffuse surface at some point p, it is stored in a structure called a *sonel map* while a new sonel is generated and reflected diffusely. Upon encountering a specular surface, the sonel is reflected specularly. If the sonel encounters an occluder (edge) or, falls within the *diffraction zone* of an occluder, the sonel is reflected by choosing a random direction over the hemisphere about the incidence point on the edge. As with specularly reflected sonels, diffracted sonels are not stored.

In the second stage (the *acoustical rendering* stage), the echogram is estimated through the use of the previously constructed sonel map coupled with Monte-Carlo visibility ray tracing. The echogram is estimated by tracing acoustical visibility rays from each receiver into the scene and recording their interaction with any objects/surfaces they may encounter. When a ray intersects a diffuse surface at point p, tracing of the ray terminates and the sonel map is used to provide an estimate of the acoustical energy leaving point p and arriving at the receiver using a density estimation algorithm. The energy is scaled to account for attenuation by the medium and then added to the accumulating echogram. Specular reflections are handled using the same approach as in the sonel tracing stage whereby ideal specular reflections are assumed (e.g., angle of reflection is equal to the angle of incidence with respect to the surface normal). When an acoustical visibility ray encounters a sound source indirectly, the fraction of energy leaving the sound source and arriving at the receiver is determined, scaled to account for attenuation by the medium and the added to the accumulating echogram. Direct sound is accounted for when an acoustical visibility ray encounters a sound source prior to encountering any other object/surface. When the acoustical visibility ray is incident on an edge or falls within the *diffraction zone* of a surface, the ray is diffracted using a modified version of the Huygens-Fresnel principle.

Sonel mapping overcomes many of the fundamental problems associated

with deterministic approaches (e.g., exponential running times) by employing Monte-Carlo methods. Sonel mapping is based on Monte-Carlo ray tracing whereby *point sampling* is used to provide an estimate of the sound energy in a model. In addition, instead of relying on a deterministic approach, sonel mapping employs a Russian roulette approach to determine which type of interaction does occur at each sonel/surface interaction point. Using a Russian roulette approach, a single interaction occurs at each sonel/surface interaction point as opposed to multiple interactions inherent with many deterministic approaches. This leads to a tremendous savings in computational cost in addition to allowing for the possibility of exploring arbitrarily long reflection paths, paths that may not necessarily be explored using deterministic techniques. The sonel mapping approach enjoys various advantages over deterministic based techniques, including i) the ability to handle arbitrary geometry, ii) low memory consumption, iii) can handle procedural geometry, iv) can handle any type of reflection model v) does not require a pre-computation of the representation for the solution and vi) the result is correct except for variance (noise). Furthermore, the use of a Russian roulette approach allows for the possibility of exploring arbitrarily long paths that may not necessarily be explored using other, deterministic approaches while eliminating the potential exponential running times inherent in many deterministic approaches. Moreover, with Russian roulette, the accuracy of the simulation can be improved by increasing

the number of samples initially emitted from the sound source. Although this leads to an increase in computation time, an efficiency vs. accuracy trade-off can nevertheless be made.

In addition to modeling of specular and diffuse reflections, sonel mapping addresses the modeling of diffraction effects. Acoustical diffraction is accomplished using a modified version of the Huygens-Fresnel principle [79]. The Huygens-Fresnel principle assumes a propagating wavefront is composed of a number of secondary sources. This fits nicely into the sonel mapping probabilistic framework whereby acoustical wave propagation is approximated by propagating sonels from a sound source and tracing them through the environment. Diffraction effects are approximated in a very simple and efficient manner allowing computation at interactive rates. Although the Huygens-Fresnel principle is a rather simple approach, it can satisfactorily describe a large number of diffraction configurations in an efficient manner.

Sonel mapping exploits the fact that in many situations, the environment remains static for some time by storing and reusing information regarding the sound energy in an environment. Sonels are stored in a data structure called the *sonel map* and can be re-used later as needed. This avoids potentially complex and computationally expensive operations required to model the propagation of sound energy from continuously being performed leading to further gains in efficiency.

7.2 Future Work

Although the sonel mapping approach introduced in this dissertation overcomes many of the limitations associated with existing acoustical modeling approaches, it has also identified issues that warrant further investigation. This section describes potential additions, modifications and improvements that could be pursued in the future.

7.2.1 Short Term Extensions

More realistic distribution functions for the emission of sonels throughout various phases of the sonel mapping algorithm could be explored. Directional distribution functions could be employed at the sound source to model sound sources with other than omni-directional directivity. This would allow for more realistic modeling of sound sources such as loudspeakers, musical instruments and human speech. Distribution functions could also be employed in the acoustical rendering stage where the sonel map is used to estimate the diffusely reflected energy reaching a receiver. Currently it is assumed that sonels used to provide this estimate (e.g., all sonels within a pre-defined distance of the incidence point) will actually reach the receiver when in fact they may not necessarily do so. An energy distribution function would more realistically describe the energy reaching the receiver from such a diffuse reflection point.

Another approach worth further investigation is the combination of Russian roulette and a deterministic termination criterion such as the energy discontinuity percentage (EDP) [52] whereby a sonel is terminated after its energy level has decreased beyond some pre-defined amount. The EDP could be used to accurately model the early and most important portion of the impulse response while Russian roulette could be used to model the latter portion. The latter portion of the impulse response is typically approximated as exponentially decaying noise in order to limit the computational cost associated with modeling it. However, using Russian roulette, it may be possible to obtain an efficient approximation to the latter portion of the impulse response more accurate than existing approaches at a feasible computational expense. By combining both approaches, the benefits of each one can be exploited and used to provide a more accurate solution without incurring the potentially exponential increase in computational cost if a deterministic approach was used alone.

Finally, only diffusely reflected sonels are currently stored in the sonel map. Specularly reflected sonels are not stored since such sonels would not provide useful information (given that the probability of having a matching incoming sonel from the specular direction is small). Future work will investigate the storing of diffracted sonels in a manner similar to the diffusely reflected sonels possibly in a separate map. This may further lead to greater efficiencies as it avoids the re-computation of diffractive components.

7.2.2 Longer Term Extensions

The current model of acoustical energy propagation contains various assumptions and approximations. One such approximation is with respect to refraction: it is ignored altogether. Although refraction can be ignored in the majority of indoor acoustical modeling applications where the medium is homogeneous as assumed for the environments considered in this work, modeling of acoustical refraction will be addressed in future versions. Accounting for refraction effects will increase the robustness of the sonel mapping algorithm. In particular, it can be used to model outdoor environments where refraction is more likely to occur and to model environments where the medium of propagation is not necessarily solely air but may include water (e.g., under-water acoustics) or solids (e.g., steel). A technique that may prove to be useful in modeling refraction phenomena is the *participating media* [84] extension to photon mapping that allows photon mapping to model environments where the assumption that light travels through a vacuum (as assumed in the photon mapping technique) is not valid (e.g., light passing through fog and smoke).

Sonel mapping is a physical-based approach, inspired by various physical phenomena that describe the propagation of wave energy. The simulations presented in this dissertation have shown that sonel mapping does conform to the expected analytical and theoretical results. Further verification could involve comparisons between measurements made in an actual (controlled room)

and with the results of simulating the acoustics of the same room. Although measurements can be made in a "regular" (con-controlled) room, using a room whereby the room parameters can be controlled (e.g., exactly specifying the absorption and diffuse reflection coefficients) can allow for meaningful comparisons to be made. Such an approach is taken by Tsingos et al. [179] where they constructed a simple enclosure (the "Bell Labs Box") in order to validate their acoustical diffraction method by comparing actual measured impulse responses in the Bell Labs Box and the impulse response obtained by simulating the enclosure.

Despite the many simulations described in this dissertation, none of these simulations included human subjects despite the fact that the ultimate user of any acoustical modeling application and virtual environment in general including sonel mapping, is a human. Although sonel mapping may be correct with respect to physical laws, human auditory perception must also be accounted for as various physical attributes of sound may lead to differing perceptual responses across human observers. In addition, perceptual factors may also dictate that in certain situations, complete accuracy is not necessarily required and a coarse approximation may be sufficient, leading to potential increases in efficiencies. There is a vast amount of information with respect to acoustical cues used by human listeners given the long history of psycho-acoustical research that have investigated controlled studies examining the relationship

between observed sound stimuli features and the corresponding perceptual response [108]. Human perceptual factors and insensitivities have been exploited in various computer graphics applications. By employing computational resources to those portions of a scene that are observed by a human observers in contrast to features that provide little or no meaning to the overall image appearance, can result in tremendous performance gains can be achieved [114]. Despite this vast knowledge, many acoustical modeling approaches have focused on the physical attributes of sound propagation and at times completely ignore any perceptual effects. Fortunately, this is changing. For example, Martens has recognized the importance of perceptual effects with respect to auditory displays and examined the deployment of auditory display technology whereby responses are calibrated to actual responses of a the human listener [108, 110, 112]. Martens [109] also describes some of the "uses and misuses" of psychophysical methods that are used to subjectively evaluate spatial sound reproduction.

Given the importance of perceptual factors, future work will include experimental verification of sonel mapping with human subjects to determine the effectiveness of the approach to human listeners, the intended target audience. Sonel mapping is based on theoretical (physical) properties and although it does conform to theoretical results, there is not necessarily a perfect mapping between physical properties and perception. For example, the perceptual

equivalent to intensity is loudness; loudness is a subjective measure and although intensity can be measured, loudness cannot. Furthermore, loudness may not always be an accurate representation of intensity [132] as the loudness of pure tone sounds is frequency and bandwidth dependent [133]. In addition to verification of the approach, human tests may also lead to refinements to the algorithm in order to account for perceptual effects. Human tests would also allow for meaningful conclusions to be drawn with respect to the efficiency vs. accuracy trade-off in order to determine just how many sonels are required to simulate the acoustics of a particular environment. This may involve user tests whereby users are to localize a sound source while varying the number of sonels emitted from a sound source (e.g., the number of sonels could be the control variable). Since there is a direct relationship between the number of sonels emitted and the required computation time, limiting the number of emitted sonels leads to increased efficiency. Furthermore, as described in the simple room simulation of Section 5.1.4.5, as the number sonels emitted from a sound source is increased, the sound level at a receiver approaches the actual sound level. Rather than increasing the number of sonels, it may be worth examining whether there is a "correction" that can be used to bring the sonel count to an asymptote. User tests can be used to confirm the validity of such a correction. Other potential avenues worth investigating further involve examining how much accuracy with respect to physical properties, is actually required when

modeling various aspects of the algorithm (e.g., diffraction, diffusion and specular reflection) and whether perceptual effects can be incorporated in to modeling such features thereby reducing the computational requirements. Martens at al. have conducted similar investigations with respect to diffraction modeling and as described in Section 3.2.3, describe a perceptually-based solution to the diffraction of sound by an occluder of low computational cost.

Finally, the sonel mapping algorithm itself is parallel in nature when considering that each of the discrete set of frequencies in the simulation is considered separately. This parallel nature can be exploited by performing the simulation of each frequency separately in parallel and then combining the results. Although introducing such parallelism does require the use of additional hardware (e.g., processors), the cost of such hardware has dropped significantly over the last several years making such an approach definitely feasible.

Appendix A

Background: Acoustical Rendering - Putting it all Together

This appendix examines the delivery of spatial sound to a listener through headphones or loudspeakers, outlining the advantages and limitations of each approach. In the process of doing so, several loudspeaker-based three-dimensional sound systems are described. Although the delivery of spatial sound to a listener is not considered in this dissertation, the information presented in this Appendix is given for completeness.

A.1 Headphone Based Systems

Headphone based systems offer a number of advantages over loudspeaker based systems. Headphones provide a high level of channel separation, thereby minimizing any *crosstalk* that arises when the signal intended for the left (or right) ear is also heard by the right (or left) ear. Headphones isolate the listener from external sounds and reverberation which may be present in the environ-

ment [67], ensuring the acoustics of the listening environment or the listener's position in the room, do not affect the listener's perception. Despite the potential benefits headphone based systems offer, they do have their shortcomings as well. Headphones may be uncomfortable to wear and can be cumbersome. Furthermore, unless the relevant spatial information is accounted for (e.g., reverberation and HRTF information), sounds conveyed through headphones will not be "externalized" but rather will appear to be originating from within the head. This is referred to as *inside-the-head localization* (IHL) and is described further in the following section. Finally, as described in Section A.1.2, the response of the headphone introduced by the headphone transducer itself may have to be accounted for as well.

A.1.1 Inside-the-Head Localization

Inside-the-head localization (IHL) refers to the lack of externalization of a sound source resulting in the false impression that the sound is originating from inside the listener's head and can only move left and right inside the head along the interaural axis, being biased towards the rear of the head [94]. IHL results from various factors including the lack of correct environmental context (e.g., lack of reverberation and HRTF information). IHL can be greatly reduced by ensuring the sounds delivered to the listener's ears reproduce the sound as it would occur naturally, or in other words providing the listener with a "realis-

tic spectral profile of the sound at each ear" [162]. Although the externalization of a sound source is difficult to predict precisely, it does increase as the sound becomes more "natural" and contains spatial information [15] that is updated appropriately with any head movements, as in "normal" listening situations [43]. Although rare, IHL can also occur when listening to "external" sound sources in the real world, especially when the sounds are unfamiliar to the listener or when the sounds are obtained (recorded) in an anechoic environment [43].

A.1.2 Headphone Equalization

The headphone transfer function represents the characteristics of the headphone transducer itself as well as the transfer function between the headphone transducer and the ear drum (or at the point in the ear canal or outer ear where it is measured) [98]. It can be measured in a manner similar to the approach used to measure HRTFs except that headphones are placed over the ears of the target head. Unlike HRTFs, the headphone transfer function does not vary as a function of sound source location hence only one measurement is required. Once the transfer function has been obtained, equalization filters can be used to remove it from headphone conveyed sound. A detailed description of headphone equalization is provided by Møller [129].

The spectral features of the headphone transfer function can be significant

and can contain peaks and notches with magnitude and bandwidth similar to the magnitude and bandwidth of the peaks and notches of HRTFs [98]. The transfer function may influence the resulting ITD due to group delays which vary between the ears and the location of headphone placement. However, the differences resulting to the transfer function from different headphone placements will unlikely lead to any degradation of localization abilities [116, 121]. Several studies have also shown that the headphone transfer function varies across individuals with individual differences that can be substantially larger than the differences between individual HRTFs [71, 130]. Although headphone equalization is a straightforward task, its effects on sound source localization are not completely known and currently, it is unclear whether headphone equalization is actually required [202].

A.2 Loudspeaker Based Systems

In this section, several loudspeaker-based three-dimensional sound techniques are described. There are also various loudspeaker-based systems that do not necessarily incorporate "true" three-dimensional sound technologies, but are widely used especially for entertainment purposes (e.g., Quadraphonics™ [153, 154], Ambisonics [60, 68], Dolby Stereo™, Dolby Digital™ and Dolby 5.1™ [102]). Such systems do not necessarily incorporate "true" 3D sound technologies but they have nevertheless inspired development of more modern ap-

proaches including *wave field synthesis* whereby a large number of loudspeakers is used to completely and accurately recreate the sound field of a particular environment leading to high quality auralization [22]. Emphasis here is placed on systems that aim to recreate a particular sound field as it would be in the natural setting and therefore, such systems are not described further here (see [158] for further details regarding such system and recording/playback techniques in general). Irrespective of the loudspeaker techniques employed, the intended effect of these techniques is restricted to a small region of space. This small region is known as the *listener sweet spot* and deviation from this region will lead to serious degradations in system performance.

A.2.1 Transaural Audio

Rather than presenting BRIR-processed sound to a listener via headphones, loudspeakers can be used instead. However, in contrast to using headphones, there is no isolation between the signals intended for the left and right ears when loudspeakers are used. In other words, in a typical two loudspeaker (stereo) scenario, the signal received at the left and right ears is a linear combination of the signal output by the left and right loudspeakers, including any filtering effects introduced by the loudspeakers and the environment (e.g., the speaker frequency response, absorption of sound by the medium and head response) [67]. In addition to the desired signal coming from the left and right

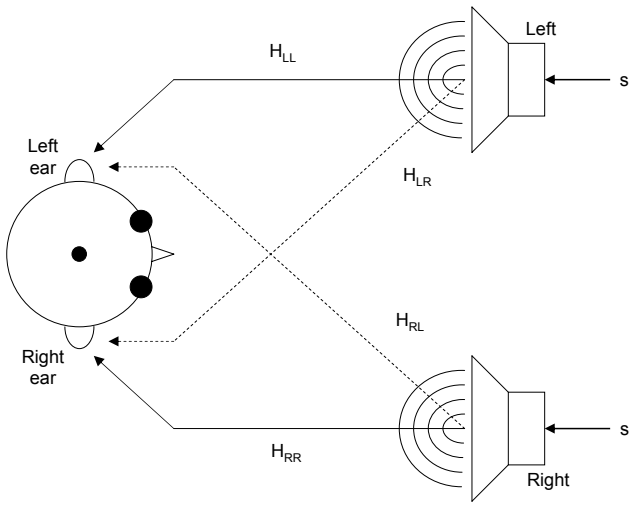


Figure A.1: Crosstalk. When using loudspeakers as opposed to headphones to convey sound to a user of a three-dimensional sound system, in addition to the desired left loudspeaker signal H_{LL} reaching the left ear e_L, a delayed and attenuated portion of the right loudspeaker signal H_{RL} will also reach the left ear. A similar situation occurs with the signal reaching the right ear e_R, where in addition to the desired signal H_{RR} from the right loudspeaker, a delayed and attenuated portion of the left loudspeaker H_{LR} will also reach the right ear.

loudspeakers H_{LL} and H_{RR} respectively, a delayed and attenuated portion of the left loudspeaker signal will reach the right ear H_{LR}, while a delayed and attenuated portion of the right loudspeaker signal will reach the left ear H_{RL} (see Figure A.1). This is known as *crosstalk* and must be removed. *Transaural audio* [38] is a loudspeaker technique that employs crosstalk cancellation to remove the unwanted cross talk signals and can overcome some of the limitations inherent with HRTF headphone based systems.

A.2.1.1 Crosstalk Cancellation

Crosstalk cancellation was first proposed by Bauer in 1961 [12] in order to allow for the delivery of HRTF based audio to a listener using a pair of loudspeak-

ers as opposed to headphones. Two years later, the first crosstalk canceller was actually implemented by Atal and Schroeder [8]. Essentially, the Atal and Schroeder crosstalk canceller involves adding a delayed and inverted version of the crosstalk signal to the opposite loudspeaker output. A delayed and inverted version of the crosstalk signal going from the right loudspeaker to the left ear H_{RL} would be added to the left loudspeaker output, while a delayed and inverted version of the crosstalk signal going from the left loudspeaker to the right ear H_{LR} would be added to the right loudspeaker output. Given that the inverted signals are 180° out of phase and delayed, if the delay is chosen such that it equals exactly the amount of time it takes for the crosstalk signal to reach the opposite ear, the crosstalk will be (ideally) completely cancelled.

In theory, crosstalk cancellation completely removes the unwanted signals thereby allowing the desired binaural signals to be delivered to the corresponding ears. In practice however, this is not the case. Given the use of HRTFs in the crosstalk canceller, its effectiveness is limited by the variability in size and shape of the human head and pinna [67]. It has a small listener sweet spot and in order for it to function properly, the listener must remain stationary in the sweet spot since movements as small as 74 - 100mm completely destroy the desired effect [135]. When the listener moves more than this amount, the HRTFs used by the crosstalk canceller may be incorrect and the time required for the crosstalk signals to reach the contralateral ears and the attenuation

factor may also change. As with headphone based systems, this problem can be greatly reduced by tracking the listener's head. Gardner [67] developed a system utilizing a magnetic head tracker in order to produce a much more realistic and greater range three-dimensional auditory display using loudspeakers. Given the dynamic updates of head movements, this system offers improved localization over existing, non-tracked loudspeaker displays as it allows for dynamic localization cues. Mouchtaris et al. [135] describe a loudspeaker based three-dimensional audio display which allows for dynamic crosstalk cancellation that uses a camera-based head tracking system and thereby eliminates the tether associated with magnetic trackers.

A.2.2 Amplitude Panning

In *amplitude panning*, the amplitude (intensity or output level) of the signal being delivered to each loudspeaker[1] is adjusted in some manner as to simulate the directional properties of the ILD. In other words, by adjusting the amplitude of the signal applied to each loudspeaker through the use of a gain factor, the listener perceives a virtual sound source emanating from some direction dependent on the gain factors [144]. Mathematically, amplitude panning is described by

$$b_i(t) = g_i(t)s_m(t), \qquad i = 1,\ldots,N \tag{A.1}$$

[1]Headphones can also be used when the number of loudspeakers is two.

where, $b_i(t)$ is the signal output by loudspeaker i at time t, $s_m(t)$ is the "unprocessed" sound applied to each of the loudspeakers at time t, g_i is the gain factor applied to the signal delivered to loudspeaker i, and finally, N is the total number of loudspeakers being used.

Various amplitude panning techniques exist which allow for a wide variety of loudspeaker set-ups including both two and three-dimensional configurations. The general idea is to compute the appropriate gain factors to create the impression of a virtual sound source at a specific position relative to the listener. The *stereophonic law of sines* [11] and the *tangent law* [19] can be used to compute the gain for each channel in the typical two-channel stereo configuration. Using *pair-wise amplitude panning* techniques [39], two-channel methods can be extended to account for N loudspeakers by choosing and outputting the sound to two loudspeakers only, in a manner similar to the conventional two-channel stereo panning technique.

Three-dimensional panning is an extension of the two-channel, two dimensional technique. In a manner similar to pair-wise amplitude panning, sound is applied to a subset of three loudspeakers only. A virtual sound source can be positioned anywhere on the triangle formed by the three loudspeakers. However, currently no general trigonometric method of three-dimensional amplitude panning for an arbitrary three-dimensional loudspeaker setup exists and the calculation of the gains applied to the loudspeakers is configuration depen-

dent [144].

A.2.2.1 Vector Base Amplitude Panning

A more recent method of calculating the gain factors is the *vector base amplitude panning* (VBAP) technique. This technique can be used with an arbitrary number of loudspeakers and supports two and three-dimensional loudspeaker configurations. It allows the loudspeakers to be placed in any position provided they are nearly equidistant around the listener and that the listening room is not very reverberant [144].

In the stereo VBAP configuration, the two channel stereo setup is treated as a two-dimensional vector base defined by two unit length vectors, each vector pointing to one of the two loudspeakers. A third unit vector points to the direction of the virtual sound source and is formulated as a linear combination of the two (appropriately scaled) loudspeaker vectors. Using simple linear algebra techniques, the two loudspeaker scaling factors (gains) can be calculated. The formulation of two-dimensional VBAP can be generalized to handle a three-dimensional loudspeaker configuration, where three equidistant loudspeakers can be thought of as being positioned on an imaginary unit radius sphere. Three loudspeaker unit vectors point from the listener position to one of the three loudspeakers and a fourth unit vector points to the position of the virtual sound source. The virtual sound source can then be mapped into a loca-

tion within the “active triangle” formed by the three loudspeakers. As with the two-dimensional stereo configuration, the vector pointing to the virtual source can be given as a linear combination of the three loudspeaker vectors and linear algebra can be used to calculate the appropriate gain applied to each loudspeaker. Two and three-dimensional VBAP can be extended to allow an arbitrary number of loudspeakers. In a two-dimensional configuration, two of the N loudspeakers that lie on the same plane as the listener (typically the horizontal plane [144]) are chosen and sound is applied to these two loudspeakers only. In the three-dimensional configuration, three of the N loudspeakers are chosen and the sound will be mapped to a location within the “active triangle” formed by the three loudspeakers.

The VBAP technique is a relatively simple and computationally efficient method which allows for the maximum virtual sound source localization accuracy possible with amplitude panning [144]. However, in the three-dimensional configuration, maximum localization accuracy (minimum virtual sound source directional error) is proportional to the dimension of the active triangle [144] and although the dimension of the active triangle can be decreased by increasing the number of loudspeakers, increasing the number of loudspeakers may not always be practical. In addition, as with all pair-wise and triplet-wise amplitude panning techniques, the virtual sound source will spread when it is panned between loudspeakers. Finally, although VBAP allows for accurate

virtual sound source localization on the azimuthal plane especially near the median plane, the localization of virtual sound sources that do not lie on the azimuthal plane (e.g., non-zero elevation) cannot be predicted as it is dependent on the individual listener although with a large number of loudspeakers elevation localization becomes acceptable [145].

A.2.3 Wave Field Synthesis

Wave field synthesis (WFS), developed at the Technical University Delft by Berkhout et al. [22, 29], involves a large number of closely-spaced loudspeakers that are fed audio signals in a manner such that a highly natural sound field is produced including the reproduction of wave front curvature that would result from real sound sources. It thus allows the simultaneous reproduction of an arbitrary number of virtual sound sources [22]. The theory behind the WFS technique is Huygens' principle which states that at every time instant, every point on the primary wavefront can be thought of as a continuous emitter of secondary wavelets and these secondary wavelets combine to produce a new wavefront in the direction of propagation. Given a wave (sound) field (that is known with respect to pressure and normal particle velocity) on a boundary surface S of a closed volume V free of any sources, the sound pressure at any point within V can be determined. In the ideal scenario, planes of loudspeakers surround the listening area and are driven with signals such that they

produce a volume flux proportional to the normal component of the particle velocity of the original wave field at each corresponding position [28]. However, for practical purposes (e.g., hardware and computational power requirements), rather than using multiple planes of loudspeakers, linear loudspeaker arrays surround the listener area instead. This does of course lead to several problems. Most notably, sound reproduction is correct only for wave field components in the horizontal plane and therefore, sound reproduction is not correct over the entire listening area [27]. Given a linear array, the input signal $E_i(\omega)$ of each loudspeaker i can be described as [28]

$$E_i(\omega) = K\sqrt{jk}V_n(r_i,\omega) \tag{A.2}$$

where, $V_n(r_i,\omega)$ is the normal component of the particle velocity at the loudspeaker position r_i, k is the wave-number and K is a constant depending on the loudspeaker sensitivity (given a loudspeaker with a flat frequency response, K is frequency independent).

In contrast to other loudspeaker-based systems whose intended effect is restricted to the listener sweet spot, WFS systems generate a wave field with natural time and space properties enveloping an extensive listening area [53] where multiple listeners are free to move about without fear of losing the correct acoustical impression. This has made WFS an attractive technique for auralization in addition to other applications including sound enhancement in

theaters, multi-purpose auditoria and reproduction of multi-channel recordings [53]. In fact, in theory, it is a superior technique [144], however, it is usually impractical due to several inherent limitations. Most importantly, it requires the distance between loudspeakers to be as small as possible in order to avoid spatial aliasing (the highest frequency that can be represented is inversely proportional to the spacing between loudspeakers) [144, 185]. This results in the requirement of a large number of loudspeakers and extensive computational requirements.

A.2.4 Spherical Microphone Arrays

Rather than generating three-dimensional audio using the methods discussed in the previous sections (HRTF-based techniques, amplitude panning, WFS etc.), another technique involves recording the sound field of a particular environment using an array of microphones and then reproducing it at a later time over a region of space and over the entire frequency band [1] with the ultimate goal of reconstructing the original sound field [126]. Various microphone array configurations are available including linear, circular and planar, whose theoretical analysis is well developed and have been applied to a variety of applications such as speech enhancement in conference rooms and auralization of sound fields measured in concert halls [147]. Spherical microphone arrays, whereby a number of microphones are mounted on the surface of a sphere, have

also been introduced and used for a variety of applications including, spatial beamforming, sound recordings with high spatial detail and sound field measurements and analysis. Unlike other array configurations, the spherical symmetry inherent in a spherical microphone array allows for three-dimensional analysis [148]. Various sampling approaches are available including *equiangle sampling* [56], *Gaussian sampling* and *nearly uniform sampling* [148]. However, regardless the sampling technique used, in order to avoid aliasing, sampling must be band-limited (e.g., limited harmonic order) [148] and the number of microphones required to sample up to the Nth order harmonic of a signal is at least $(N+1)^2$ [148]. Although in theory one can sample up to any order harmonic, due to the limited technology currently available, sampling has typically been restricted to measuring the zeroth and first order of a sound field.

The idea of recording a sound field and reproducing it a later time is certainly not new. In fact, Ambisonics, introduced in the early 1970s is such a microphone technique that can be simulated to perform a synthesis of spatial audio [144]. However, until recently, such available systems were restricted to recording the zeroth and first order harmonics of a sound fields. Only recently has a system capable of recording second order sound fields been introduced [143]. Abhayapala and Ward presented the theory (using spherical harmonics analysis) and guidelines for a higher order system and provided an example of a third order system for operation in the frequency range of 340Hz to

3.4kHz [1]. More recently, Rafaely presented a spherical-harmonics-based design and analysis for a spherical microphone array framework, covering the effects factors including array order, input noise, microphone positioning and spatial aliasing [148].

Appendix B

Monte Carlo Methods

This appendix provides a brief introduction to stochastic methods relevant to this work. In particular, Monte-Carlo integration and Russian roulette are described.

B.1 Monte Carlo Integration

Monte-Carlo integration methods originated at the Los Alamos National Laboratory during the development of the atomic bomb just after World War II [97]. Using stochastic sampling, Monte-Carlo methods are used to estimate integrals which may be difficult to solve analytically. They have been applied in a variety of applications, ranging from neutron transport problems, radiative heat transfer, queuing theory and computer graphics, where they can be

used to accurately compute illumination throughout a scene [173]. Given an integral of the form

$$I = \int_{\Omega} f(x)dx \tag{B.1}$$

where, $f(x)$ is a function evaluated over the domain Ω, Monte-Carlo integration provides an approximation to the integral I by randomly sampling the integral and averaging the results. Mathematically, the Monte-Carlo approximation I_m to the original integral I is given by [173]

$$I_m = \frac{1}{N}\sum_{i=1}^{N}\frac{f(\xi_i)}{p(\xi_i)} \tag{B.2}$$

where, $\xi\ \varepsilon\ \Omega$ is a uniformly distributed random number, $p(\xi_i)$ is the probability density function (pdf) describing the probability of the occurrence of ξ_i and N is the total number of samples taken. As the number of samples is increased, the approximation becomes more accurate and when considering the limit, the approximation will approach the actual value (e.g. converge)

$$\lim_{N\to\infty} I_m = I. \tag{B.3}$$

Although the Monte-Carlo estimate does converge, it does so slowly. In particular, the convergence rate is proportional to $\sqrt{1/N}$ [87] or in other words, in order to reduce the error by one half, four times as many samples are required.

Various techniques are available to improve the convergence rate and hence reduce the variance between the actual and approximated integral values. A common technique is *importance sampling* where essentially, the samples are concentrated on important portions of the function [87]. Another variance reduction technique is *stratified sampling* where in its simplest form the domain of the original function is divided into N non-overlapping sub-domains. At least one of the samples is chosen from each of the sub-domains, thus ensuring all samples are taken from across the domain rather than potentially, from a small region of the domain. Although the convergence rate of Monte Carlo methods is slow, their convergence is independent of dimension and therefore Monte Carlo methods are preferable for approximating multi-dimensional integrals [173].

The field of stochastic sampling in general and Monte-Carlo methods in particular, is large and far more complex than the brief overview presented here. The interested reader is referred to the many general references regarding Monte Carlo methods, including [87, 157] and references describing Monte-Carlo methods for the use of light transport can be found in [173, 184].

B.1.1 The Russian Roulette Approach

Russian roulette is an importance sampling technique used to increase the efficiency of an estimator [173]. It was initially introduced to the field of particle physics simulation [72] to terminate random paths whose contributions

were estimated to be small or in other words, to terminate paths that have entered "non-important" regions of the domain [184]. Arvo and Kirk [7] introduced Russian roulette to the field of computer graphics by incorporating it into their stochastic ray tracing method as a means of terminating recursive rays. Russian roulette ensures that the path length (reflection count) is kept at a manageable size yet due to its probabilistic nature, allows for paths of an arbitrary size to be explored. With respect to image synthesis, this allows for the generation of an unbiased image.

Mathematically, given an integral $I = \int f(x)dx$, an estimator I_m for the integral I and an *acceptance* probability P, with Russian roulette, a uniform random number $\xi \varepsilon [0 \ldots 1]$ is introduced to determine whether the estimator I_m is to be evaluated or not. The Russian roulette estimate I_r can then be made as follows [84]

$$I_r = \begin{cases} I_m/P & \text{if } \xi < P \\ 0 & \text{otherwise.} \end{cases} \tag{B.4}$$

The resulting Russian roulette estimate I_r is unbiased [84]. In other words, its expected value equals the expected value of the integral I, hence, the accuracy of the Russian roulette estimate can be increased (and thus the variance reduced) by increasing the number of samples. Despite the unavoidable increase in variance in comparison to the original estimator I_m, under the as-

sumption that the probability P can be computed quicker than the estimate provided by I_m, the Russian roulette estimate I_r can be computed faster than I_m [173]. In other words, at the cost of higher variance, the efficiency of the estimator is increased.

Appendix C

The Huygens-Fresnel Principle

In this appendix, a detailed mathematical derivation of the Huygens-Fresnel principle following the derivation given by Hecht [79] is presented. Two additional derivations ("finding the position of a secondary source in a Fresnel zone" in Section C.2 and "sampling positions within a Fresnel zone" in Section C.3) relevant to the Huygens-Fresnel principal and developed by the author specifically for this dissertation are also included in this appendix.

C.1 Mathematical Details

Huygens' principle, developed by Christian Huygens in 1678, is based on the wave theory of light. Referring to Figure C.1, Huygens' principle states that every point on the primary wavefront can be thought of as a continuous emitter of secondary wavelets (sources) and these secondary wavelets combine to produce a new wavefront in the direction of propagation [198].

Huygens' principle is itself not completely correct since, if each of the secondary

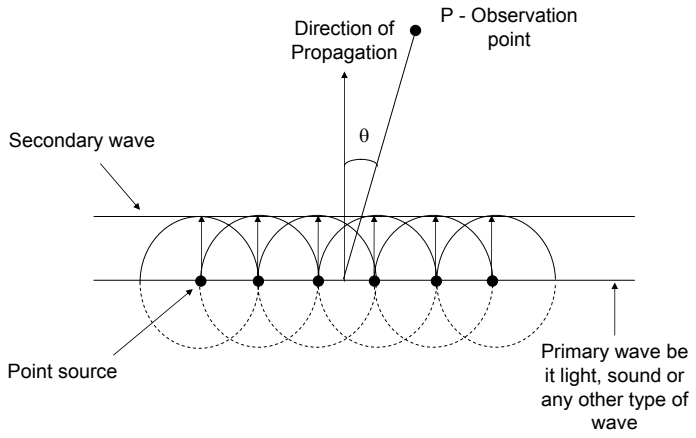


Figure C.1: Huygens' principle. Every point on the primary wavefront can be thought of as a continuous emitter of secondary wavelets (sources) and these secondary wavelets combine to produce a new wavefront in the direction of propagation.

wavelets were emitted uniformly in all directions then, in addition to the forward propagating wavefront, a wavefront propagating in the reverse direction would also be observed when in fact it is not. As inferred by Fresnel and later formulated by Kirchoff, these secondary wavelets are emitted in a direction dependent manner based on an *obliquity* or *inclination* factor $K(\theta)$ [79]

$$K(\theta) = \frac{1}{2}(1+\cos(\theta)) \tag{C.1}$$

where, as illustrated in Figure C.1, θ is the angle made with the normal to the direction of propagation of the primary wavefront. Huygens' principle and Fresnel's modification are collectively known as the Huygens-Fresnel principle and can describe various diffraction configurations in a simple manner [79].

Referring to Figure C.2, consider a sound source (S) and receiver (R) in free space (e.g., no obstacles between them). Having originated at S at time

$t = 0$ with an amplitude $E_\circ$, at a time t' later the wave will have propagated a distance ρ and its amplitude will be $E_\circ/\rho$. The wavefront at time t' can be described as

$$E = \frac{E_\circ}{\rho} cos(\omega t' - k\rho) \tag{C.2}$$

where, $\omega = 2\pi f$ is the angular frequency and $k = 2\pi\lambda$ is the wave-number (λ represents wavelength). This expanding wavefront can be divided into a number of ring-like regions, collectively known as *Fresnel zones* [79]. The boundary of the i^{th} Fresnel zone corresponds to the intersection of the wavefront with a sphere of radius $r_\circ + i\lambda/2$ centered at the receiver where, $r_\circ$ is equal to the distance between the receiver and the expanding wavefront after the expanding wavefront has traversed a distance of ρ from the sound source. In other words, the distance from the receiver to each adjacent zone differs by half a wavelength ($\lambda/2$).

Each Fresnel zone is finite in extent and as illustrated in Figure C.3, a differential ring-shaped area dS can be defined within a zone. The secondary sources (wavelets) within dS are coherent and are assumed to emit in phase with the primary wave. The secondary sources travel a distance r to reach the receiver at a time t, all of them arriving there with the same phase $\omega t - k(\rho + r)$. The strength of the secondary sources per unit area on dS, denoted by E_A, is proportional to $E_\circ/\rho$ within a constant factor Q (e.g., $E_A = QE_\circ/\rho$, where $Q =$

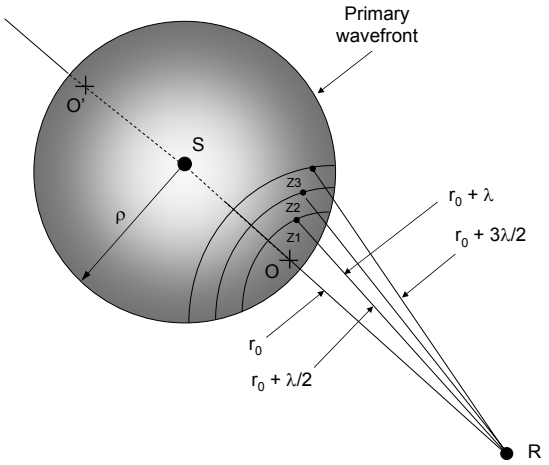


Figure C.2: Fresnel zones.

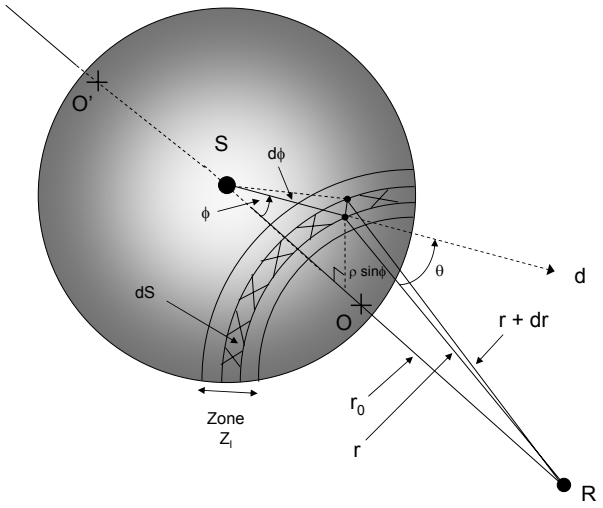


Figure C.3: Fresnel zone geometry.

$1/\lambda$) [79].

The energy dE reaching the receiver from all the secondary sources on dS is given as

$$dE = K(\theta)\frac{E_A}{r}\cos[\omega t - k(\rho + r)]dS \tag{C.3}$$

where, the obliquity factor $K(\theta)$ is assumed to be constant throughout dS and

throughout the entire Fresnel zone. Referring to Figure C.3, dS itself can be given as a function of r

$$dS = \rho d\varphi 2\pi(\rho \sin\varphi) \tag{C.4}$$

and after applying the law of cosines yields

$$r^2 = \rho^2 + (\rho + r_\circ)^2 - 2\rho(\rho + r_\circ)\cos\varphi. \tag{C.5}$$

Keeping ρ and $r_\circ$ constant and differentiating Equation C.5 above gives

$$2rdr = 2\rho(\rho + r_\circ)\sin\varphi d\varphi. \tag{C.6}$$

By rearranging Equation C.6 above, $d\varphi$ can be expressed as

$$d\varphi = \frac{2rdr}{2\rho(\rho + r_\circ)\sin\phi} \tag{C.7}$$

and by using the value of $d\varphi$, dS is given as

$$dS = 2\pi\frac{\rho}{(\rho + r_\circ)}rdr. \tag{C.8}$$

Finally, the energy E_i arriving at the receiver from the i^{th} Fresnel zone (Z_i) can be determined by integrating over all differential areas across Z_i

$$E_i = K_i(\theta)2\pi\frac{E_A\rho}{(\rho+r_\circ)}\int_{r_i-1}^{r_i}\cos[\omega t - k(\rho+r)]dr. \tag{C.9}$$

After performing the integration

$$E_i = \frac{-K_i(\theta)E_A\rho\lambda}{(\rho+r_\circ)}\sin[\omega t - k\rho - kr]_{r=r_{i-1}}^{r=r_i} \tag{C.10}$$

and since $r_{i-1} = r_\circ + (i-1)\lambda/2$ and $r_i = i\lambda/2$, Equation C.10 can be evaluated, leading to

$$E_i = (-1)^{i+1}\frac{2K_i(\theta)E_A\rho\lambda}{(\rho+r_\circ)}\sin[\omega t - k(\rho+r_\circ)]. \tag{C.11}$$

The distance between adjacent zones differs by $\lambda/2$ and therefore, according to Equation C.11, depending on whether i is even or odd, the energy term will be positive or negative respectively. As a result, the energy reaching the receiver from adjacent zones will be out of phase by one half of a wavelength and thus cancel each other. The total energy E_{total} reaching the receiver from the sound source is determined by accumulating the energy from each of the m zones

$$E_{total} = E_1 + E_2 + E_3 + \cdots + E_m. \tag{C.12}$$

Since the sign of the energy of each zone alternates, Equation C.12 can be reformulated as

$$E_{total} = |E_1| - |E_2| + |E_3| - \cdots \pm |E_m|. \tag{C.13}$$

When m is odd, Equation C.13 can be expressed as

$$\begin{aligned} E_{total} &= \frac{|E_1|}{2} + \left(\frac{|E_1|}{2} - |E_2| + \frac{|E_3|}{2} \right) + \left(\frac{|E_3|}{2} - |E_4| + \frac{|E_5|}{2} \right) + \cdots \\ &\quad + \left(\frac{|E_{m-2}|}{2} - |E_{m-1}| + \frac{|E_m|}{2} \right) + \frac{|E_m|}{2} \end{aligned} \tag{C.14}$$

and when m is even, Equation C.13 can be expressed as

$$\begin{aligned} E_{total} &= \frac{|E_1|}{2} - \frac{|E_2|}{2} - \left(\frac{|E_2|}{2} - |E_3| + \frac{|E_4|}{2} \right) \\ &\quad - \left(\frac{|E_4|}{2} - |E_5| + \frac{|E_6|}{2} \right) + \cdots \\ &\quad + \left(\frac{|E_{m-3}|}{2} - |E_{m-2}| + \frac{|E_{m-1}|}{2} \right) - \frac{|E_{m-1}|}{2} + |E_m|. \end{aligned} \tag{C.15}$$

There are now two cases to consider with respect to E_i and its two neighbors E_{i-1} and E_{i+1}: E_i is either greater than or less than the arithmetic mean of its two neighbors. When it is less e.g., when

$$|E_i| > \frac{|E_{i-1} + E_{i+1}|}{2}, \tag{C.16}$$

then each of the terms in the brackets is negative and therefore, from Equation C.15

$$E_{total} < \frac{|E_1|}{2} + \frac{|E_m|}{2} \tag{C.17}$$

and from Equation C.16

$$E_{total} > |E_1| - \frac{|E_2|}{2} + \frac{|E_{m-1}|}{2} + |E_m|. \tag{C.18}$$

The variation between adjacent zones can be ignored given that the obliquity factor $K(\theta)$ goes from 0 to 1 over large number of zones. For the same reason, Equation C.18 above is given as

$$E > \frac{|E_1|}{2} + \frac{|E_m|}{2}. \tag{C.19}$$

Considering Equations C.17 and C.19, E_{total} can be approximated as

$$E_{total} \approx \frac{|E_1|}{2} + \frac{|E_m|}{2}. \tag{C.20}$$

Similarly, the same conclusion can be drawn when

$$|E_i| < \frac{|E_{i-1}| + |E_{i+1}|}{2}. \tag{C.21}$$

Referring back to Equation C.13, when m of the last term $|E_m|$ is even, the same approach as above results in

$$E_{total} \approx \frac{|E_1|}{2} - \frac{|E_m|}{2}. \tag{C.22}$$

According to Fresnel, the obliquity factor $K(\theta)$ was such that the last Fresnel zone occurs at $\theta = 90^\circ$. In other words,

$$K(\theta) = 0 \ \text{ for } \ \pi/2 \leq |\theta| \leq \pi \tag{C.23}$$

and therefore, it can be deduced that the disturbance generated by the entire unobstructed wavefront is approximately equal to one half of the contribution of the first zone [79]. Mathematically,

$$E_{total} \approx \frac{|E_1|}{2} \tag{C.24}$$

when $|E_m|$ reduces to zero since $K_m(\pi/2) = 0$ (when considering Fresnel's initial formulation).

Fresnel's original obliquity factor formulation was later revised by Kirchoff with the introduction of Kirchoff's *correct obliquity factor* that was similar to Fresnel's formulation except it divides the entire spherical wavefront into zones (e.g., $\theta = 180^\circ$ in contrast to Fresnel who considered only half the spherical wavefront (e.g., $\theta = 90^\circ$). When considering Kirchoff's "correct" obliquity factor, the obliquity factor corresponding to the last (m^{th}) zone is equal to zero (e.g., $K_m(\pi) = 0$) and therefore $|E_m| = 0$ also resulting in $E_{total} \approx |E_1|/2$.

C.2 Finding the Position of a Secondary Source Within a Fresnel Zone

The distance between secondary sources in adjacent zones (e.g., between zones Z_1 and Z_2) is $\lambda/2$. Referring to Figure C.3 and to the original Huygens-Fresnel geometry, Equation C.5 provides an expression for r^2, where r is the distance between the receiver and the secondary source in a particular Fresnel zone. By rearranging Equation C.5, an expression for ϕ, the angle between the line connecting the sound source and receiver and the line from the sound source to the secondary source, can be determined

$$cos\phi = \frac{r^2 - \rho^2 + (\rho + r_\circ)^2}{-2\rho(\rho + r_\circ)}. \tag{C.25}$$

Equation C.6 from the original Huygens-Fresnel geometry provides an expression for the value of $2rdr$ (e.g., $2rdr = 2\rho(\rho + r_\circ)\sin\varphi d\varphi$), where dr is the difference in distance between the receiver and the secondary sources between adjacent differential areas dS. Since a particular Fresnel zone is comprised of several differential areas, the value of dr is not necessarily equal to $\lambda/2$. However, here dr is set to a value of $\lambda/2$ thus representing an adjacent Fresnel zone as opposed to an adjacent differential area within a zone. With the values of both dr and ϕ, by rearranging Equation C.6, an expression for $d\phi$ is obtained

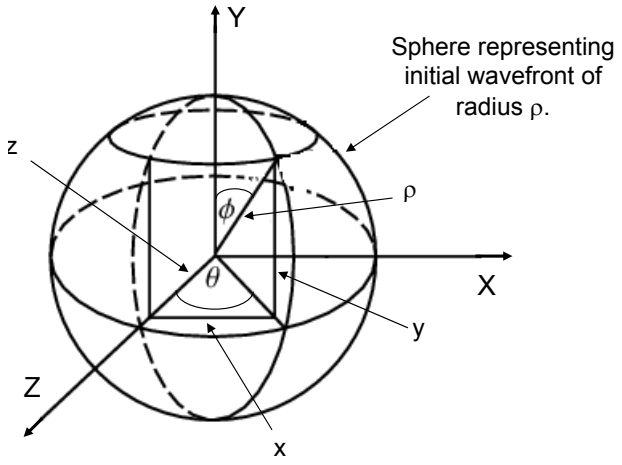


Figure C.4: Spherical coordinates.

$$d\phi = \frac{2rdr}{2\rho(\rho + r_\circ)sin\phi}. \tag{C.26}$$

Referring to Figure C.3, since the elevation angle θ and the radius of the initial wavefront ρ remain constant, the position of a secondary source in the adjacent zone can now be determined by solving for each of its x,y,z coordinates using the equations for the Cartesian coordinates (see Figure C.4) of the sphere along with the previously computed value of $d\phi$

$$x = x_s + (\rho sin(\theta)sin(\phi + d\phi)) \tag{C.27}$$

$$y = y_s + (\rho sin(\theta)cos(\phi + d\phi)) \tag{C.28}$$

$$z = z_s + (\rho cos(\theta)). \tag{C.29}$$

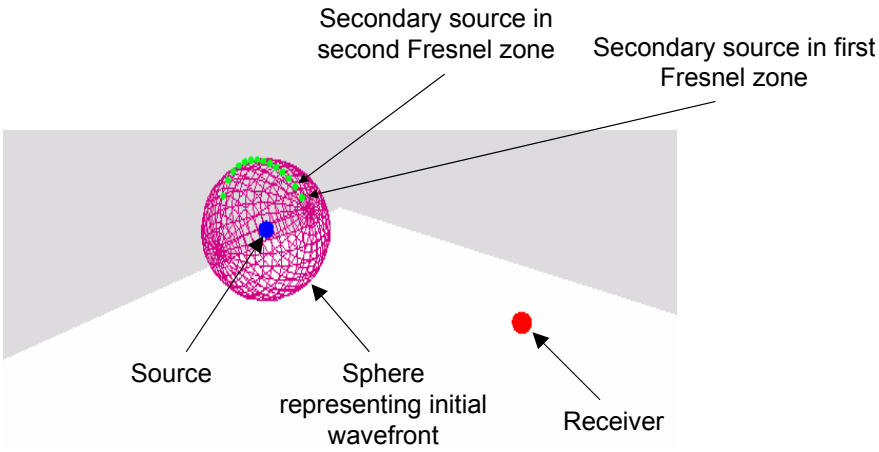


Figure C.5: Sampling a secondary source within a particular Fresnel zone considering a 63Hz sound source.

Figure C.5 provides a graphical illustration of the sampling of a secondary source in each Fresnel zones for a 63Hz sound source.

C.3 Sampling Positions Within a Fresnel Zone

Given the position of a secondary sources within the i^{th} Fresnel zone (denoted by p_i), other positions within the zone can be uniformly sampled (e.g., equal spacing between sampled positions)[1]. Since position p_i is known, its vertical and horizontal angles (ϕ_p and θ_p respectively) relative to the sound source are given as

$$\phi_p = cos^{-1}(\frac{y_p - y_s}{\rho}), \quad \theta_p = tan^{-1}\left(\frac{x_p - x_s}{z_p - z_s}\right). \tag{C.30}$$

The uniform sampling of the i^{th} Fresnel zone is accomplished in several steps.

[1]Positions can also be sampled following other, non-uniform distributions as well although in this dissertation, only uniform sampling is considered.

Initially, samples along a "ring" of constant vertical angle $\phi = \phi_p$ are obtained while varying the horizontal angle θ from 0° to 360° in constant intervals (see Figure C.6(a)). The angular resolution between each sample along this ring of constant ϕ (denoted by θ_{res}) is dependent on the total number of samples considered

$$\theta_{res} = \frac{2\pi}{N_{vis}}. \tag{C.31}$$

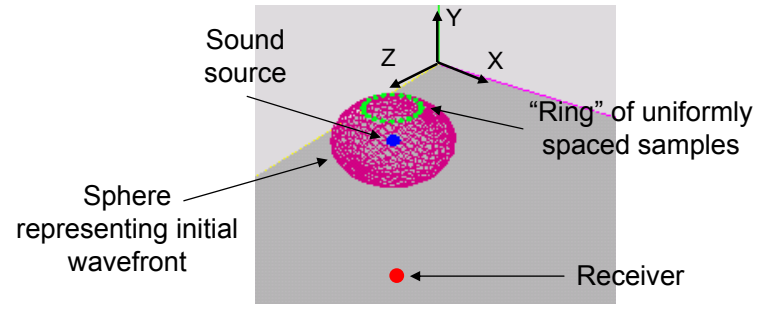
Positions along the ring of constant ϕ are sampled as follows

$$x = x_s + (\rho sin(\theta) sin(\pi - \phi_p)) \tag{C.32}$$

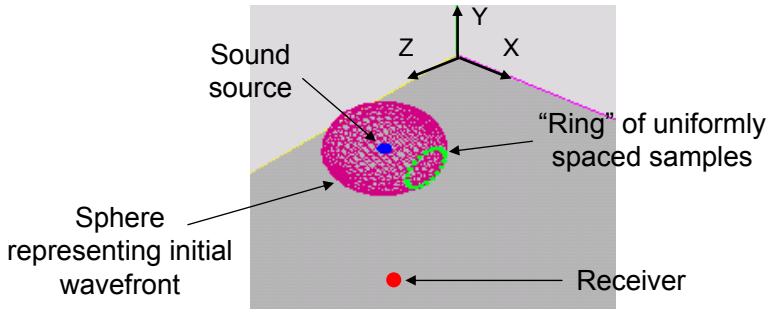
$$y = y_s + (\rho sin(\theta) cos(\pi - \phi_p)) \tag{C.33}$$

$$z = z_s + (\rho cos(\pi - \phi_p)) \tag{C.34}$$

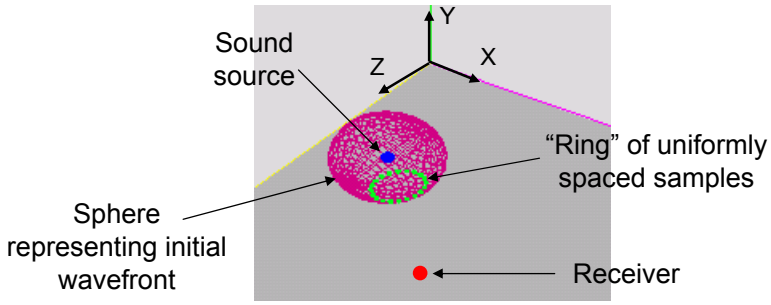
and before sampling each additional position, θ is incremented by θ_{res}. Once a position along the ring of constant ϕ is determined two rotations are performed in order to move the sampled position to the i^{th} Fresnel zone. The first rotation rotates the sampled position about the z-axis by an angle equal to $-\phi_s/2$ (where, ϕ_s is the vertical angle of the secondary source position in the i^{th} Fresnel zone), thus moving the sampled position to a "new" position p_{rz} (see Figure C.6(b)).



(a) Ring of samples of constant ϕ.



(b) Rotation of ring about the z-axis.



(c) Rotation of ring about the y-axis.

Figure C.6: Uniformly sampling positions within a Fresnel zone. (a) A "ring" of samples of constant vertical angle ϕ. (b) The "ring" is rotated about the z-axis by and amount equal to $-\pi_s/2$ (where, $-\pi_s/2$ is the is the secondary source vertical angle.). (c) The "ring" is rotated again but this time about the y-axis by an amount equal to $-(\pi/2-\theta_{rec})$ (where θ_{rec} is the horizontal angle between the receiver and the sound source).

The second rotation rotates p_{rz} about the y-axis by an angle equal to $-(\pi/2-\theta_{receiver})$ (where $\theta_{receiver}$ is the horizontal angle between the receiver and the sound source), thus moving p_{rz} to the i^{th} Fresnel zone (see Figure C.6(c)).

Bibliography

[1] T. D. Abhayapala and D. B. Ward. Theory and design of high order sound field microphones using spherical microphone array. In *Proceedings of the 2002 International Conference on Acoustics Speech and Signal Processing*, pages 1949–1952, Orlando, FL. USA, May 13-17, 2002.

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

[3] J. Amanatides. Ray tracing with cones. In *Proceedings of the 11^{th} Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 1984)*, pages 129–135, Minneapolis, MN. USA, 1984.

[4] D. B. Anderson and M. A. Casey. The sound dimension. *IEEE Spectrum*, pages 46–50, March 1997.

[5] Y. Ando. *Concert Hall Acoustics*. Springer-Verlag, Berlin, Germany, 1985.

[6] A. Appel. Some techniques for shading machine renderings and solids. In *Proceedings of the American Federation of Information Processing Societies (AFIPS) Spring Joint Computer Conference (SJCC)*, volume 32, pages 37–45, Atlantic City, NJ. USA, 1968.

[7] J. Arvo and B. Kirk. Particle transport and image synthesis. In *Proceedings of the 17^{th} Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 1990)*, pages 63–66, Dallas, TX. USA, August, 1990.

[8] B. S. Atal and M. R. Schroeder. Apparent sound source translator. U.S. Patent $3,236,949$, February 23 1963.

[9] H. E. Bass, L. C. Sutherland, and A. J. Zuckerwar. Atmospheric absorption of sound: update. *Journal of the Acoustical Society of America*, 88(4):2019–2021, 1990.

[10] D. W. Batteau. The role of the pinna in human localization. *Proceedings of the Royal Society of London*, pages 158–180, 1967.

[11] B. Bauer. Phasor analysis of some stereophonic phenomena. *Journal of the Acoustical Society of America*, 33(11):1536–1539, 1961.

[12] B. Bauer. Stereophonic earphones and binaural loudspeakers. *Journal of the Audio Engineering Society*, 9(2):148–151, 1961.

[13] B. G. Baumgart. Winged edge polyhedron representation. Technical Report AIM-179 (CS-TR-74-320), Stanford University, Palo Alto CA. USA, October 1972.

[14] E. B. Becker, G. F. Carey, and J. T. Oden. *Finite Elements, An Introduction, Volume 1*. Prentice Hall Publishers, Englewood Cliffs, NJ. USA, 1981.

[15] D. R. Begault. Perceptual effects of synthetic reverberation on three-dimensional audio systems. *Journal of the Audio Engineering Society*, 40(11):895–904, 1992.

[16] D. R. Begault. Auditory and non-auditory factors that potentially influence virtual acoustic imagery. In *Proceedings of the Audio Engineering Society* 16^{th} *International Conference on Spatial Sound Reproduction*, pages 1–14, Rovanciemi, Finland, 1999.

[17] R. Begault. *3-D Sound for Virtual Reality and Multimedia*. Academic Press Professional, Cambridge, MA. USA, 1994.

[18] G. V. Bekesy. *Experiments in Hearing*. McGraw Hill, New York, NY. USA, 1960.

[19] J. C. Bennett, K. Barker, and F. O. Edeko. A new approach to the assessment of stereophonic sound system performance. *Journal of the Audio Engineering Society*, 33(5):314–321, 1985.

[20] J. L. Bentley. Multidimensional binary search trees used for associative searching. *Communications of the ACM*, 18(9):509–517, 1975.

[21] J. L. Bentley. Multidimensional binary search trees in database applications. *IEEE Transactions on Software Engineering*, 5(4):333–340, 1979.

[22] A. J. Berkhout, D. de Vries, and P. Vogel. Acoustic control by wave field synthesis. *Journal of the Acoustical Society of America*, 93(5):2764–2778, 1993.

[23] M. Bertram, E. Deines, J. Mohring, J. Jegorovs, and H. Hagen. Phonon tracing for auralization and visualization of sound. In *Proceedings of the IEEE Visualization 2005*, pages 151–158, Minneapolis, MN. USA., 2005.

[24] M. A. Biot and I. Tolstoy. Formulation of wave propagation in infinite media by normal coordinates with an application to diffraction. *Journal of the Acoustical Society of America*, 29(3):381–391, 1957.

[25] J. Blauert. *Spatial Hearing: The Psychophysics of Human Sound Localization*. MIT Press, Cambridge, MA. USA, 1983.

[26] J. Blauert. *The Psychophysics of Human Sound Localization*. MIT Press, Cambridge, MA. USA, revised edition, 1996.

[27] M. M. Boone. Acoustic rendering with wave feld systhesis. In *Proceedings of the ACM SIGGRAPH and Eurographics Campfire on Acoustic Rendering for Virtual Environments*, pages 37–45, Snowbird, UT. USA, May 26-29, 2001.

[28] M. M. Boone and W. P. J. de Bruijn. On the applicability of distributed mode loudspeaker panels for wave field synthesis based sound reproduction. In *Proceedings of the 108th Convention of the Audio Engineering Society*, Paris, France, February 19-22, 2000.

[29] M. M. Boone, D. de Vries, and P. F. van Tol. Spatial sound field reproduction by wave field synthesis. *Journal of the Audio Engineering Society*, 43(12):1003–1012, 1995.

[30] M. Born and E. Wolf. *Principles of Optics*. Pergamon Press, New York, NY. USA, 7 edition, 1999.

[31] D. S. Brungart. Control of perceived distance in virtual audio displays. In *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.*, pages 1101–1104, Hong Kong, October 29-November 1, 1998.

[32] D. S. Brungart and W. M. Rabinowitz. Auditory localization of nearby sources. Head related transfer functions. *Journal of the Acoustical Society of America*, 106(3):1465–1479, 1999.

[33] D. S. Brungart and W. R. Rabinowitz. Auditory localization in the near-field. In *Proceedings of the International Conference on Auditory Displays (ICAD 1996)*, Palo Alto, CA. USA, November 4 - 6, 1996.

[34] D. S. Brungart and K. R. Scott. The effects of production and presentation level on the auditory distance perception of speech. *Journal of the Acoustical Society of America*, 110(1):425–440, 2001.

[35] W. Buxton. Using our ears: An introduction to the use of nonspeech audio cues. In E. Farrell, editor, *Proceedings of the SPIE: Extracting Meaning from Complex Data: Processing, Display, Interaction*, volume 1259, pages 124–127, Santa Clara, CA. USA, February 14-16, 1990.

[36] P. T. Calamia, U. P. Svensson, and T. A. Funkhouser. Integration of edge-diffraction calculations and geometrical-acoustics modeling. In *Proceedings of Forum Acusticum 2005*, Budapest, Hungary, August 29-September 2, 2005.

[37] S. Carlile. *Virtual Auditory Space: Generation and Application*. R. G. Landes Company, Austin, TX. USA, 1996.

[38] M. Casey, W. Gardner, and S. Basu. Vision steered beamforming and transaural rendering for the artificial life interactive video environment (ALIVE). In *Proceedings of the 99th Convention of the Audio Engineering Society*, New York, NY. USA, 1996.

[39] J. Chowning. The simulation of moving sound sources. *Journal of the Audio Engineering Society*, 19(1):2–6, 1971.

[40] J. M. Chowning. Digital sound synthesis, acoustics and perception: A rich intersection. In *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00).*, Verona, Italy, December, 2000.

[41] C. L. Christensen and J. H. Rindel. Predicting acoustics in class rooms. In *Proceedings of the 34th International Congress and Exposition on Noise Control Engineering (Inter-Noise 2005)*, Rio de Janeiro, Brazil, August 7-10 , 2005.

[42] M. Cohen and N. Koizumi. Virtual gain for audio windows. *Presence: Teleoperators and Virtual Environments*, 7(1):53–66, 1999.

[43] M. Cohen and E. Wenzel. The design of multidimensional sound interfaces. In W. Barfield and T. Furness, editors, *Virtual Environments and Advanced Interface Design*, chapter 8, pages 291–346. Oxford University Press Inc., New York, NY. USA, 1995.

[44] M. F. Cohen, S. E. Chen, J. R. Wallace, and D. P. Greenberg. A progressive refinement approach for fast radiosity image generation. In *Proceedings of the 15th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 1988)*, pages 75–84, Atlanta, GA. USA, August, 1988.

[45] M. F. Cohen and D. P. Greenberg. The hemi-cube: a radiosity solution for complex environments. In *Proceedings of the 12th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 1985)*, volume 19, pages 31–40, San Francisco, CA. USA, 1985.

[46] M. F. Cohen and J. R. Wallace. *Radiosity and Realistic Image Synthesis*. Morgan Kaufmann Publishers, Inc., San Francisco, CA. USA, 1993.

[47] P. D. Coleman. An analysis of cues to auditory depth perception in free space. *Psychological Bulletin*, 60:302–315, 1963.

[48] L. Cremer and H. A. Müller. *Principles and Applications of Room Acoustics*, volume 1. Applied Science Publishers LTD., Barking, Essex, UK, 1978.

[49] N. Dadoun, D. G. Kilpatrick, and J. P. Walsh. The geometry of beam tracing. In *Proceedings of the Symposium on Computational Geometry*, pages 55–71, Baltimore, MD. USA, June 1985.

[50] B. I. Dalenbäck. Room acoustic prediction based on a unified treatment of diffuse and specular reflection. *Journal of the Acoustical Society of America*, 100(2):899–909, 1996.

[51] B. I. Dalenbäck, M. Kleiner, and P. Svensson. A macroscopic view of diffuse reflection. *Journal of the Audio Engineering Society*, 42(10):793–806, 1994.

[52] S. M. Dance and B. M. Shield. Modeling of sound fields in enclosed spaces with absorbent room surfaces. Part I: Performance spaces. *Applied Acoustics*, 58(1):1–18, 1999.

[53] D. de Vries and M. M. Boone. Wave field synthesis and analysis using array technology. In *Proceedings of the 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 15–18, New Paltz, NY. USA, October 17-20, 2004.

[54] C. S. Desai and J. F. Abel. *Introduction to the Finite Element Method*. Von Nostrand Reinhold, New York, NY. USA, 1972.

[55] K. Doel, P. G. Kry, and D. K. Pai. FOLEYAUTOMATIC: Physically-based sound effects for interactive simulation and animation. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 2001)*, pages 537–544, Los Angeles, CA. USA, August 12-1,7 2001.

[56] J. R. Driscoll and D. M. Healy. Computing Fourier transforms and convolutions on the 2-sphere. *Advances in Applied Mathematics*, 15(2):202–250, 1994.

[57] R. O. Duda and W. L. Martens. Range dependence of the response of a spherical head model. *Journal of the Acoustical Society of America*, 104(5):3048–3058, 1998.

[58] J. J. Embrechts. Broad spectrum diffusion model for room acoustics ray-tracing algorithms. *Journal of the Acoustical Society of America*, 107(4):2068–2080, 2000.

[59] F. A. Everest. *Master Handbook of Acoustics*. McGraw-Hill, New York, NY. USA, 4 edition, 2001.

[60] P. Fellgett. Ambisonics. Part one. General system description. *Studio Sound*, 17(8):20–22, 1975.

[61] H. G. Fisher and S. J. Freedman. The role of the pinna in auditory localization. *Journal of Auditory Research*, 8(1):15–26, 1968.

[62] H. Fletcher and W. A. Munson. Loudness, its definition, measurement and calculation. *Journal of the Acoustical Society of America*, 5(2):82–108, 1933.

[63] T. Funkhouser, N. Tsingos, I. Carlbom, G. Elko, G. Pingali, M. Sondhi, and J. E. West. A beam tracing approach to acoustic modeling for interactive virtual environments. In *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 1998)*, pages 21–32, Orlando, FL. USA, July, 1998.

[64] T. Funkhouser, N. Tsingos, I. Carlbom, G. Elko, M. Sondhi, J. E. West, G. Pingali, P. Min, and A. Ngan. A beam tracing method for interactive architectural acoustics. *Journal of the Acoustical Society of America*, 115(2):739–756, 2004.

[65] J. Garas. *Adaptive 3D Sound Systems*. Kluwer Academic Publishers, Norwell, MA. USA, 2000.

[66] M. B. Gardner. Distance estimation of 0° or apparent 0° oriented speech signals in anechoic space. *Journal of the Acoustical Society of America*, 45(1):47–53, 1968.

[67] W. Gardner. *3-D Audio Using Loudspeakers*. Kluwer Academic Publishers, Norwell, MA. USA, 1998.

[68] M. Gerzon. Ambisonics. Part two: studio techniques. *Studio Sound*, 17(8):24, 26, 28–30, 1975.

[69] R. A. Goldstein and R. Nagel. 3-D visual simulation. *Simulation*, 16(1):25–31, 1971.

[70] C. M. Goral, K. E. Torrance, and D. P. Greenberg. Modeling the interaction of light between diffuse surfaces. In *Proceedings of the 11^{th} Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 1984)*, volume 18, pages 213–222, Orlando, FL. USA, 1984.

[71] D. Hammershøi and H. Møller. Methods for binaural recording and reproduction. *Acta Acustica*, 88(3):302–311, 2002.

[72] J. M. Hammersley and D. C. Handscomb. *Monte Carlo Methods*. Chapman and Hall, New York NY. USA, 1964.

[73] T. J. Hargreaves, T. J. Cox, and Y. W. Lam. Surface diffusion coefficients for room acoustics: far field measures. *Journal of the Acoustical Society of America*, 108(4):1710–1720, 2000.

[74] C. M. Harris. Absorption of sound in air versus humidity and temperature. *Journal of the Acoustical Society of America*, 40(1):148–159, 1966.

[75] W. M. Hartmann. The physical description of signals. In B. C. J. Moore, editor, *Hearing.*, Handbook of Perception and Cognition, chapter 1, pages 1–40. Academic Press Inc., San Diego, CA. USA, 1995.

[76] W. M. Hartmann. Listening in a room and the precedence effect. In R. H. Gilkey and T. R. Anderson, editors, *Binaural and Spatial Hearing in Real and Virtual Environments*, pages 191–210. Lawrence Erlbaum Associates, Mahwah, NJ. USA, 1997.

[77] W.H. Hartmann. How we localize sound. *Physics Today*, pages 24–29, November 1999. http://www.aip.org/pt/nov99/locsound.html.

[78] D. D. Hearn and M. P. Baker. *Computer Graphics C Version*. Prentice-Hall, Upper Saddle River, NJ. USA, 2 edition, 1997.

[79] E. Hecht. *Optics*. Pearson Education Inc., San Francisco, CA. USA, 4 edition, 2002.

[80] P. Heckbert and P. Hanrahan. Beam tracing polygonal objects. In *Proceedings of the 11th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 1984)*, volume 18, pages 119–127, Orlando, FL. USA, July, 1984.

[81] M. Hodgson. Evidence of diffuse surface reflections in rooms. *Journal of the Acoustical Society of America*, 89(2):765–771, 1991.

[82] M. Hodgson. Ray-tracing evaluation of empirical models for predicting noise in industrial workshops. *Applied Acoustics*, 64(11):1033–1048, 2003.

[83] D. Foley J, A. van Dam, S. K. Feiner, J. F. Hughes, and R. L. Phillips. *Introduction to Computer Graphics*. Addison-Wesley Publishing Co., Reading, MA. USA, 1994.

[84] H. W. Jensen. *Realistic Image Synthesis Using Photon Mapping*. A. K. Peters, Natick, MA. USA, 2001.

[85] H. W. Jensen and N. J. Christensen. Photon maps in bidirectional monte carlo ray tracing of complex objects. *Computers and Graphics*, 19(2):215–224, 1995.

[86] J. Kajiya. The rendering equation. In *Proceedings of the 13th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 1986)*, volume 20, pages 143–150, Dallas, TX. USA, 1986.

[87] M. H. Kalos and P. A. Whitlock. *Monte-Carlo Methods*, volume 1. John Wiley and Sons., New York, NY. USA, 1986.

[88] B. Kapralos, M. Jenkin, and E. Milios. Auditory perception and spatial (3D) auditory systems. Technical Report CS-2003-07, Department of Computer Science, York University, Toronto, Ontario, Canada, July 2002.

[89] B. Kapralos, M. Jenkin, and E. Milios. Sonel mapping: acoustic modeling utilizing an acoustic version of photon mapping. In *Proceedings of the IEEE International Workshop on Haptics Audio Visual Environments and their Applications (HAVE 2004)*, pages 1–6, Ottawa, Ontario, Canada, October 2-3, 2004.

[90] B. Kapralos, M. Jenkin, and E. Milios. Acoustical diffraction modeling utilizing the Huygens-Fresnel principle. In *Proceedings of the IEEE International Workshop on Haptics Audio Visual Environments and their Applications (HAVE 2005)*, pages 39–44, Ottawa, Ontario, Canada, October 1-2, 2005.

[91] B. Kapralos, M. Jenkin, and E. Milios. Acoustical modeling using a Russian roulette strategy. In *Proceedings of the 118th Convention of the Audio Engineering Society*, Barcelona, Spain, May 28-31 2005.

[92] B. Kapralos, M. Jenkin, and E. Milios. Sonel mapping: A stochastic acoustical modeling system. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Toulouse, France, May 14-19, 2006.

[93] J. B. Keller. Geometrical theory of diffraction. *Journal of the Optical Society of America*, 52(2):116–130, 1962.

[94] G. Kendall. A 3D sound primer: Directional hearing and stereo reproduction. *Computer Music Journal*, 19(4):23–46, 1995.

[95] M. Kleiner and D. DalenbackandP.Svensson. Auralization - an overview. *Journal of the Audio Engineering Society*, 41(11):861–875, 1993.

[96] U. Kristiansen, A. Krokstad, and T. Follestad. Extending the image method to higher-order reflections. *Journal of Applied Acoustics*, 38(2-4):195–206, 1993.

[97] A. Krokstad, S. Strom, and S. Sorsdal. Calculating the acoustical room response by the use of a ray tracing technique. *Journal of Sound and Vibration*, 8(1):118–125, 1968.

[98] A. Kulkarni and S. Colburn. Variability in the characterization of the headphone transfer-function. *Journal of the Acoustical Society of America*, 107(2):1071–1074, 2000.

[99] A. Kulowski. Algorithmic representation of the ray tracing technique. *Applied Acoustics*, 18(6):449–469, 1985.

[100] H. Kuttruff. *Room Acoustics*. Spon Press, London, England, 4 edition, 2000.

[101] K. H. Kuttruff. Auralization of impulse responses modeled on the basis of ray-tracing results. *Journal of the Audio Engineering Society*, 41(11):876–880, 1993.

[102] Dolby Laboratories. Dolby digital - the sound of the future here today. Technical report, Dolby Laboratories, San Francisco, CA. USA, 1998.

[103] Y. W. Lam. A comparison of three different diffuse reflection modeling methods used in room acoustic computer models. *Journal of the Acoustical Society of America*, 100(4):2181–2192, 1996.

[104] T. Lewers. A combined beam tracing and radiant exchange computer model of room acoustics. *Applied Acoustics*, 38(2-4):161–178, 1993.

[105] T. Lokki, P. Svensson, and L. Savioja. An efficient auralization of edge diffraction. In *Proceedings of the Audio Engineering Society 21st International Conference on Architectural Acoustics and Sound Reinforcement*, pages 317–325, St. Petersburg, Russia, June 1-3, 2002.

[106] W. L. Martens. Efficient auralization of small, cluttered spaces: Simulating sonic obstructions at close range. In S. Kuwano and T.Kato, editors, *Proceedings of the 7th Western Pacific Regional Acoustics Conference*, pages 317–320, Kumamoto, Japan, October 3-5 2000.

[107] W. L. Martens. Binaural range display with independent control over loudness: Presenting nearby whispers in the dark. In *Proceedings of the International Conference on Artificial Reality and Telexistence (ICAT))*, pages 100–109, University of Tokyo, Japan, December 5-7 2001.

[108] W. L. Martens. Psychophysical calibration for controlling the range of a virtual sound source: Multidimensional complexity in spatial auditory display. In *Proceedings of the 2001 International Conference on Auditory Display*, pages 1–12, Espoo, Finland, July 29 - August 1 2001.

[109] W. L. Martens. Uses and misuses of psychophysical methods in the evaluation of spatial sound reproduction. In *Proceedings of the 110th Convention of the Audio Engineering Society*, Amsterdam, Netherlands, May 12-15 2001.

[110] W. L. Martens. Rapid psychophysical calibration using bisection scaling for individualized control of source elevation in auditory display. In *Proceedings of the 2002 International Conference on Auditory Display (ICAD 2002)*, pages 1–8, Kyoto, Japan, July 2-5 2002.

[111] W. L. Martens. Perceptual evaluation of filters controlling source direction: Customized and generalized hrtfs for binaural synthesis. *Acoustical Science and Technology*, 24(5):220–232, 2003.

[112] W. L. Martens and J. Herder. Perceptual criteria for eliminating reflectors and occluders from the rendering of environmental sound. In *137th Meeting of the Acoustical Society of America and the 2nd Convention of the European Acoustics Association*, page S53, Berlin, March 14-19 1999. Acoustical Society of America and European Acoustics Association.

[113] W. L. Martens, J. Herder, and Y. Shiba. A filtering model for efficient rendering of the spatial image of an occluded virtual sound source. In *Proceedings of the 137th Meeting of the Acoustical Society of America and the 2nd Convention of the European Acoustics Association*, page S54, Berlin, Germany, March 1999.

[114] W. L. Martens and K. Myszkowski. Psychophysical validation of the visible differences predictor for global illumination applications. In C. M. Wittenbrink and A. Varshney, editors, *Late Breaking Hot Topics Proceedings of the IEEE Visualization '98)*, pages 49–52. IEEE Computer Society, October 18-23 1998.

[115] W. L. Martens and W. Woszczyk. Guidelines for enhancing the sense of presence in virtual acoustic environments. In H. Thwaites, editor, *Proceedings of the 9th International Conference on Virtual Systems and Multimedia*, pages 306–313, Montreal, Quebec, Canada, October 15-17 2003.

[116] R. L. Martin, K. I. Mcanally, and M. A. Senova. Free-field equivalent localization of virtual audio. *Journal of the Audio Engineering Society*, 49(1/2):14–22, 2001.

[117] G. Martins. *A Hybrid Model for Simulating Diffuse First Reflections in Two-dimensional Synthetic Acoustic Environments*. PhD thesis, Schulich School of Music, McGill University, Montreal, Quebec. Canada, March 2001.

[118] L. Martins, J. H. Rindel, and C. L. Christensen. Predicting the acoustics of ancient open-air theatres: the importance of calculation methods and geometrical details. In *Proceedings of the Joint Baltic-Nordic Acoustics Meeting 2004*, pages 1–8, Mariehamn, Aland, Finland, June 8-10, 2004.

[119] A. Le Bot and A Bocquillet. Comparison of an integral equation on energy and the ray-tracing technique in room acoustics. *Journal of the Acoustical Society of America*, 108(4):1732–1740, 2000.

[120] Riitta Väänänen. *Parameterization, Auralization and Authoring of Room Acoustics for Virtual Reality Applications*. PhD thesis, Helsinki University of Technology, Helsinki, Finland, May 10 2003.

[121] K. I. Mcanally and R. L. Martin. Variability in the headphone-to-ear-canal transfer function. *Journal of the Audio Engineering Society*, 50(4):263–266, 2002.

[122] M. Mehta, J. Johnson, and J. Rocafort. *Architectural Acoustics Principles and Design*. Prentice Hall, Inc., Upper Saddle River, NJ. USA, 1999.

[123] D. H. Mershon, W. L. Ballenger, W. L. Little, P.L. Mcmurtry, and J. L. Buchanan. Effects of room reflectance and background noise on perceived auditory distance. *Perception*, 18(3):403–416, 1989.

[124] D. H. Mershon and J. N. Bowers. Absolute and relative cues for the auditory perception of egocentric distance. *Perception*, 8(3):311–322, 1979.

[125] D. H. Mershon and L. E. King. Intensity and reverberation as factors in the auditory perception of egocentric distance. *Perception and Psychophysics*, 18(6):409–415, 1975.

[126] J. Meyer and G. W.Elko. A spherical microphone array for spatial sound recording. *Journal of the Acoustical Society of America*, 111(5):2346, 2002.

[127] J. C. Middlebrooks. Narrow-band sound localization related to external ear acoustics. *Journal of the Acoustical Society of America*, 92(5):2607–2624, 1992.

[128] R. N. Miles. Sound field in a rectangular enclosure with diffusely reflecting boundaries. *Journal of Sound and Vibration*, 92(2):203–226, 1984.

[129] H. Møller. Fundamentals of binaural technology. *Applied Acoustics*, 36(1-2):171–218, 1992.

[130] H. Møller, D. Hammershoi, C. B. Jensen, and M. F. Sorensen. Transfer characteristics of headphones measured on human ears. *Journal of the Audio Engineering Society*, 43(4):203–217, 1995.

[131] M. Monks, B. M. Oh, and J. Dorsey. Acoustic simulation and visualization using a new unified beam tracing and image source approach. In *Proceedings of the 101st Convention of the Audio Engineering Society*, Los Angeles, CA. USA, November 8-11, 1996.

[132] B. C. J. Moore. *An Introduction to the Psychology of Hearing*. Academic Press Limited, San Diego, CA. USA, 3 edition, 1989.

[133] B. C. J. Moore, B. R. Glassberg, and Thomas Baer. A model for the prediction of thresholds, loudness and partial loudness. *Journal of the Audio Engineering Society*, 45(4):224–239, 1997.

[134] R. F. Moore. *Elements of Computer Music*. Prentice-Hall, Englewood Cliffs, NJ. USA, 1990.

[135] A. Mouchtaris, P. Reveliotis, and C. Kyriakakis. Inverse filter design for immersive audio rendering over loudspeakers. *IEEE Transactions on Multimedia*, 2(2):77–87, 2000.

[136] M. Naguib and H. Wiley. Estimating the distance to a sound: mechanisms and adaptations for long-range communications. *Animal Behavior*, 62(5):825–837, 2001.

[137] S. H. Nielsen. Auditory distance perception in different rooms. *Journal of the Audio Engineering Society*, 41(10):755–770, 1993.

[138] E. Nosal, M. Hodgson, and I. 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, 2004.

[139] J. F. O'Brien, P. R. Cook, and G. Essl. Synthesizing sounds from physically based motions. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 2001)*, pages 529–536., Los Angeles, CA. USA, August 12-17, 2001.

[140] S. R. Oldfield and S. A. Parker. Acuity of sound localization: a topography of auditory space II: Pinna cues absent. *Perception*, 13(5):601–617, 1984.

[141] A. M. Ondet and J. L. Barbry. Modeling of sound propagation in fitted workshops using ray tracing. *Journal of the Acoustical Society of America*, 85(2):787–796, 1989.

[142] A. Papoulis and U. Pillai. *Probability, Random Variables, and Stochastic Processes*, volume 4. McGraw-Hill, New York, NY. USA, 2001.

[143] M. A. Poletti. A unified theory of horizontal holographic sound systems. *Journal of the Audio Engineering Society*, 48(12):1155–1182, 2000.

[144] V. Pulkki. *Spatial Sound Generation and Perception by Amplitude Panning Techniques*. PhD thesis, Department of Electrical and Communications Engineering, Helsinki University of Technology, Helsinki, Finland, August 2001.

[145] V. Pulkki and M. Karjalainen. Directional quality of 3-D amplitude panned virtual sources. In *Proceedings of the 2001 International Conference on Auditory Display (ICAD 2001)*, pages 239–244, Espoo, Finland, July 29-August 1, 2001.

[146] R. Rabenstein, O. Schips, and A. Stenger. Acoustic rendering of buildings. In *5^{th} International Conference on Building Simulation*, Prague, Czech Republic, September 8-10, 1997.

[147] B. Rafaely. Plane-wave decomposition of the sound field on a sphere by spherical convolution. *Journal of the Acoustical Society of America*, 116(4):2149–2157, 2004.

[148] B. Rafaely. Analysis and design of spherical microphone arrays. *IEEE Transactions on Speech and Audio Processing*, 13(1):135–143, 2005.

[149] B. Rakerd and W. M. Hartmann. Localization of sound in rooms, II: the effects of a single reflecting surface. *Journal of the Acoustical Society of America*, 78(2):524–533, 1985.

[150] Rhintek. Computer aided room acoustics (CARA©). http://www.cara.de.

[151] J. H. Rindel. Modeling in auditorium acoustics - from ripple tank and scale models to computer simulations. In *Proceedings of Forum Acusticum*, pages 1–8, Sevilla, Spain, September 16-20 , 2002.

[152] D. W. Robinson and R. S. Dadson. A re-determination of the equal-loudness relations for pure tones. *British Journal of Applied Physics*, 7:166–181, 1956.

[153] H. Robjohns. You are surrounded: Surround sound explained - part 2. *Sound on Sound*, September 2001.

[154] H. Robjohns. You are surrounded: Surround sound explained - part 3. *Sound on Sound*, October 2001.

[155] T. D. Rossing, R. F. Moore, and P. A. Wheeler. *The Science of Sound*. Benjamin Cummings, San Francisco, CA. USA, 2002.

[156] G. Rougeron, F. Gaudaire, Y. Gabilet, and K. Bouatouch. Simulation of the indoor propagation of a 60GHz electromagnetic wave with a time-dependent radiosity algorithm. *Computers and Graphics*, 26(1):125–141, 2002.

[157] R. Y. Rubenstein. *Simulation and the Monte-Carlo Method*. John Wiley and Sons., New York, NY. USA, 1981.

[158] F. Rumsey. *Spatial Audio*. Focal Press, Woburn, MA. USA, 2001.

[159] W. C. Sabine. Theater acoustics. *The American Architect*, 104:257, 1913.

[160] L. Savioja. *Modeling Techniques for Virtual Acoustics*. PhD thesis, Helsinki University of Technology, Telecommunications Software and Multimedia Laboratory, Helsinki, Finland, 1999.

[161] M. R. Schroeder. New method for measuring reverberation time. *Journal of the Acoustical Society of America*, 37(3):409–412, 1965.

[162] M. N. Semple. Sounds in a virtual world. *Nature*, 396(6713):723–724, December 1998.

[163] J. Shi, A. Zhang, J. Encarnacão, and M. Göbel. A modified radiosity algorithm for integrated visual and auditory rendering. *Computers and Graphics*, 17(6):633–642, 1993.

[164] R. D. Shilling and B. Shinn-Cunningham. Virtual auditory displays. In K. Stanney, editor, *Handbook of Virtual Environment Technology*, pages 65–92. Lawrence Erlbaum Associates, Mahwah, NJ. USA, 2002.

[165] B. G. Shinn-Cunningham. Distance cues for virtual auditory space. In *Proceedings of the IEEE 2000 International Symposium on Multimedia Information Processing*, pages 227–230, Sydney, Australia, December, 2000.

[166] B. G. Shinn-Cunningham. Learning reverberation: considerations for spatial auditory displays. In *Proceedings of the International Conference on Auditory Displays (ICAD 2000)*, pages 126–134, Atlanta, GA. USA, December, 2000.

[167] D. Shreiner, M. Woo, J. Neider, and T. Davis. *OpenGL Programming Guide: The Official Guide to Learning OpenGL Version 1.4*. Addison-Wesley, Boston, MA. USA., 4 edition, 2004.

[168] A. Sibbald. Chaotic waves for 3D audio. Technical report, Sensaura Ltd., Middlesex, UK, 2001.

[169] F. X. Sillion and C. Puech. *Radiosity & Global Illumination*. Morgan Kaufmann Publishers Inc., San Francisco, CA. USA, 1994.

[170] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London, England, 1986.

[171] W. H. Slattery and J. C. Middlebrooks. Monaural sound localization: acute versus chronic unilateral impairment. *Hearing Research*, 75(1/2):38–46, 1984.

[172] J. W. Strutt. On our perception of sound direction. *Philosophical Magazine*, 13:214–232, 1907.

[173] F. Suykens. *On Robust Monte Carlo Algorithms for Multi-pass Global Illumination*. PhD thesis, Department of Computer Science, Katholieke Univeriteit Leuven, Leuven, Belgium, 2002.

[174] U. P. Svensson, R. I. Fred, and J. Vaderkooy. An analytic secondary source model of edge diffraction impulse responses. *Journal of the Acoustical Society of America*, 106(5):2331–2344, 1999.

[175] U. P. Svensson and U. R. Kristiansen. Computational modeling and simulation of acoustic spaces. In *Proceedings of the 22nd International Conference on Virtual, Synthetic and Entertainment Audio*, pages 11–30, Espoo, Finland, 2002.

[176] J. R. Thompson and R. A. Tapia. *Nonparametric Function Estimation, Modeling and Simulation*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA. USA, 1990.

[177] W. R. Thurlow, J. W. Mangels, and P. S. Runge. Head movements during sound localization. *Journal of the Acoustical Society of America*, 42(2):489–493, 1967.

[178] R. R. Torres, P. Svensson, and M. Kleiner. Computation of edge diffraction for more accurate room acoustics auralization. *Journal of the Acoustical Society of America*, 109(2):600–610, 2001.

[179] N. Tsingos, I. Carlbom, G. Elko, T. Funkhouser, and B. Kubli. Validation of acoustical simulations in the "Bell Labs Box". *IEEE Computer Graphics and Applications*, 22(4):28–37, July 2002.

[180] N. Tsingos, T. Funkhouser, A. Ngan, and I. Carlbom. Modeling acoustics in virtual environments using the uniform theory of diffraction. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 2001)*, pages 545–552, Los Angeles CA. USA, 2001.

[181] N. Tsingos and J. D. Gascuel. Soundtracks for computer animation: Sound rendering in dynamic environments with occlusion. In *Proceedings of Graphics Interface '97*, pages 9–16, Kelowna, BC. Canada., May 21-23, 1997.

[182] N. Tsingos and J.D. Gascuel. Fast rendering of sound occlusion and diffraction effects for virtual acoustic environments. In *104th Convention of the Audio Engineering Society*, pages 1–14, Amsterdam, The Netherlands, May 16-19, 1998.

[183] M. van der Voorden, L. Nijs, G. Vermeir, and G. Jansens. The applicability of ray-tracing based simulation tools to predict sound pressure levels and reverberation times in coupled spaces. In *Proceedings of the 6th International Building Performance Simulation Association (IBPSA) Conference*, volume 3, pages 1211–1218, Kyoto, Japan, September 13-15, 1999.

[184] E. Veach. *Robust Monte Carlo Methods for Robust Light Transport Simulation*. PhD thesis, Department of Computer Science, Stanford University, Stanford, CA. USA, 1997.

[185] E. N. G. Verheijen. *Sound Reproduction by Wave Field Synthesis*. PhD thesis, Technical University Delft, The Netherlands, 1998.

[186] M Vorländer. Simulation of the transient and steady-state sound propagation in rooms using a new combined ray-tracing/image-source algorithm. *Journal of the Acoustical Society of America*, 86(1):172–178, 1989.

[187] H. Wallach. The role of head movements and vestibular and visual cues in sound localization. *Experimental Psychology*, 27:339–368, 1940.

[188] H. Wallach, E. B. Newman, and M. R. Rosenzweig. The precedence effect in sound localization. *Journal of Psychology*, 52:315–336, 1949.

[189] D. B. Ward and G. W. Elko. A new robust system for 3D audio using loudspeakers. In *Proceedings of the 2000 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2000)*, volume 2, pages II781 –II784, 2000.

[190] R. M. Warren. *Auditory Perception: A New Analysis and Synthesis*. Cambridge University Press, New York, NY. USA, 1983.

[191] E. M. Wenzel, M. Arruda, and D. J. Kistler. Localization using nonindividualized head-related transfer functions. *Journal of the Acoustical Society of America*, 94(1):111–123, 1993.

[192] E. W. Wenzel, J. D. Miller, and J. S. Abel. A software-based system for interactive spatial sound synthesis. In *Proceedings of the 2000 International Conference on Auditory Display (ICAD 2000)*, pages 151–156, Atlanta, GA. USA, April 1-5 2000.

[193] E. W. Wenzel, J. D. Miller, and J. S. Abel. Sound lab: a real-time, software-based system for the study of spatial hearing. In *Proceedings of the 108th Convention of the Audio Engineering Society*, Preprint 5140, Paris, France, February 19-22, 2000.

[194] T. Whitted. An improved illumination model for shaded display. *Communications of the ACM*, 23(6):343–349, 1980.

[195] F. L. Wightman and D. J. Kistler. Headphone simulation of free-field listening. I: stimulus synthesis. *Journal of the Acoustical Society of America*, 85(2):858–867, 1989.

[196] F. L. Wightman and D. J. Kistler. Sound localization. In W. Yost, A. Popper, and R. Fay, editors, *Springer Handbook of Auditory Research: Human Psychophysics*, volume 3, pages 155–192. Springer-Verlag Inc., New York NY. USA, 1993.

[197] F. L. Wightman and D. J. Kistler. Factors affecting the relative salience of sound localization cues. In R. H. Gilkey and T. R. Anderson, editors, *Binaural and Spatial Hearing in Real and Virtual Environments*, chapter 1, pages 1–23. Lawrence Erlbaum Associates, Mahwah, NJ. USA, 1997.

[198] R. Wolfson and J. M. Pasachoff. *Physics with Modern Physics*. HarperCollins College Publishers, New York, NY. USA, 2 edition, 1995.

[199] Z. Xiangyang, C. Ke'an, and S. Jincai. On the accuracy of the ray-tracing algorithms based on various sound receiver models. *Applied Acoustics*, 64(4):433–441, 2003.

[200] L. N. Yang and B. M. Shield. Development of a ray tracing computer model for the prediction of the sound field in long enclosures. *Journal of Sound and Vibration*, 229(1):133–146, 2000.

[201] P. Zahorik. Assessing auditory distance perception using virtual acoustics. *Journal of the Acoustical Society of America*, 111(4):1832–1846, 2002.

[202] D. Zotkin, R. Duraiswami, and L. Davis. Rendering localized spatial audio in a virtual auditory space. *IEEE Transactions on Multimedia*, 6(4):553–564, 2004.