PERCEPTUALLY DRIVEN INTERACTIVE SOUND PROPAGATION FOR VIRTUAL ENVIRONMENTS

Atul Rungta

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Computer Science.

Chapel Hill
2018

Approved by:

Dinesh Manocha

Ming Lin

Roberta Klatzky

Marc Niethammer

Gary Bishop

ABSTRACT

Atul Rungta: Perceptually Driven Interactive Sound Propagation for Virtual Environments
(Under the direction of Dinesh Manocha)

Sound simulation and rendering can significantly augment a user‘s sense of presence in virtual environments. Many techniques for sound propagation have been proposed that predict the behavior of sound as it interacts with the environment and is received by the user. At a broad level, the propagation algorithms can be classified into reverberation filters, geometric methods, and wave-based methods. In practice, heuristic methods based on reverberation filters are simple to implement and have a low computational overhead, while wave-based algorithms are limited to static scenes and involve extensive precomputation. However, relatively little work has been done on the psychoacoustic characterization of different propagation algorithms, and evaluating the relationship between scientific accuracy and perceptual benefits.

In this dissertation, we present perceptual evaluations of sound propagation methods and their ability to model complex acoustic effects for virtual environments. Our results indicate that scientifically accurate methods for reverberation and diffraction do result in increased perceptual differentiation. Based on these evaluations, we present two novel hybrid sound propagation methods that combine the accuracy of wave-based methods with the speed of geometric methods for interactive sound propagation in dynamic scenes. Our first algorithm couples modal sound synthesis with geometric sound propagation using wave-based sound radiation to perform mode-aware sound propagation. We introduce diffraction kernels of rigid objects, which encapsulate the sound diffraction behaviors of individual objects in the free space and are then used to simulate plausible diffraction effects using an interactive path tracing algorithm. Finally, we present a novel perceptual driven metric that can be used to accelerate the computation of late reverberation to enable plausible simulation of reverberation with a low runtime overhead. We highlight the benefits of our novel propagation algorithms in different scenarios.

To Mummy & Papa

ACKNOWLEDGMENTS

First of all, I would like to thank my adviser Prof, Dinesh Manocha for giving me the opportunity to work with him and providing me with the necessary tools to pursue my ideas. His guidance and support were instrumental in me accomplishing my research goals.

I would also like to thank my committee member Prof. Roberta Klatzky for her exceptional guidance in helping me design my experiments. For someone of her stature to be so approachable and helpful was indeed a privilege. Further, I am grateful to my committee members Prof. Ming Lin, Prof. Marc Niethammer, and Prof. Gary Bishop for their feedback in making sure my dissertation meets the highest standards.

I would like to thank my co-author Ravish Mehra for being a mentor to me as I was starting out and introducing me to the field of sound simulation. Further, I am grateful to my other co-authors for their help and insight. Apart from that, these last five years would not have been an enjoyable experience had it not been for my amazing friends in the GAMMA group and the CS department; the constant companionship, deep technical (and sometimes personal) discussions, the late night fooseball, ping-pong, and heated, irrelevant discussions have made this an unforgettable experience, which I shall forever look back upon fondly.

Apart from my friends at UNC, my friends from Pune (India) have continued to play a huge role in my life and I owe my sanity and well-being over the last five years to them. Aaditya, Vaibhav, Anu, Neha, and Mansa are the reason I kept traveling to California, to blow-off steam after a strenuous paper submission. Here's to many more fun trips together.

My wife Vagisha is perhaps the most patient person I know. She was very understanding during her time here and managed really well on the limited resources we had at our disposable. This really helped me focus on my PhD and get it done as soon as was possible. I love you!

Finally, and most importantly, I would like thank my parents who are always there for me, no matter what. They have been incredibly supportive and loving and none of this would have been possible without them. I dedicate this dissertation to them.

Onward and upward!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

Sound is one of the most important senses available for conveying information about the world around us. It constitutes a fundamental mode of communication, entertainment, and education, among other things. Mathematically, sound is a pressure wave creating compression and rarefaction when traveling through a medium. These waves create changes in pressure that, upon reaching our ears, cause the eardrum to vibrate, which creates the sensation we call sound. Human hearing is only sensitive to a subrange of sound wave frequencies called the *human hearing range* $(20Hz - 20kHz)$.

Sound has many associated phenomena that have been studied both physically and psychophysically. Physically speaking, sound propagation constitutes the emanation of sound from a source, and transmission through an environment undergoing phenomena such as diffraction, reflections, scattering before reaching the listener. Once the sound waves reach the listener, they create the perception of sound and the study of these perceptual effects constitutes the psychophysical aspect of sound. The field of psychophysics that deals with studying the perception of sound is called psychoacoustics.

1.1 Motivation

In the last few years, virtual reality has seen a resurgence that has been bolstered by the availability of affordable, off-the-shelf head-mounted displays. Studies have shown that sound forms a critical component of virtual environments and that it has an augmentative effect on the user‘s sense of presence and immersion (Dubois et al., 2009; Larsson et al., 2002a; Hendrix and Barfield, 1996). In training simulations such as treatment of post-traumatic stress disorder (PTSD), studies have shown the importance of accurate sound cues (Rothbaum et al., 1999) to maintain training fidelity. Combat training simulations in VR also benefit from accurate sound simulation (Hughes et al., 2006). Video games are as popular as ever and, although much focus has been put on creating the most realistic visuals, they benefit greatly from sound. Accurate sound propagation can enhance the gamer’s sense of realism in the environment and, improve gameplay and immersion. Augmented reality (AR) is poised to become popular in the years to come, and along better visuals, better sound will enhance the user's sense of realism. Sound propagation can be used to generate realistic environmental sound effects such as diffraction and reverberation for virtual sound sources placed in a real environment. Teleconferencing systems can benefit from sound propagation by considering the geometry and materials of the rooms in which the callers are located, making the remote callers sound like they are in the same room. Big architectural projects such as building auditoriums, hospitals, and airports can benefit tremendously from realistic sound simulation on their CAD models to achieve the desired aural characteristics before the projects are implemented, saving a lot of time and money. Similarly, classrooms can benefit from sound propagation to gauge the intelligibility of speech in all parts of a classroom before it is constructed.

Given the importance of sound, much research has gone into modeling sound propagation and the associated effects to capture the interaction of sound with the environment as it emanates from a source, undergoing diffraction, reflection, and scattering. These techniques can be broadly categorized as: (a) parametric filter-based, (b) geometric, and (c) wave-based. Each of these techniques offers a trade-off in terms of scientific accuracy and computation cost/requirements. Parametric filters or heuristic techniques form the simplest and cheapest way to simulate sound propagation, especially reverberation, in an environment. These techniques use digital filters that must be tuned manually and this is normally done based on the sound designer's intuition. Although these filters are not physically accurate, their low-compute and memory overheads make them widely used in video games and virtual acoustic systems. Geometric techniques work under the simplifying assumption that sound travels in straight lines. This allows them to use methods such as image-sources, ray tracing, beam tracing, and frustum tracing to simulate the interaction of sound with the environment. These techniques are accurate for high sound frequencies and can accurately simulate phenomena such as reverberation. However, low-frequency phenomena such as diffraction cannot be easily modeled. These methods have relatively low-compute requirements leading to their increased use in computer games and virtual reality systems. The final category of sound propagation system includes wave-based or numerical techniques. These constitute the most accurate methods of sound propagation because they solve the underlying wave equation (acoustic wave equation) that governs the propagation of sound. These methods can accurately simulate all underlying phenomena associated with sound propagation but have extremely high compute and memory requirements largely precluding their use in interactive applications. Therefore, each of the sound propagation techniques offers a trade-off in terms of accuracy and computational cost.

While sound propagation numerically simulates acoustic phenomena, an equally important aspect of modeling sound is the perceptual study of these phenomena. Some of the earliest experiments in modern psychoacoustics were done by Hermann von Helmholtz (Von Helmholtz, 1912), who studied human hearing and the perception of music. Lord Rayleigh, in addition to his pioneering work on the physics of sound, did ground-breaking work in sound localization propounding his duplex theory (Rayleigh, 1907). With the advent of telephony, psychoacoustic measures concerning auditory thresholds, intensity discrimination, and frequency discrimination were established to reduce the bandwidth requirements of telephone lines (Fletcher and Galt, 1950; Fletcher, 1953). As sound made its way into virtual environments, a whole new realm of applied psychoacoustics was developed dealing with not only the perception of sound in these environments but also how the environments *themselves* affect the perception of sound. A typical example of this type of work is experiments on sound localization in virtual environments (Begault and Trejo, 2000) and the associated studies of the head-related transfer functions (HRTF) (Wenzel et al., 1993; Algazi et al., 2001; Kistler and Wightman, 1992) that are either numerically or experimentally synthesized. An interesting example of the virtual environment itself affecting a psychoacoustical phenomenon is that of auditory distance perception which is the ability to discern the distance to a sound source based on the sound. In real-life, auditory distance perception is compressive (Bronkhorst and Houtgast, 1999), but when coupled with a virtual environment, the compression characteristics change (Zahorik, 2002).

1.2 Psychoacoustic Evaluation of Interactive Sound Propagation Algorithms

As mentioned above, sound propagation has seen a plethora of research yielding methods that offer varying trade-offs between scientific accuracy and computational cost. Despite this, not much work has been done in establishing another crucial trade-off: *perceptual accuracy* and cost. Because of this gap, the psychoacoustical differentiation capabilities of sound propagation algorithms remain largely unstudied and the question whether increased scientific accuracy afford increased *perceptual differentiation* goes unanswered. Given the associated sound propagation phenomena, e.g., diffraction, reverberation, and the different techniques available to simulate them, one obvious question arises as to whether a more expensive but scientifically accurate method performs better perceptually compared to a cheaper but approximate method.

Evaluation of Approximate and Accurate Diffraction Simulation Diffraction is a very important, yet psychoacoustically understudied, phenomenon referring to the bending of sound around obstacles. Diffraction helps a listener hear sounds not in the line- of-sight and is observed in everyday life (Tsingos et al., 2001a). The acoustic wave equation explains the phenomenon; hence, wave-based methods simulate it automatically. Geometric methods, on the other hand, work on the simplifying assumption of rectilinear propagation of sound cannot account for it intrinsically, must incorporate diffraction externally. This has been done by incorporating methods such as the uniform theory of diffraction. (Kouyoumjian and Pathak, 1974). With the advent of better algorithms and faster hardware, realtime diffraction is now used for interactive geometric propagation (Schissler et al., 2014a; Tsingos et al., 2001a). Few psychoacoustical studies have been performed on diffraction. (Torres and Kleiner, 1998) conducted a study investigating the audibility of diffracted sound for a simple open geometry and found that diffracted sound is audible in non-shadow regions. (Torres et al., 2001a) performed listening tests to compute the audibility of edge diffraction in a stage house and found that first order diffraction is significantly more audible than second order diffraction. However, none of these studies considered interactive sound propagation algorithms or virtual environments and they were done in actual, physical environments. (Mehra et al., 2015) conducted experiments to evaluate the subject localization performance using wave-based and geometric propagation and found that diffraction helped the subjects localize faster. However, this study did not consider diffraction as a separate phenomenon from other acoustic effects, necessitating the comparative study of diffraction as computed by an accurate (wave-based) and an approximate (UTD) method.

Evaluation of Approximate and Accurate Reverberation Simulation Unlike diffraction, reverberation is a well-studied phenomenon in psychoacoustics, both in real life and in virtual reality. Reverberation has myriad perceptual effects on humans. (Zannini et al., 2011) found that sound localization performance decreased with increasing reverberation times. (Hartmann, 1983) established how the localization accuracy decreases in the same room if it is reflective compared to absorptive. (Rakerd and Hartmann, 1985) showed the detrimental effect of reverberation on localization, even at low frequencies. Other perceptual effects of reverberation include the reduction of speech clarity. (Knudsen, 1932) showed how reverberation can reduce the number of sounds that are heard correctly. (Galster, 2007) established the relationship between the speech transmission index (STI) and the reverberation conditions. This relationship shows the significant decrease in speech intelligibility in small, highly-reflective rooms. Despite negatively affecting localization and speech intelligibility, reverberation is known to a have positive impact on externalization, environment size estimation, and auditory distance perception. (Begault, 1992) conducted studies to show that spatial reverberation increases externalization. (Hameed et al., 2004) assessed reverberation parameters (RT_{60} and DRR) and their effect on room size estimation, finding that RT_{60} is the most important in this regard. (Cabrera et al., 2005) performed studies in real and virtual rooms to investigate the role of various parameters, while (Pop and Cabrera, 2005) performed experiments in three real rooms to find a negative relation between sound level and room size perception. Reverberation parameters such as direct-to-reverberant ratio (DRR) are known to increase auditory distance perception (Mershon and King, 1975). (Richards and Wiley, 1980) showed the effect of reverberation on auditory distance perception in outdoor environments. (Bronkhorst and Houtgast, 1999) formulated a computational model of auditory distance perception as a function of the DRR showing that increases in simulated reflections of sounds in virtual environments can result in increased auditory distance perception.

Given the strong effect reverberation has on human auditory perception, most virtual environments incorporate it either using approximate reverberation filters or a geometric acoustic system. Reverberation filters use a combination of nested all-pass and comb filters to produce a series of decaying echoes. These filters require setting reverberation parameters such as RT_{60} and DRR to model reverberation. Since RT_{60} requires either real-world measurements or accurate simulation to estimate, a well-known empirical formula called the Sabine's equation is used to estimate RT_{60}. Sabine's equation establishes the relationship between the volume, surface area, and absorptivity of the room and is given by:

$$RT_{60} \approx 0.1611 sm^{-1} \frac{V}{Sa}, \tag{1.1}$$

where V is the total volume of the room in m^3, S is total surface area in m^2, a is the average absorption coefficient of the room surfaces, and Sa is the total absorption in sabins. Geometric acoustic systems, on the other hand, simulate the effect of reverberation from first principles by performing high-order specular and diffuse reflections in the environment. However, no existing work looks at the perceptual comparison of accurate (geometric acoustics) and approximate (reverberation filter) methods. A comparative study is thus required to validate whether the more expensive, more accurate geometric acoustic methods offer better perceptual performance compared to the cheap, approximate digital filters.

1.3 Accurate Simulation of Sound Propagation Effects

Modeling sound propagation involves simulating the constituent phenomena such as scattering, diffraction, early reflections, and late reverberation. As mentioned above, research in simulating these phenomena has yielded three distinct categories of methods: heuristic or parametric filters, geometric methods, and wave-based methods.

Typically, parametric filters are used to simulate reverberation based on reverberation parameters of the environment; hence, these filters are also called artificial reverberators.These reverberators typically fall into three distinct categories: delay networks, convolutional, and room models. Out of the three, delay networks are by far the most commonly used because of their low computational complexity. The first artificial reverberator was introduced by Schroeder (Schroeder and Logan, 1961) and is an example of a delay network based reverberator. It uses digital nested allpass filters in combination with a parallel bank of comb filters to produce a series of decaying echoes. These filters require parameters such as reverberation time (RT_{60}) and direct-to-reverberant ratio (DRR) to tune the allpass and comb filters. An important improvement to Schroeder's filter was made by Moorer (Moorer, 1979) who added one-pole filters to Schroeder's filter to enable setting (RT_{60}) in a frequency-dependent way. Another important type of delay network filter is the Feedback Delay Network (FDN) introduced by Jot and Chaigne (Jot and Chaigne, 1991). Despite its age, the Schroeder filter remains the most widely used artificial reverberator due to its simplicity and low compute requirements.

Geometric methods, are high-frequency approximations that work under the simplifying assumption that sound travels in straight lines. The most commonly used geometric techniques are based on image-source methods (Allen and Berkley, 1979; Borish, 1984a) and ray-tracing (Krokstad et al., 1968a; Vorländer, 1989). In addition, beam tracing (Funkhouser et al., 1998a) and frustum tracing have also been used (Taylor et al., 2009b; Chandak et al., 2008). Geometric methods tend to be efficient and are thus used in interactive virtual environments to simulate specular and diffuse reflections. More recently, (Schissler et al., 2014a) use temporal coherence to significantly accelerate the computation of high-order diffuse reflections at interactive rates. Although these methods can simulate high-frequency acoustic phenomena accurately, their simplifying assumption breaks down at low frequencies where phenomena such as diffraction become common. In order to address this, methods such as the uniform theory of diffraction (UTD) (Kouyoumjian and Pathak, 1974) and Biot-Tolstoy-Medwin (BTM) (Biot and Tolstoy, 1957) have been incorporated into interactive geometric acoustic systems (Tsingos et al., 2001a; Antani et al., 2012b).

Wave-based methods solve the underlying mathematical formulation that governs sound phenomena (the acoustic wave-equation) numerically to generate accurate results. The acoustic wave-equation is a second order partial differentiation equation given in the time domain as:

$$\nabla^2 p - \frac{1}{c^2}\frac{\partial p^2}{\partial^2 t} = F(\mathbf{x}, t) \quad \mathbf{x} \in \Omega, \tag{1.2}$$

where ∇^2 is the Laplacian, p is pressure, $c = 343ms^{-1}$ is the speed of sound, $F(\mathbf{x}, t)$ is the forcing term corresponding to the source, and Ω is the domain of interest. Equivalently, it can be expressed in the frequency domain as the Helmholtz equation:

$$\nabla^2 p + \frac{\omega^2}{c^2} p = 0, \quad \mathbf{x} \in \Omega, \tag{1.3}$$

where $p = p(\mathbf{x}, \omega)$ is the complex-valued pressure field, ω is the angular frequency, c is the speed of sound in the medium, and ∇^2 is the Laplacian operator.

Methods solving the frequency domain Helmholtz equation are based on the finite element method (FEM) (Zienkiewicz et al., 2006) which discretizes the entire domain volume, and the boundary element method (BEM), which only discretizes the surface or boundary of the domain in question. (Cheng and Cheng, 2005; Liu and Nishimura, 2006; Gumerov and Duraiswami, 2009). Time-domain methods for solving the acoustic wave-equation include methods such as the finite-difference time domain (FDTD) (Yee, 1966; Taflove and Hagness, 2005), which uses finite-differences and iterates over the time steps. Stability requirements for these methods mandate time step sizes that make these methods scale linearly with the volume of the scene and, by the fourth-order with the increasing frequency of simulation. Other, more efficient, time-domain methods include adaptive rectangular decomposition (ARD) (Raghuvanshi et al., 2009) and the pseudo-spectral time domain (PSTD) (Liu, 1997). These methods can accurately model all acoustic phenomena but their high precomputation and runtime costs limit their application to interactive sound propagation in static virtual environments.

Considering the distinct advantages and disadvantages of the different techniques available for sound propagation modeling, we need methods that provide perceptual accuracy and computational efficiency for high-fidelity virtual environments.

Modal Source Propagation Research in sound simulation can be broadly categorized into two different groups: sound synthesis and sound propagation. While sound propagation deals with how sound interacts with the environment, sound synthesis deals with how a source itself produces sound based on its physical characteristics when provided with an external impulse, e.g., hitting a bell (O'Brien et al., 2002a; Raghuvanshi and Lin, 2006). While many standalone algorithms have been developed for each of these categories, these methods tend to be mutually exclusive. Rigid sound sources vibrate on being applied on impact, but the final sound we hear is these vibrations that propagate through the environment. Existing methods can model the propagation of these vibrations in *freespace* (James et al., 2006a), but no method can handle the propagation of the complex vibrational modes through an environment while accounting for the specific directivities of vibrational modes on the object. In other words, there is a need for a method that can *couple* the modally synthesized sound with sound propagation to give us mode-aware sound propagation.

Diffraction refers to the bending of sound waves around obstacles when the object dimensions are comparable to the wavelength. Few virtual acoustic systems incorporate diffraction because of the high complexity of simulating it. Geometric acoustic systems use methods such as the uniform theory of diffraction (UTD)(Kouyoumjian and Pathak, 1974) and Biot-Tolstoy-Medwin (BTM) (Biot and Tolstoy, 1957) to approximate diffraction. These methods have been incorporated into existing geometric acoustic systems (Tsingos et al., 2001a; Antani et al., 2012b), but these are approximate methods that only work for very simple objects. Further, the cost of these methods increases quadratically off of which sound must diffract. Wave-based techniques incorporate all the associated wave effects, including diffraction. As explained above, however, these methods have very high computation requirements, especially during the precomputation stage, and can only work in realtime virtual environments for scenes with limited number of sources and limited dynamism (Mehra et al., 2013; Raghuvanshi et al., 2009). Some hybrid methods (Yeh et al., 2013) combine the speed of geometric methods with the accuracy of numerical methods but suffer from some of the same drawbacks as the wave-based methods, i.e. large compute requirements for precomputation. Therefore, it is important to develop methods for simulating diffraction that have both a low precomputation and a low runtime overhead while having an accuracy comparable to wave-based methods.

Reverberation is the phenomenon of repeated reflections of sound in an enclosed environment. Reverberation is one of the most important phenomena in sound propagation and has a multitude of perceptual effects on hearing as mentioned in the section above. Given the importance of this phenomenon, most virtual acoustic systems incorporate reverberation to various degrees of accuracy. The first attempt at simulating reverberation was through the use of artificial reverberators (Schroeder and Logan, 1961), which are parametric filters that use combinations of all-pass and comb filters to create delayed and attenuated versions of the input signal to simulate repeated reflections. These methods work well when the parameters of the filters are tuned based on the actual physical parameters such as the RT_{60} and DRR, but estimating these parameters requires running simulations or doing actual acoustical measurements. Hence the parametric filters are normally tuned based on the sound designer's intuition of what the reverberation would sound like in that particular space making these methods inaccurate. However, artificial reverberators are still widely used in virtual acoustic systems such as game engines because of their simplicity and low computational overhead. Another way in which reverberation is simulated in virtual environments is through geometric acoustic systems. Since reverberation is repeated reflections of sound in an environment as the sound decays, geometric acoustic systems simulate this by repeating reflections of the geometric primitive (rays, frusta, beams) reducing the energy of the primitive with each bounce based on the material parameters of the surface. Typically, these systems can only compute a few orders of specular reflections for realtime environments (Chandak et al., 2008; Funkhouser et al., 2004; Vorländer, 1989), but recent developments in geometric acoustics have made it possible to get accurate reverberation using high-order specular and diffuse reflections(Schissler et al., 2014a; Schissler and Manocha, 2017) at interactive rates for multiple sources. Although this makes accurate reverberation possible at interactive rates, these methods need at least the computational power of a mid-range desktop machine. With virtual reality moving to low-powered devices (e.g., mobile), there is a need to simulate reverberation that is both plausible and low in its computational requirements.

1.4 Challenges

Although sound propagation research has made tremendous progress, the high-quality interactive simulation of acoustic effects in dynamic environments remains a challenging problem. First, no rigorous evaluations exist that compare the perceptual performance of various propagation algorithms in virtual environments. These evaluations are critical to establishing the relationship between the scientific accuracy of these algorithms and their perceptual accuracy. Second, although source representation and source directivity have been incorporated into existing propagation systems (Mehra et al., 2014a; Ren et al., 2012), none of these methods consider the propagation of individual modes of a vibrating rigid object in an interactive framework. (James et al., 2006a) introduced the idea of using multipole basis functions to compute the propagated sound of these individual modes, but this work only considers freespace propagation. While the idea of multipole basis functions can be extended to environmental sound propagation within the framework of a wave-based system, the high frequencies often generated by small rigid-bodies would be well beyond the capabilities of the current generation of wave-based techniques. Current geometric methods offer the best way to couple modal sounds with a propagation engine for all mode frequencies by weighting the rays using the multipole basis functions. This presents a significant challenge in terms of maintaining interactivity because tens of thousands of rays would be needed to get artifact free sound, and each one of these rays would have to be weighted using these complex and expensive multipole basis functions.

Diffraction has long been challenging to simulate accurately, especially for geometric systems. The state-of-the-art geometric system (Schissler et al., 2014a) incorporates the uniform theory of diffraction method and can handle high-orders of diffraction at interactive rates. However, it can only do so for objects with few edges or simplified versions of complex objects. Wave-based methods, as mentioned, handle diffraction for any object shape, but their exorbitant precomputation costs make them impractical for most applications. Further, these methods often only work in static scenes and can handle very few sources interactively. The state-of-the-art hybrid method (Yeh et al., 2013) offers the speed of geometric methods with the accuracy of wave-based methods, but it also suffers from very high precomputation costs that can reach on the order of weeks on desktop machines. This necessitate the use of high-performance compute clusters, which makes these methods impractical. We thus need a method that can significantly reduce the precomputation time for computing diffraction without sacrificing accuracy and that can work interactively at runtime.

Finally, accurate reverberation modeling has come a long way with current geometric methods such as (Schissler et al., 2014a), becoming capable of computing very high orders of specular and diffuse reflections by exploiting temporal and spatial coherence. These geometric methods can run at interactive rates on a desktop machine for a large number of sources. Parametric digital filters, although not physically-based, can be tuned to sound very similar to the actual reverberant field (provided we know the values of parameters such as reverberation time RT_{60} and direct-to-reverberant ratio DRR) and have very low compute requirements. As virtual reality moves to low-powered devices, ray-tracing based solutions are not viable as they are too expensive to compute. Parametric filters offer a cheap solution to the problem provided accurate reverberation parameters are provided. One simple solution to this problem would be to use ray tracing at a number of sample points and precompute the reverberation parameters for each of those points. Depending on the size of the scene, however, thousands of such sample points might be needed for precomputation and even with a fast sound propagation system, computing the full reverberation characteristics of all those points might take too long for it to be practical. We thus need a way of computing reverberation on low powered devices that avoids expensive ray tracing at runtime and *precomputation*, but that provides accurate reverberation parameters for tuning the digital reverberation filters.

1.5 Thesis Statement

The perceptual differentiation of acoustic effects such as modal source propagation, diffraction, and reverberation is enhanced by accurate simulation, and can be modeled efficiently in immersive virtual environments.

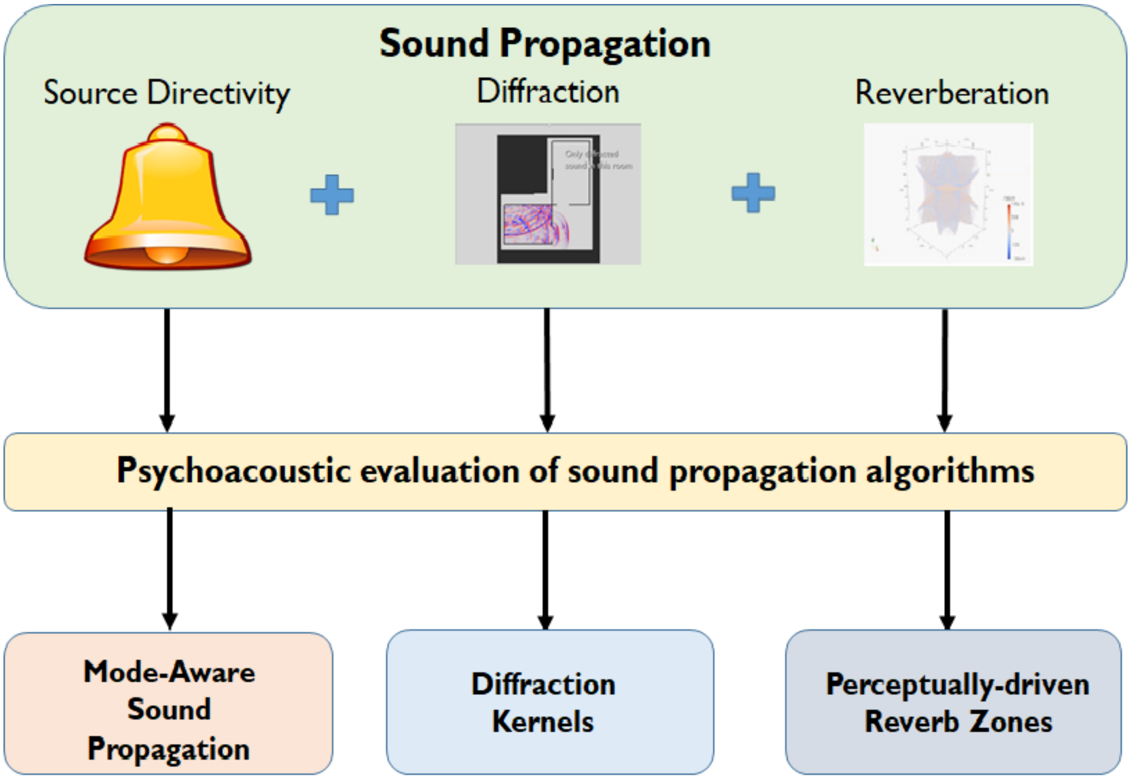
Sound Propagation
Source Directivity
Diffraction
Reverberation
Psychoacoustic evaluation of sound propagation algorithms
Mode-Aware Sound Propagation
Diffraction Kernels
Perceptually-driven Reverb Zones

Figure 1.1: **Overview of our work:** The figure shows the various components of our work. We consider sound propagation as a combination of three distinct phenomena: Source directivity, diffraction, and reverberation. Then, user evaluations tell us if numerical accuracy matters perceptually. These results are then utilized to design more perceptually accurate and efficient algorithms

In this dissertation, we propose a set of experiments that evaluate the perceptual accuracy of various existing algorithms that simulate phenomena such as diffraction and reverberation. Based on the results of these experiments, we propose a set of efficient techniques to model challenging acoustic phenomena with a focus on interactive virtual environments. All the methods developed have runtime computation on the order of a few milliseconds and have been integrated with commercial game engines to highlight their efficiency in a variety of challenging scenarios. Figure 1.1 gives a high-level overview of the thesis' flow.

1.6 Main Results

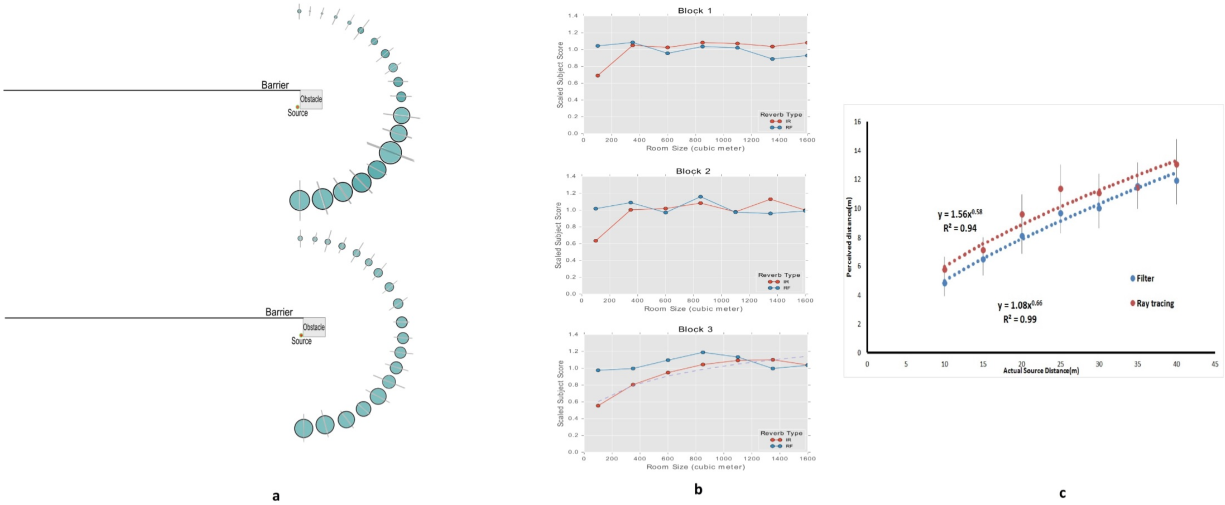

Figure 1.2: **Overview of the results of our user studies:**. a) The loudness perception for a sound source behind an obstacle and multiple listener positions around it in an open environment. The figure on top is the result for geometric diffraction (UTD), while the one below is for wave-based diffraction. b) The results of the room size perception experiment using two different reverberation algorithms. c) The results of the auditory distance perception experiment using two different reverberation algorithms.

1.6.1 Psychoacoustic Evaluation of Interactive Sound Propagation Algorithms

We conducted three different user studies comparing listeners perceptual responses to both accurate and approximate propagation algorithms simulating two key acoustic effects: diffraction and reverberation. Our first set of experiments evaluated diffraction computed using two methods: an accurate wave-based method and an approximate uniform theory of diffraction (UTD) method. The second and third sets of experiments evaluated reverberation computed using an accurate ray tracing based method and an approximate filter-based method. The first of these experiments evaluated the effect of the reverberation method on the perception of room size while the second experiment evaluated the effect of the two reverberation conditions on acoustic distance perception. In all three cases we saw a positive effect of the more physically-accurate algorithm over the approximate algorithm. (Figure 1.2)

1.6.2 Mode-aware Sound Propagation

Sound simulation research has focused on either sound synthesis or sound propagation, and many standalone algorithms have been developed for each domain. Although the current sound propagation methods can handle myriad sound sources and model their directional characteristics, none of these methods account for the modal characteristics of rigid sound sources that vibrate on impact (e.g., hitting a bell). The final sound we hear in such cases is the vibrations that propagate through the environment. In other words, the sounds we hear from these objects are coupled, i.e. the sound is synthesized when a source has an impulse applied to it and propagates through the environment in a manner defined by the vibrational characteristics of the sound (or their modes). In this work, we present an integration approach for coupling modal sound synthesis for rigid bodies with an interactive sound propagation system. The main contributions of this work are as follows:

1. First method that enables per-mode coupling of synthesis and propagation through single-point multipole expansion.

2. Interactive mode-adaptive propagation technique based on perceptually-driven Hankel function approximation.

3. High degree of dynamism to model dynamic surface vibrations, sound radiation, and propagation for moving sources and listeners.

Figure 1.3 highlights our approach. We first perform modal analysis on the object to compute its vibrational modes and their respective frequencies. The modes and the frequencies are used to compute the directivity patterns of the object using a frequency-domain, acoustic wave equation solver (BEM) which are then stored in compact basis functions. At runtime, these basis functions are sampled by weighting each ray depending on its angle and the distance traveled to give a mode-adaptive propagation algorithm. Given that a scene needs a large number of rays ($\sim 10^4$) to avoid sampling artifacts, computing the basis function for each

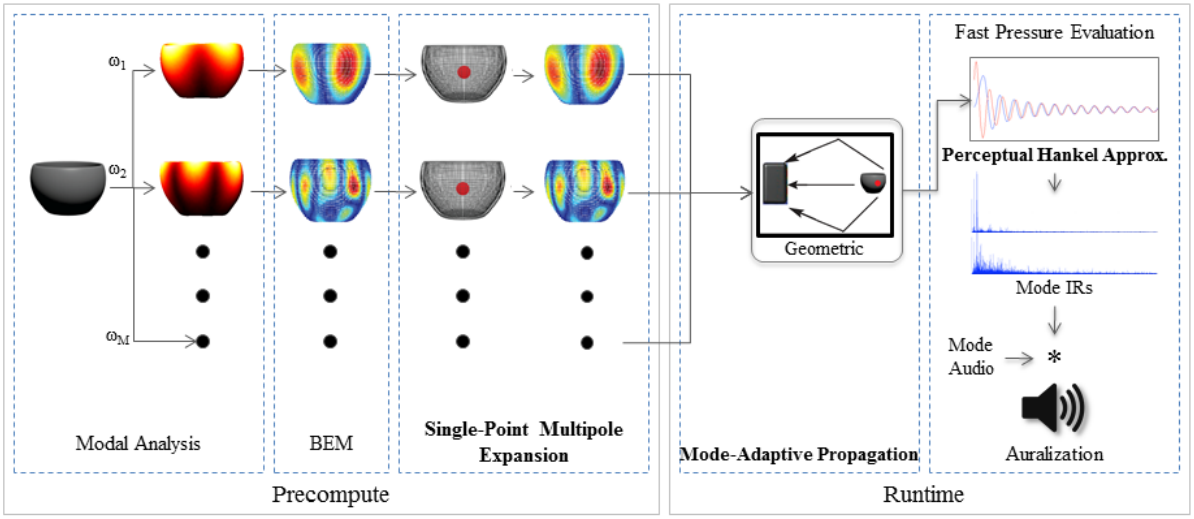

Figure 1.3: **Overview of our coupled synthesis and propagation pipeline for interactive virtual environments:**. The first stage of precomputation comprises the modal analysis. The figures in red show the first two sounding modes of the bowl. We then calculate the radiating pressure field for each of the modes using BEM, place a single multipole at the center of the object, and approximate the BEM evaluated pressure. In the runtime part of the pipeline, we use the multipole to couple with an interactive propagation system and generate the final sound at the listener. We present a new perceptual Hankel approximation algorithm to enable interactive performance.

one of them is too slow for interactive runtime. The main bottleneck is the large number of Hankel function evaluations (our basis function is a combination of spherical harmonics and Hankel functions). To keep our system interactive, we present a novel approximation scheme to significantly reduce the cost of evaluating the Hankel function when the listener is far enough away from the source. We use the threshold of hearing to compute the distance at which switching to an approximate version of the Hankel function would produce no noticeable difference. This scheme gives a 3x - 7x speedup in evaluating the basis function making our system interactive even in complex environments. We also conducted a preliminary, online user-study to evaluate whether our Hankel function approximation causes any perceptible loss of audio quality. The results indicate that the subjects were unable to distinguish between the audio rendered using the approximate function and the audio rendered using the full Hankel function in our benchmarks. The overall approach allows for high degrees of dynamism - it can support dynamic sources, dynamic listeners, and dynamic directivity simultaneously. We have integrated our system with the Unity game engine and demonstrate the effectiveness of this fully-automatic technique for audio content creation in complex indoor and outdoor scenes.

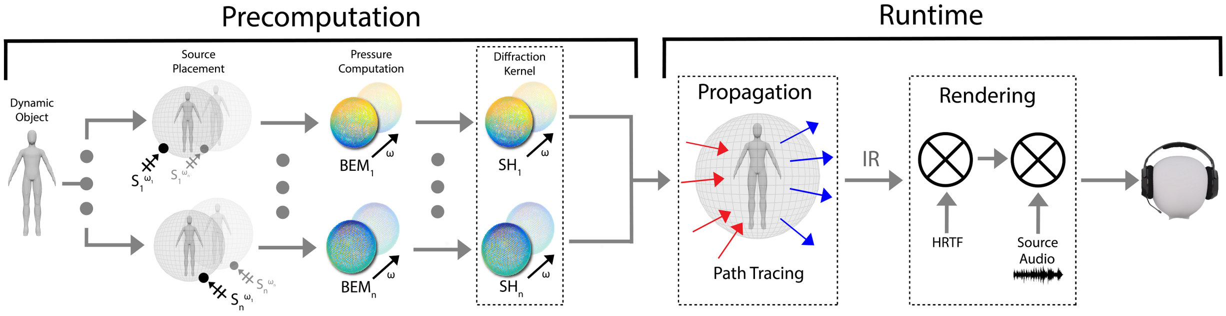

Figure 1.4: **Interactive Sound Propagation and Rendering:** We highlight different stages of our novel sound propagation and rendering pipeline, which uses per-object diffraction kernels. In the precomputation stage, we adaptively perform BEM simulations for certain directions (computed using our novel source placement algorithm) and measure the outgoing pressure fields produced by the scattering of plane waves at various frequencies. These pressure fields encode the scattering as a function of frequency and the input and output directions. These fields are then converted into an efficient spherical harmonic representation called the diffraction kernel. At runtime, the diffraction kernel is coupled with an interactive path tracing algorithm to simulate sound propagation and auralization in dynamic scenes.

1.6.3 Diffraction Kernels for Interactive Sound Propagation in Dynamic Environments

In this work, we introduce the notion of diffraction kernels to incorporate plausible diffraction in an interactive sound propagation algorithm. Diffraction kernels encapsulate the diffraction sound interaction behaviors of individual objects in the free field by representing them in a compact spherical harmonic basis function. We precompute the diffraction characteristic using an accurate wave-based method. Since diffraction is a direction-dependent phenomenon, a significant amount of precomputation is required to capture the effect from all possible incident angles. To remedy this, we introduce a novel source placement algorithm that significantly reduces the precomputation time required to compute the complete diffracted field of an object. Our source placement algorithm clusters the incident source positions that are likely to produce similar or symmetric diffracted fields and computes the diffracted field for only one source per cluster, thus giving us a substantial speedup in the precomputation stage. The main contributions of this work are as follows:

- Adaptive source-placement algorithm that significantly reduces the precomputation time by comparing the view-dependent shape signatures of an object.

- Handling of highly tessellated or smooth objects while modeling diffraction and occlusion effects.

- General approach that can be integrated into existing geometric sound propagation methods.

Figure 1.4 shows the approach. We consider a densely sampled sphere around the object representing all possible incident angles for diffraction computation. Using our source placement algorithm we compute the incident angles that are expected to produce similar diffracted fields and cluster them together. An accurate wave-based solver (BEM) is run for one representative point per cluster and the diffracted fields for the other points in the cluster are interpolated from the representative point. The computed diffracted field is then represented in a spherical harmonic basis that we call diffraction kernels. We have performed extensive evaluation of the method and find that our approach results in a significant speedup (up to 130x) in the precomputation time and introduces a mean absolute error (MAE) of $<$ 2dB even for complex, highly-tessellated objects. The diffraction kernels are easily integrable with interactive sound propagation algorithms to produce plausible diffraction effects for highly complex objects.

1.6.4 Perceptual Characterization of Early and Late Reflections for Auditory Displays

Even though geometric sound propagation is interactive on PC-based systems, its compute and memory requirements are still too high for devices with limited resources such as mobile-based VR devices such as Oculus Go and Samsung GearVR. This becomes particularly evident in reverberant environments where the decay of sound energy is slow requiring multiple reflections to compute the reverberation. As an alternative, parametric reverb filters can approximate reverberation efficiently, but must be tuned by hand to match the reverberation characteristics of a given environment. This makes these filters a feasible option for resource-constrained platforms, but tuning of the filters poses a significant challenge since reverberation characteristics often vary within an environment requiring one to establish multiple reverb zones with different reverberation parameters. A viable approach to the problem could be the dense sampling of the scene at a multitude of points and using a high-fidelity sound propagation system to estimate the reverberation parameters (e.g., RT_{60}) at each of these sample points and then using these RT_{60} values to Since computing late reverberation is a computationally expensive process, such naive preprocessing could take a long time.

We approach this problem by observing that while late reverberation is expensive to compute, early reflection computation is relatively cheap, typically requiring an order of magnitude fewer rays than late reverberation. However, early reflections do not directly estimate reverberation parameters such as RT_{60}. We introduce a novel, perceptually derived metric $P - Reverb$ that relates the just-noticeable difference (JND) of the early reflections to the late reverberation in terms of the mean-free path (MFP) of the environment. We conduct two extensive user evaluations that establish the JND of the early reflections and late reverberation in terms of the mean-free path of the environment. We then relate the two JNDs in terms of the mean-free path giving us the $P - Reverb$ metric. We demonstrate the use of the metric in speeding up the precomputation of late reverberation parameters (e.g., RT_{60}). The main contributions of this work are as follows:

- A novel, perceptually derived metric relating early reflections to the late reverberation ($P - Reverb$)
- Two extensive user-studies establishing the JND of early reflections and JND of late reverberation in terms of mean-free path of the environment.
- Significant speed-up in the precomputation of RT_{60} using $P - Reverb$ metric.

1.7 Organization

The rest of this dissertation is organized as follows:

Chapter 2 presents various user studies evaluating the perceptual differentiation capabilities of current sound propagation methods by comparing their efficacy in modeling acoustic phenomena showcasing the advantages of using physically-accurate algorithms.

Chapter 3 presents an efficient method to couple modal sound synthesis with sound propagation using an efficient, perceptually-driven approximation of the Hankel function.

Chapter 4 presents an adaptive source placement algorithm for reducing the computation time of acoustic scattering problems. We also highlight the performance and effectiveness of this approach by coupling our algorithm to a geometric sound propagation system to compute diffraction effects from complex objects in real-time.

Chapter 5 presents two extensive user evaluations to derive a novel metric that relates the just-noticeable difference (JND) of early reflections to the JND of late reverberation in terms of the mean-free path of the environment. We highlight the accuracy of the metric and its ability to significantly accelerate the precomputation of late reverberation parameter (RT_{60}).

Chapter 6 concludes the dissertation and includes a discussion on the limitations and future challenges in the area of simulating sound propagation effects.

CHAPTER 2: Psychoacoustic Evaluation of Interactive Sound Propagation

2.1 Introduction

Sound plays a vital role in increasing the degree of realism in virtual environment (VE) systems (Begault et al., 1994; Larsson et al., 2002a) and other interactive applications such as video games. This observation has motivated the development of different sound propagation methods that are used to simulate how sound waves, emitted from a source, travel through an environment and interact with the objects before reaching a listener. These methods are used to model well-known acoustic phenomena such as diffraction, reverberation (comprising early reflections and late reverberation), and scattering. At a broad level, sound propagation methods are categorized into geometric (Krokstad et al., 1968a; Borish, 1984b; Allen and Berkley, 1979) and wave-based (Zienkiewicz, 2005; Yee, 1966; Cheng and Cheng, 2005) algorithms. While these computational techniques have been studied for many decades in different fields, only recent advancements in terms of new algorithms and fast hardware have enabled the development of interactive propagation systems that are useful for VE. These include interactive geometric methods based on ray tracing and beam tracing (Schissler et al., 2014a; Tsingos et al., 2001a; Taylor et al., 2009b; Funkhouser et al., 2004) that can simulate approximate diffraction, early reflections, and high-order late reverberation. Furthermore, advancements in scientific solvers (Mehra et al., 2013; Raghuvanshi et al., 2009; Webb and Gray, 2013; Mehra et al., 2015) have made it possible to compute highly accurate solutions to the wave equation for large domains, and thereby perform interactive sound propagation.

Given the recent developments in interactive sound propagation algorithms, it is imperative to evaluate their perceptual effectiveness. Psychoacoustics researchers have focused on evaluating the perceptual effects of many of these acoustic phenomena and many important results have been published on how different propagation phenomena affect our perception of the environment (Fastl and Zwicker, 2007). Most of these studies were conducted in either real-world environments or in very simple virtual environments that could only simulate limited acoustic effects. The advent of interactive and accurate sound propagation techniques makes the task of perceptually evaluating these phenomena simpler and less expensive.

In this chapter, we mainly focus on two of the aforementioned phenomena: diffraction and late reverberation (hereafter referred to as reverberation). Early reflections were not considered here, as they are easy to simulate and have been widely studied in the literature (Haas, 1951; Djelani and Blauert, 2001). Reverberation is also a well-studied phenomenon in psychoacoustics and enhances immersion (Kuttruff, 2007). Despite this, one of its fundamental effects, namely, to convey the size of the environment remains relatively unexamined. Reverberation is typically approximated using artificial filters, and such filters are widely used in computer games and VE (Jot and Chaigne, 1991).

Therefore, we evaluate the reverberation computed in a physically accurate manner using an interactive, state-of-the-art geometric propagation system to a pre-computed (Schroeder-type) filter and evaluate their relative effectiveness in telling us the size of an environment. Perceptual effects of diffraction, although important (Torres et al., 2001b), are seldom evaluated in virtual environments primarily due to the complexity of modeling diffraction. Most interactive geometric sound propagation systems approximate edge diffraction based on Uniform Theory of Diffraction (UTD) (Tsingos et al., 2001a). Therefore, we evaluate the perceptual performance of diffraction effects computed using UTD with a numerically accurate solver that directly solves the wave equation. We report two separate comparative studies to evaluate whether *increased numerical accuracy of sound propagation translates to perceptual differentiation*.

In our diffraction study, we construct a virtual test scene similar to (Kawai, 1981) and perform a psychoacoustical evaluation. We evaluate the diffracted sound field around an obstacle by placing subjects along a semi-circle for the two methods: UTD and wave-based. The subjects are asked to rate the perceived loudness for different positions along the semi-circle. Our results show that wave-based diffraction results in a diffracted field that decays nearly linearly with an increasing diffraction angle, as compared to the UTD-based diffraction, which shows erratic behavior.

Prior psychoacoustics studies in reverberation normally focus on evaluating how the human response to the environment varies with the changing reverberation parameters (e.g., RT_{60}). Our study builds upon these evaluations and seeks to compare the effectiveness of two competing methods to model reverberation: statistical filters and physically-accurate path tracing. We use free-magnitude estimation to compute the magnitude of the internal responsiveness of the subjects to the change in the physical dimensions of the environment and thereby, its reverberation characteristics. Our results show poor size discrimination for both the methods but show increasing discrimination ability for the physically accurate one as the subjects become more familiar with the task.

The rest of the chapter is organized as follows. In Section 2 we give a brief overview of the two acoustic effects and the corresponding techniques and algorithms used in the study to model them and their evaluation. Sections 3 and 4 describe the two studies, including their designs and procedures. Section 5 describes our evaluation metrics. In Sections 6 and 7 we discuss our results and their applications.

2.2 Background

In this section, we give a brief overview of the two sound phenomena, the methods used to model them, and the related work in psychoacoustic evaluation of these phenomena.

2.2.1 Diffraction

Diffraction refers to the bending of a wave around an obstacle when the obstacle's dimensions are comparable to the wavelength of the wave. Diffraction helps a listener hear sounds not in the line-of-sight and is observed in everyday life (Tsingos et al., 2001a). Diffraction is explained through the wave equation; hence wave-based methods emulate it automatically. Geometric methods, on the other hand, assume rectilinear propagation of sound. As these methods do not account for the bending of sound rays around an obstacle, this effect must be incorporated separately. Doing so is difficult and computationally expensive, which is why most virtual environments avoid incorporating diffraction even though the theories for approximating its effects have existed for decades (Biot and Tolstoy, 1957; Kouyoumjian and Pathak, 1974). With the advent of better algorithms and fast hardware, real-time diffraction is now used for interactive geometric propagation (Tsingos et al., 2001a; Schissler et al., 2014a).

2.2.2 Reverberation

Reverberation forms the later part of the impulse response(IR) within closed environments. It is caused by successive reflections or 'echoes' as they diminish in intensity.

Reverberation forms a critical part of the acoustics of an environment and directly correlates with the size and the clarity of sound in the environment. For these reasons, reverberation plays a very important role in architectural acoustics, especially while constructing auditoriums and concert-halls. This has led to considerable research in characterizing reverberation, and a number of parameters such as the reverberation time (RT_{60}) and clarity index (C_{50} and C_{80}) have been formulated (Kuttruff, 2007). We concern ourselves with RT_{60} in this study. RT_{60} is defined as the time it takes for the sound to decay by 60dB.

Given the importance of reverberation to architectural/room acoustics, and complexity of simulating it from physical principles, empirical methodologies have been developed to artificially simulate reverberation in virtual acoustics such as digital and convolution reverberators. These reverberators are parametrized using a number of values, the most important of which is the RT_{60}. We use a digital Schroeder filter for our evaluation as these filters are the most common form of digital reverberators in use today. A well-known empirical formula used to estimate RT_{60} is Sabine's equation, which gives the relationship between the RT_{60} of a room, its volume, and the total absorption by:

$$RT_{60} \approx 0.1611sm^{-1}\frac{V}{Sa}, \tag{2.1}$$

where V is the total volume of the room in m^3, S is total surface area in m^2, a is the average absorption coefficient of the room surfaces, and Sa is the total absorption in sabins. Another, more accurate, way to estimate the RT_{60} of a room is to look at the room's impulse response (RIR) and directly compute the time it took for the sound to decay by 60 dB as specified in ISO 3382-1:2009. This method involves a reverse cumulative trapezoidal integration to estimate the decay of the impulse response and using a linear least-squares fit to estimate the slope between 0 dB and -60 dB. This is the method we used to compute the RT_{60} for our digital reverberator.

2.2.3 Geometric Acoustics

Geometric acoustics methods work under the assumption that the wavelength of sound is much smaller (i.e., higher frequency) than the objects in the scene. This assumption allows these methods to assume that sound waves travel in straight lines and thereby use ray tracing (Vorländer, 1989; Krokstad et al., 1968a) and its variants such as beam tracing (Funkhouser et al., 2004). Other methods such as image sources (Allen and Berkley, 1979; Borish, 1984b) have also been employed. In order to approximate diffraction, geometric methods use formulations based on uniform theory of diffraction (UTD) or Biot-Tolstoy-Medwin (BTM). We focus on UTD-based diffraction, as that has been used for interactive applications.

2.2.3.1 Uniform Theory of Diffraction

The Uniform theory of diffraction (Kouyoumjian and Pathak, 1974) is a high frequency approximation of the phenomenon of diffraction. These methods were initially developed for the propagation of light, but later used for sound (Tsingos et al., 2001a). UTD assumes that a diffracting edge is of infinite length and acts as a secondary sound source. Another assumption made by UTD is that the source and listener are far away as compared to the wavelength of sound. According to UTD, an incoming sound ray hitting an infinite wedge results in a cone of diffracted rays, and a single ray with the shortest distance to the listener (in a homogeneous medium) forms the diffracted field.

2.2.4 Numerical or Wave-based Acoustics

Sound propagation is governed by the acoustic wave equation (in time domain):

$$\nabla^2 p - \frac{1}{c^2}\frac{\partial p^2}{\partial t^2} = F(\mathbf{x}, t) \quad \mathbf{x} \in \Omega, \tag{2.2}$$

where ∇^2 is the Laplacian, p is pressure, $c = 343ms^{-1}$ is the speed of sound, $F(\mathbf{x}, t)$ is forcing term corresponding to the source, and Ω is the domain of interest.

Solving Eq. 2.2 gives us the sound pressure P at any point in the domain. Unfortunately, closed-form solutions to the wave equation only exist for the simplest of domains, and most solvers use numerical techniques. However, the complexity of these methods increases as a fourth power of the frequency.

2.2.4.1 Adaptive Rectangular Decomposition

This technique was developed by (Raghuvanshi et al., 2009) and constitutes a domain-decomposition technique to solve the wave equation (Eq. 2.2) in homogeneous media. The underlying principle of this technique is based on the observation that the wave-equation can be solved analytically inside a rectangular domain. Therefore, this technique first decomposes the domain (scene) into a set of connected rectangles and computes the pressure in each of those rectangles analytically. The pressure is then transferred between rectangles using a finite-difference stencil. For more details on the technique, refer to (Raghuvanshi et al., 2009).

2.2.5 Related Work

Reverberation and its effects have been widely studied in psychoacoustics. (Zannini et al., 2011) explored the effect of source localization in reverberant conditions and found that the localization deteriorates with changing reverberation times. (Hartmann, 1983) found that localization accuracy decreases in a reflecting room compared to the same absorbing room. (Rakerd and Hartmann, 1985) showed that reverberation has a considerable effect on localization even at low frequencies. (Giguère and Abel, 1993) found localization was poorer in reverberant environments as compared to absorbing ones. Other studies involving reverberation have analyzed the impact of reverberation on speech clarity. (Knudsen, 1932) showed that the presence of reverberation reduces the number of sounds that are heard correctly. (Galster, 2007) found the relation between the Speech transmission index (STI) as a function of signal-to-noise ratio and the reverberant conditions and found speech intelligibility decreases significantly in a small, highly reflective rooms. Although reverberation decreases localization accuracy, it is known to have positive effects with respect to externalization. Externalization is defined as the perception of the sound source emanating from a point in the world in contrast to internalization where the sound appears to be emanating from within one's own head. (Begault, 1992) conducted studies that showed that spatial reverberation increased externalization. Another important perceptual effect of reverberation is that of size estimation. (Hameed et al., 2004) conducted studies to assess the effect of two reverberation parameters: the reverberation time RT_{60} and the direct-to-reverberant energy ratio (D/R) ratio and found that RT_{60} is the most important parameter in room size estimation. (Cabrera et al., 2005) performed experiments using real and virtual rooms and investigated the role of various parameters including the reverberation time. They found reverberation strongly affects the room size perception. (Pop and Cabrera, 2005) tested the relationship between acoustical characteristics of a room and its perceived size for three real rooms. They found that a negative relation between sound level and room size.

Most studies on diffraction have been numerical in nature (Kawai, 1981; Chu et al., 2007). Few psychoacoustical studies have been performed on diffraction. (Torres et al., 2001b) conducted a study investigating the audibility of diffracted sound for a simple open geometry and found that diffracted sound is audible in non-shadow regions. (Torres et al., 2001a) concluded that reflected-diffracted combinations are significant and audible even for minute spectral changes. (Torres and Kleiner, 1998) perform listening tests to compute the audibility of edge diffraction in a stage house and found that first order diffraction is significantly more audible than second order diffraction. (Mehra et al., 2015) conducted experiments to evaluate the subject localization performance using wave-based and geometric propagation and found that diffraction played a part in helping the subjects localize faster. However, this study did not consider diffraction as a separate phenomena from other acoustic effects.

2.3 Experiment 1: Diffraction Study

2.3.1 Participants

Sixteen subjects participated in this study with informed consent. The ages ranged from 22 to 28 (Mean = 24.6 with SD = 1.4, 2 females and 14 males). The participants were recruited from the staff and students at a university campus. All participants reported normal hearing.

2.3.2 Apparatus

The setup consisted of a Dell T7600 workstation with the sound delivered through a pair of Beyerdynamic DT990 PRO headphones. The subjects wore a blindfold. The software consisted of in-house code to auralize precomputed IRs. All code was written in C++.

2.3.3 Stimuli

The source consisted of a pre-recorded sound of a ringing bell. Since diffraction is a frequency-dependent phenomenon and more prominent at low frequencies, the bell clip was low-pass filtered with a cut-off frequency of 300 Hz. The choice of frequency range was motivated by:

- The edge-diffraction phenomena is most prominent in the 20 Hz - 350 Hz range. (Torres et al., 2001b)
- The loudness characteristic of sound is least complex (near linear) with respect to frequency in the frequency range chosen by us. Fig. 2.2

The subjects were placed at a radius of 5m from the sound source along a semi-circle covering two orders of diffraction as shown in Fig 4.9. The scene was an open scene making sure no reflected sound reached the subjects. The semi-circle was sampled at 10 ° giving 18 positions plus one for the direct sound region for a total of 19 positions. The impulse response for each of these 19 positions and the two methods of diffraction were pre-computed and stored on file. The stimuli consisted of 38 combinations (19 positions x 2 diffraction types) of convolved source and IR signals. Each trial consisted of the subject being placed at one these 19

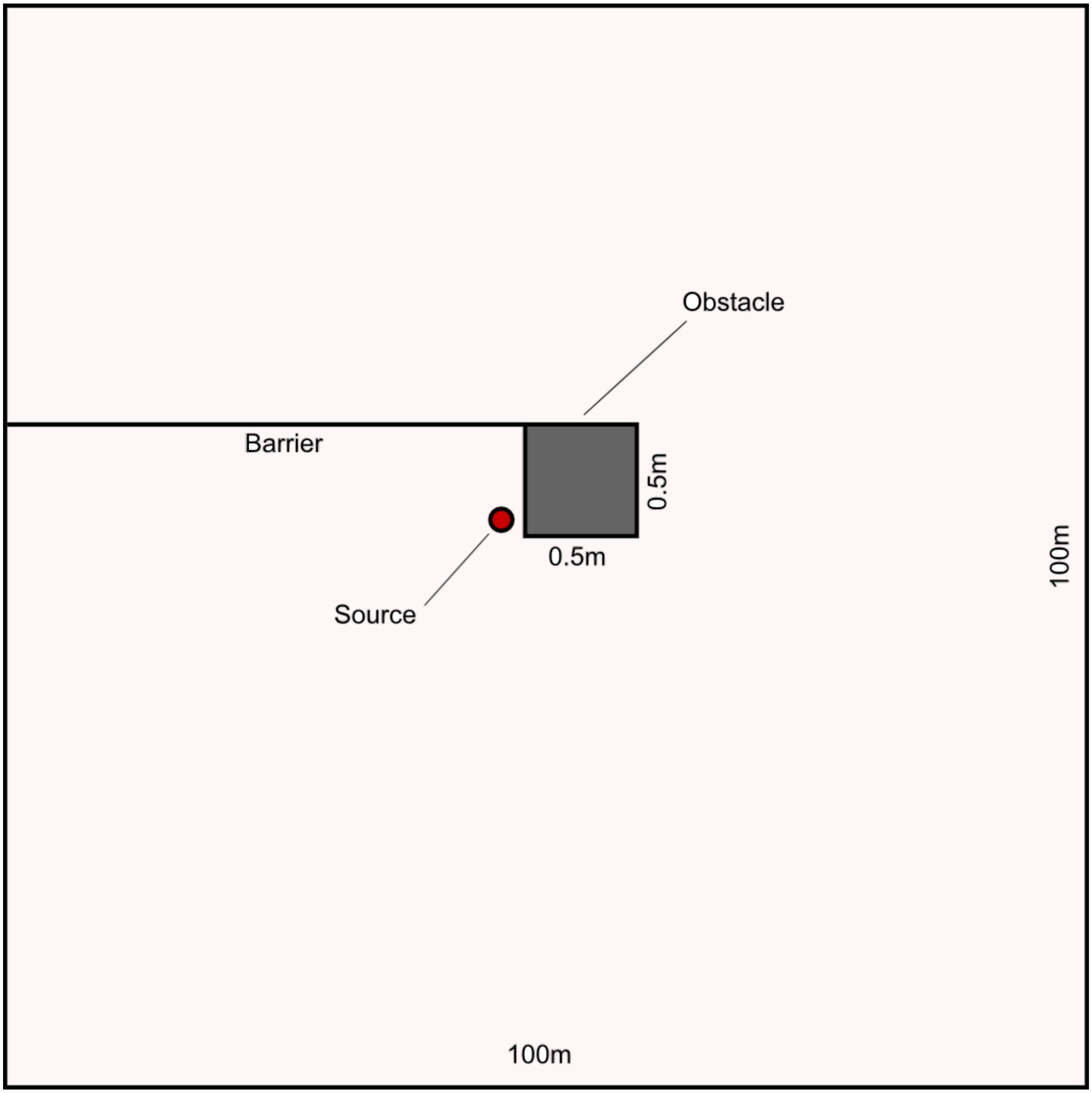

Figure 2.1: The figure shows the setting used for characterizing diffraction. The obstacle was 0.5m x 0.5m. The solid barrier prevents the sound from reaching the listener from behind the obstacle. The barrier and the obstacle are fully absorptive to prevent any reflections and making sure only the diffracted sound reached the listener. The figure is not drawn to scale for better viewing of the experimental setup.

positions randomly with the audio either auralized using the UTD-based or the wave-based diffraction, the order of which was also randomized.

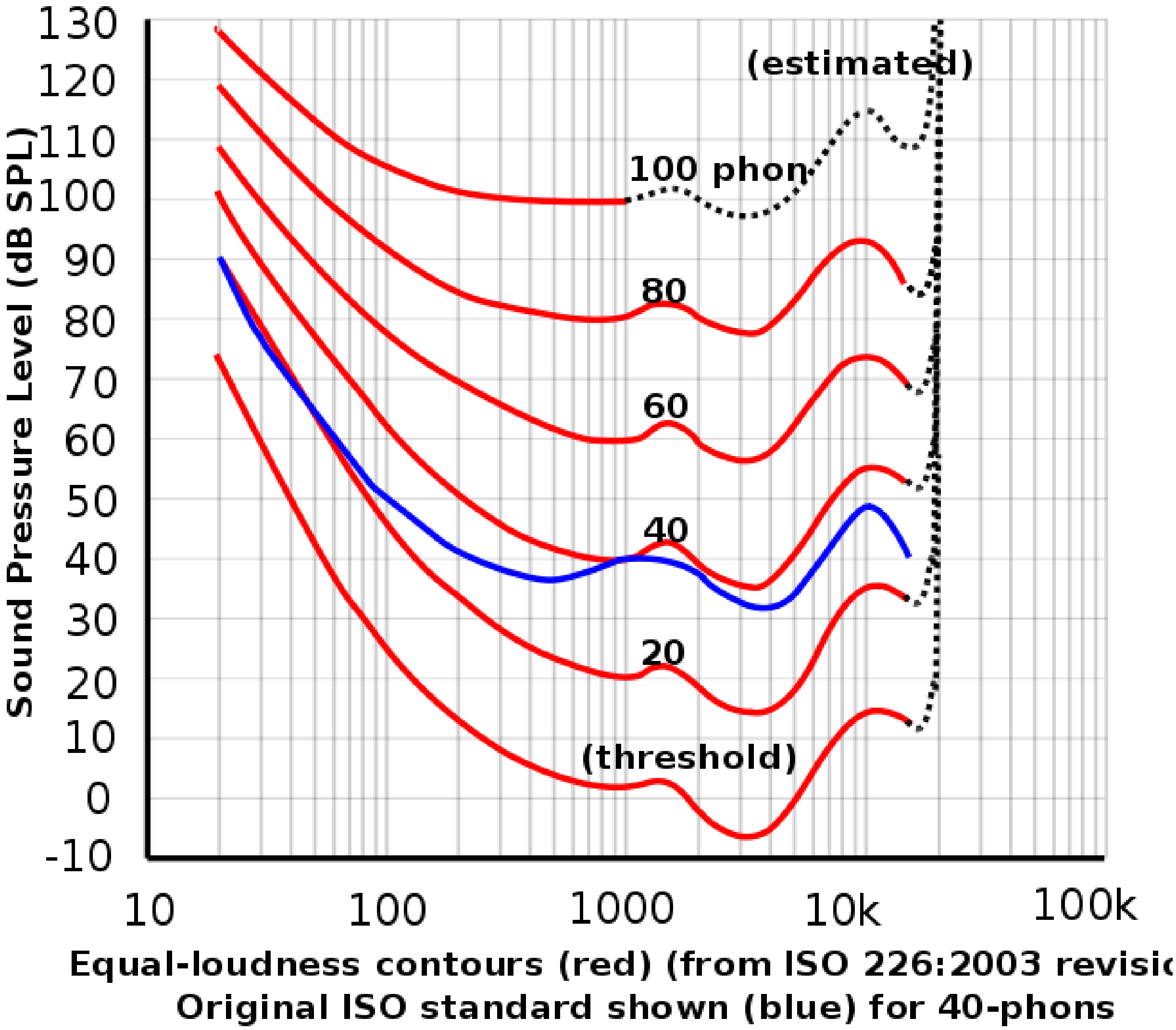

Figure 2.2: ISO equal-loudness contours with frequency in Hz.

2.3.4 Design and Procedure

We designed a virtual environment to ensure only the diffracted sound reached the subject. The environment was *effectively open*, the ground was perfectly absorptive, and the source was not kept in the line-of-sight.(Fig.4.9). Ideally, we would need to have an infinitely high obstacle and an infinitely long barrier as shown in Fig. 4.9 to make sure that the sound reaching the listener is the one that diffracts around the lateral plane of the obstacle. Otherwise, the sound would diffract around the top of the obstacle and the edge of the barrier. These dimensions can obviously not be achieved in practice and in order to **simulate** an open environment with seemingly infinite features, we enclosed the scene with a perfectly absorptive box. The box was 100m x 12m x 100m.

Another advantage of having an enclosure on the scene is that it acts as an infinitely far away region for wave-based methods. Wave-based methods inherently cannot handle open scenes and need an artificial boundary to *act* as infinity. In order to make sure this artificial infinity doesn't reflect back into the scene, an absorbing boundary condition called Perfectly Matched Layer (PML) (Berenger, 1994) is used at the boundaries. In order to add the PML boundary, the aforementioned bounding box is enclosed within a slightly bigger bounding box. The space between these boxes acts as the absorbing layer.

This makes the environment effectively open with respect to sound propagation and would give the same result as in a truly open, similar environment. Moreover, the environment also ensured that computing the propagated sound field with a wave-based propagation system was computationally tractable.

The barrier was added to one side of the cuboid to ensure that no sound could reach the subject from the other side. (Fig.4.9). The barrier, too, was perfectly absorbing to avoid spurious reflections in the scene. We tested up to two orders of diffraction. It should be noted that the concept of diffraction 'order' does not apply to wave-based methods but does in the case of geometric methods. An order of diffraction is defined as the number of edges sound has to diffract around in order to reach the listener. Typically, the cost of geometric diffraction methods such as UTD increases exponentially with increasing order.

This was a within-subject study with the same participants for each method of diffraction. The subjects wore a blindfold and the audio was monaural sound delivered using headphones. Before starting the experiments, subjects were played a sample of the sound source to make sure they were familiar with the sound source.

A total of 16 participants took part in each group. For each of the 19 positions, the subjects were asked to rate the loudness of the sound heard on an arbitrary, non-physical scale ranging from 1-20. The loudness scale was explained before the start of the experiment. The extrema of the scale were compared to a verbal standard: 1 was the loudness of a falling leaf while 20 was the loudness of someone shouting nearby. The sounds for the two methods were level-matched by matching the root mean square (RMS) of the signals in the direct sound region.

A block consisted of 38 trials(19 positions x 2 diffraction methods). There were three blocks per subject, giving a total of 114 readings. The virtual placement of the subjects was controlled by the experimenters, who pressed a key that placed the subjects at a random position along the semi-circle and randomly chose one of two diffraction methods. The subjects were allowed to take as many breaks as needed. Subjects took an average of 25-30 minutes for the entire experiment.

2.3.5 Results

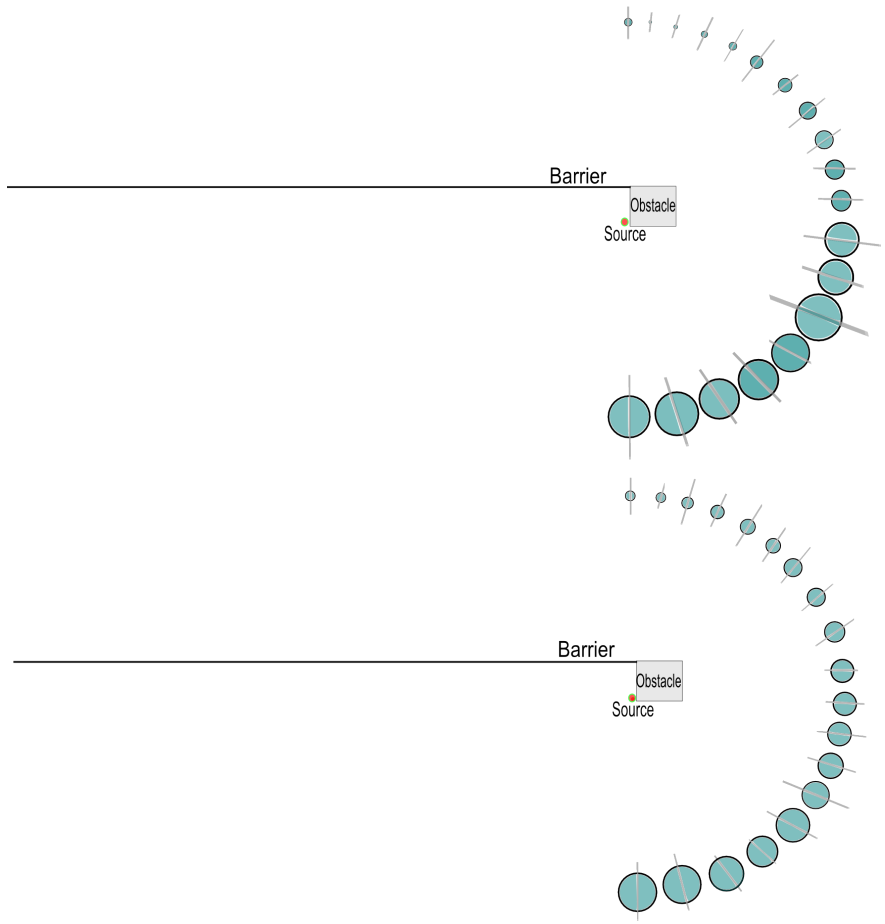

Figure 2.3: Mean subject ratings for different positions along a semi-circle for two methods of diffraction, plus one direct sound region. The cross-hatches show the between-subjects standard deviation of the responses. The upper figure shows the results for the UTD-based diffraction while the lower one shows results for wave-based diffraction. Wave-based diffraction shows a strongly linear decay while the UTD-based diffraction shows considerable deviation from linear decay.

To equate use of the scale for analysis, each subject's ratings were averaged over the three blocks, normalized by the subject's mean score (over all listener positions and diffraction methods), and then scaled by the grand-mean. The results for the are shown in Fig. 2.3, which shows the average values with the standard errors for different listener positions around the obstacle for the two methods of diffraction. As can be seen in the figure, the decay trend for the wave-based diffraction is more uniform than the UTD-based one, which shows substantial variation in the perceived loudness.

The significance of these trends was confirmed by a two-way, repeated measures ANOVA with factors diffraction method and position. It showed a significant 2-way interaction: position, F(18,270) = 48.64, $p < 0.001$, indicating variations in the positional response across the two methods.

To analyze the additional effect of diffraction order within a factorial design, the direct sound condition was eliminated, and a subsequent three-way ANOVA with factors diffraction method, diffraction angle (the 18 non-direct angles), and order was performed. It showed a significant 3-way interaction, F(8, 120) = 34.52, $p < .001$, indicating that the relation between diffraction order and diffraction angle changes with the diffraction method. All 2-way interactions were also significant: diffraction angle by diffraction method, F(8, 120) = 15.83, $p < 0.001$, diffraction angle by order, F(8, 120) = 20.10, $p < 0.001$, and diffraction method by order, F(1, 15) = 370.00, $p < 0.001$. The ANOVA also yielded significant main effects for diffraction angle F(8,120) = 288.99, $p < 0.001$ and diffraction order F(1, 15) = 672.54, $p < 0.001$.

2.4 Experiment 2: Reverberation Study I

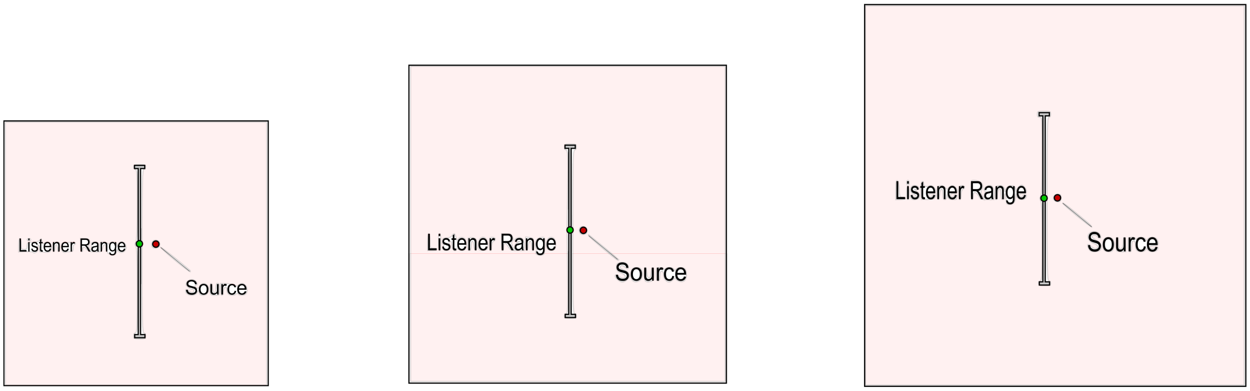

Figure 2.4: The reverberation experiment setup consisted of seven cubes of increasing dimension three of which are shown here. The listener path length in all the cubes was the same.

2.4.1 Participants

Twelve subjects took part in this study, all males. The ages ranged from 18 to 23 (Mean = 21 with SD = 1.5). All reported normal hearing. The participants were recruited from the staff and students at a university campus.

2.4.2 Apparatus and Stimuli

The setup consisted of a MacBook Pro laptop with the sound delivered through a pair of Beyerdynamic DT990 PRO headphones. The subjects were blindfolded.

Since reverberation forms the later part of the impulse response of a room, the initial part of the IR has to be remain consistent across the two cases to isolate reverberation as the only variable. In order to make sure that **only** the reverberation part of the impulse response is compared, we use the same system based on (Schissler et al., 2014a) to perform our task. This allows us to make sure that the early part of the IR is constant for both cases. In the first case, the IR is computed as is, without any modification corresponding to computing the reverberation using full impulse response. In the second case, we disable the physically computed reverberation and use a Schroeder-type filter (Schroeder, 1962) to compute the reverberation and add it to the final IR. The materials of the rooms were chosen to be gypsum board for all the surfaces including the ceiling and the floor. This was done because gypsum boards are commonly used in wall-paneling and hence would retain the subjects' familiarity with everyday environments. The sound source was a male voice.

Reverberation filters must be hand-tuned by setting the RT_{60} and Direct-to-reverberant (D/R) ratio in order to get the reverberation for an environment. Typically, these parameters are set by the audio designer based on his/her perception of what the environment would sound like. In order to remove this subjectivity, we decided to compute the parameters from the actual room parameters. The impulse response for each of the rooms was obtained and the parameters were estimated from the response as explained in Section 2. Subjective ratings of room size were obtained using the method of free-magnitude estimation.

The stimuli consisted of seven cubes of increasing dimensions ($100m^3 - 1600m^3$ in increments of $250m^3$) for each of the two reverberation conditions(Fig. 2.4). The sound source was placed near the center of the room slightly to the right. The subjects were allowed to move along a fixed path on the floor in order to hear the sound at different positions and get an average estimate of the reverberation in the room.

2.4.3 Design and Procedure

This was a between-subject study with the same participants for both methods of reverberation. The subjects wore a blindfold and headphones. The sound was rendered in stereo.

Before starting the study, the blindfolded subjects were played examples of sound originating in a small sized room and a large sized one. This was done so that the subjects could get an idea about the reverberation characteristics of the two ends of the room size spectrum and scale their internal response accordingly. They then underwent a training round in which each of the rooms were played in random order with one of the two reverberation methods. This was done in order to let the subjects come up with their own scale for scaling the room size. It is important to clarify that the subjects were *not asked to judge the size of the room in* m^3, instead, they were asked to give a dimensionless number representing how large they thought the rooms were.

Each subject rated the complete set of 7 rooms x 2 reverberation conditions with the order of the rooms randomized, on each of three separate blocks, giving a total of 42 measurements per subject (7 rooms x 2 reverberation conditions x 3 blocks). The subjects always started at the same initial position (relative to the room) and were asked to use one key on the keyboard to move forward and another to move backward. On reaching the end of the walking distance on either side, they were notified by the experimenter and asked to move in the opposite direction. The walking distance was kept constant irrespective of the size of the room, so that the subjects could not get a sense of the room size by the distance walked. The subjects were allowed to take as many breaks as needed. The average duration of the experiment was one hour. No fatigue was reported.

2.4.4 Results

To analyze the data, we first normalized it for each subject to account for the different scales adopted by the different subjects (Zwislocki and Goodman, 1980). This was done by first taking the mean of the subjects' scores for the three trial rounds. Then, each subject's mean score was computed over all the seven rooms and the two reverberation conditions. This mean was used to normalize the scores. Finally, the normalized scores were scaled by the grand mean over all subjects to give a sense of the scale used.

Fig. 2.5 shows the average magnitudes for each of the seven rooms for the two reverberation conditions across different blocks.

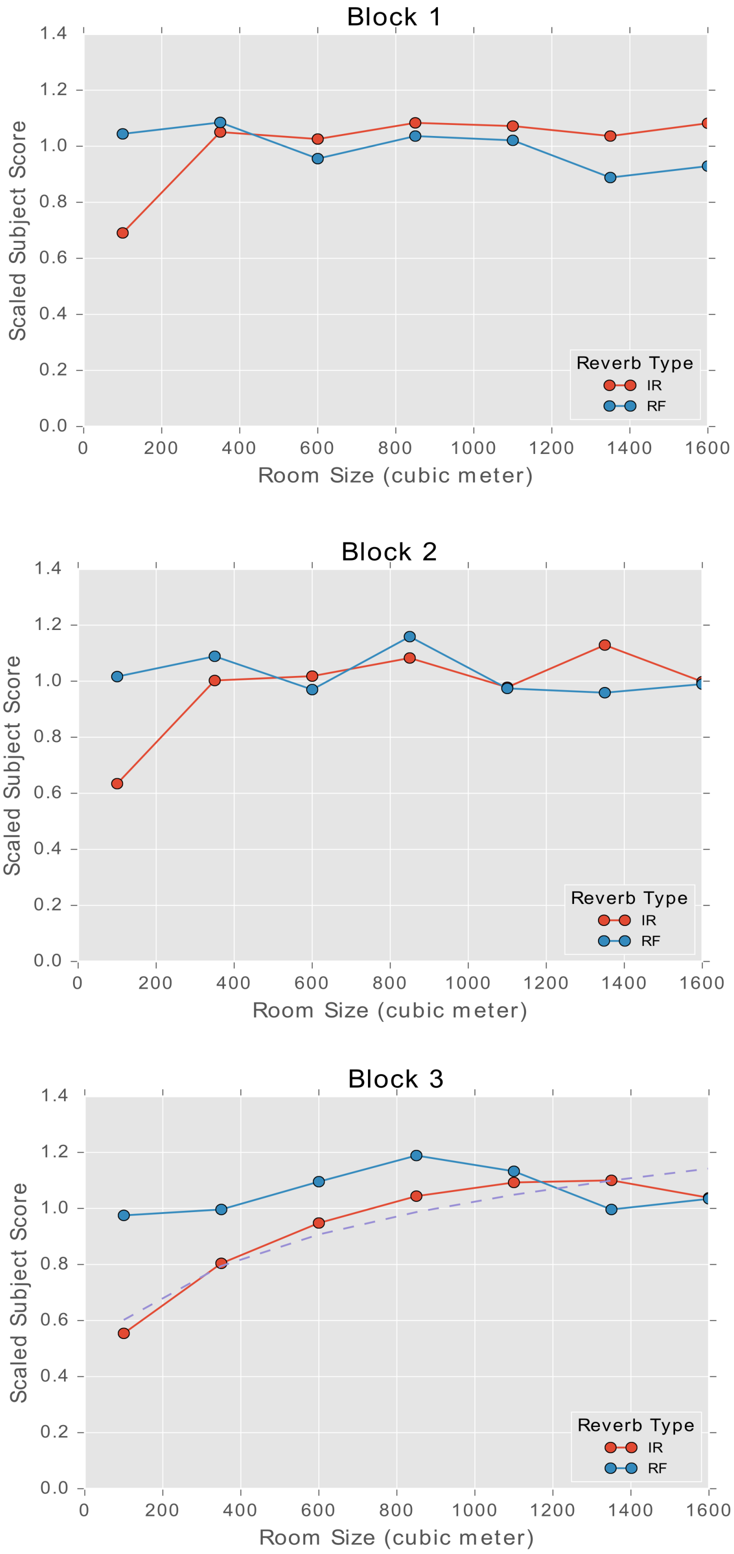

Figure 2.5: The average subjective magnitude given by the subjects for the two reverberation conditions across the three blocks of trials. The IR-based reverberation starts doing better as the experiment progresses and starts showing a logarithmic relation between the volume of the room and perceived reverberant intensity (Purple dotted line in Block 3 is the log fit). The reverberation filter shows no such learning effect.

The 3-way ANOVA with reverberation method, room size, and block showed three significant effects: First, there was a main effect of room size, F(6,66) = 2.63, p = 0.025, indicating that overall, participants did differentiate reverberant quality. Second, the interaction between room size with reverberation method was significant, F(6,66) = 4.01, p = 0.003, indicating that room size had different effects on the two reverberation methods , and finally, room size interacted with block, F(2,22) = 4.029, p = 0.034, indicating a change in room size perception with experience making the judgments. The interaction involving method led us to examine the two reverberation methods separately.

A 2-way ANOVA for the reverberation filter with trial block and room size failed to show any significance for block (F(2,22) = 0.794, p = 0.466), room size, F(6,66) = 0.759, p = 0.605, or room size with block, F(12,132) = 0.77, p = 0.678, thus indicating that neither the room size nor the trial block mattered. Essentially, the subjects were unable to discriminate different room sizes and did not show any improvement even when the experiment was repeated across blocks.(see Fig. 2.5)

In contrast, the same 2-way ANOVA for the reverberation computed using full impulse response showed significance for block, F(2,22) = 5.08, p = 0.016, and room size, F(6,66) = 5.855, $p < 0.001$. Although room size did not significantly interact with block (F(12,132) = 0.709, p = 0.741) (Fig. 2.5), it is clear that the changes with experience were confined to the smaller room sizes. A log fit to the block 3 mean data accounts for a substantial amount of the variance in the means across room size (91.5%). Thus, after two blocks of practice with the full impulse response method, subjects were differentiating the intensity of reverberation in an approximately logarithmic relation to room volume. Such compressive scaling of perceptual dimensions is commonly found, and the logarithmic relation known as Fechners law (Wolfe et al., 2014)

2.5 Experiment 3: Reverberation Study II

2.5.1 Participants

Seventeen participants took part in the study with informed consent. Their ages ranged from 19 to 47 (mean = 25.9 and SD = 7.4 - Four females, thirteen males). The participants were recruited from the students and staff at the university. All participants reported normal hearing.

2.5.2 Apparatus

The set up consisted of a Dell T7600 workstation and the sound was delivered via a pair of Beyerdynamic DT990 PRO headphones. The subjects were blindfolded for the study. The software to compute the RT_{60} and DRR was based on open-source MATLAB code (**?**). The calibration and auralization were done using in-house software, also written in MATLAB.

2.5.3 Stimuli

The source was a pre-recorded sound of human clapping. Since clapping is somewhat similar to an impulse, it tends to have a broad frequency content making wave-based methods impractical for virtual environments with such stimuli. The virtual environment consisted of a rectangular room $45m \times 10m \times 3m$ with highly reflective walls to create a highly-reverberant environment with an $8m$ walking path as shown in Figure 2.6. Seven omnidirectional sound sources were kept at increasing distances from the center of the path starting from $10m$ up to $40m$ in increments of $5m$. The sources were all kept at the same height of $1.7m$ from the floor. This value was chosen assuming a standard listener height of $1.7m$ in the virtual environment. The source sound power was 78dB.

2.5.4 Filter calibration

The filter was calibrated to match the reverberation characteristics of the geometric sound propagation system by appropriately scaling and splicing the early part of the impulse response ($\sim 80ms$), starting at the approximate onset time of reverberation to match the RT_{60} and DRR of the ray-traced impulse response. Further data and results on the calibration are available on the project website.

2.5.5 Design & Procedure

A rectangular room was chosen as the environment with highly reverberant environment similar to a painted, concrete room with no windows. In order to make sure we were comparing the *underlying methods* and not the specific parameters, we matched the RT_{60} and the DRR for both reverberation methods.

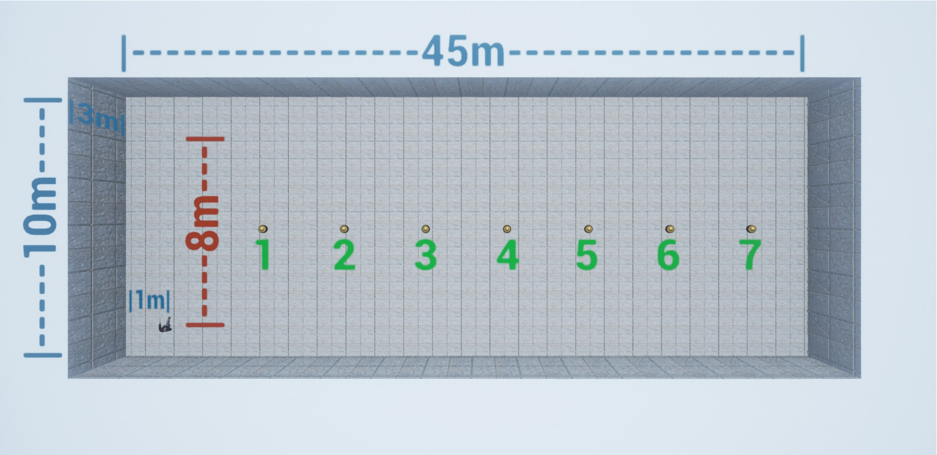

Figure 2.6: The room used for the experiment. The path marked in red is the walking path along which the subject walks. The sound sources are perpendicular to the walking path and kept at increasing distances from it. The labels 1-7 show the different source distances sampled uniformly from the range 10-40m

2.5.6 Training

Before the participants started the experiments, they completed a training task in a real-world setting. An $8m$ long walking path was constructed and the sound sources were placed at $3m$ and $6m$ from the center of the walking path, starting with $3m$. The participants were blindfolded before being led into the room so as to not give them an idea of the room dimensions. The dry (without reverberation) sound clip was played from a Harmon/Kardon HK 195 desktop speaker. The participants were asked to point at the sound source with their right hand and keep pointing at the source as they walked along the $8m$ path. Since the participants were blindfolded, they were helped by the test administrators as they walked down the path. Once they reached the end of the path, the participants were asked to give their best evaluation (in meters) as to how far from them they thought the sound source to be when it seemed closest to them. The training task was then repeated with the source moved to $6m$. The subjects were told the actual distances at the end of the training. The training exercise was not meant to be an exact replica of the experiment, as it was not possible to construct a physical room with the same kind of reverberance as the one in the virtual environment. Instead, the training was meant to give the participants a feeling for what to expect and how to make judgments.

2.5.7 Method

This was a within-subject study. The walking in the virtual environment was not controlled by the participants; instead, the 81 impulse responses per source were first convolved with sound source and then sampled such that each one of them contributed to 0.1m of the total 8m for a human traveling at an average speed of 1.39 m/s. The contributions from each of these 81 convolved impulse responses were spliced together (with interpolation) to create a sound file for each source. This sound file was played to the participants and they were asked to give the same estimate as they performed in the training, i.e., the perceived distance (in meters) of the sound when it seemed to be the closest to them. The impulse responses were spatialized using a generic HRTF-filter being applied to the direct sound and the early reflections. The participant's head orientation was fixed and they were always looking straight ahead. Each participant rated the complete set of 7 source positions $\times$ 2 reverberation methods with the order of the sources randomized for each block, giving a total of 42 (7 source positions $\times$ 2 methods $\times$ 3 blocks) judgments. The total time for the experiment, including the training, took around 15 minutes. The participants were allowed to take breaks between blocks, as required. No fatigue was reported.

2.5.8 Results

A 3-way ANOVA on block, distance, and reverberation method found a significant effect of method ($F(1, 16) = 15.29$, p < 0.01) and distance ($F(6, 96) = 29.12$, p < 0.01). All two-way interactions (block-distance, block-method, distance-method) failed to show significance. This finding indicates that the shape of distance compression is statistically the same for both reverberation methods, and the ray tracing algorithm exhibits an overall tendency to give longer distances. The null effect of block indicates that no trends are obscured by averaging over this factor.

2.6 Analysis

2.6.1 Diffraction Study

In any VE the sound field should not appear discontinuous to the listeners as that could break presence. Fig. 2.3 shows how the sound field was perceived by the subjects in the two cases. A simple linear-regression fit further solidifies the notion: The coefficient-of-determination, R^2_{UTD} evaluates to 0.91 whereas R^2_{Wave} = 0.98 for the same set of diffraction angles relative to the source.

The high R^2 values for both diffraction methods indicate that the subject responses for both the methods at least approximate a linearly-decreasing trend in terms of perceived loudness. While this is the expected trend for perceived loudness in such a setting as shown by our experiment, the UTD-based diffraction shows

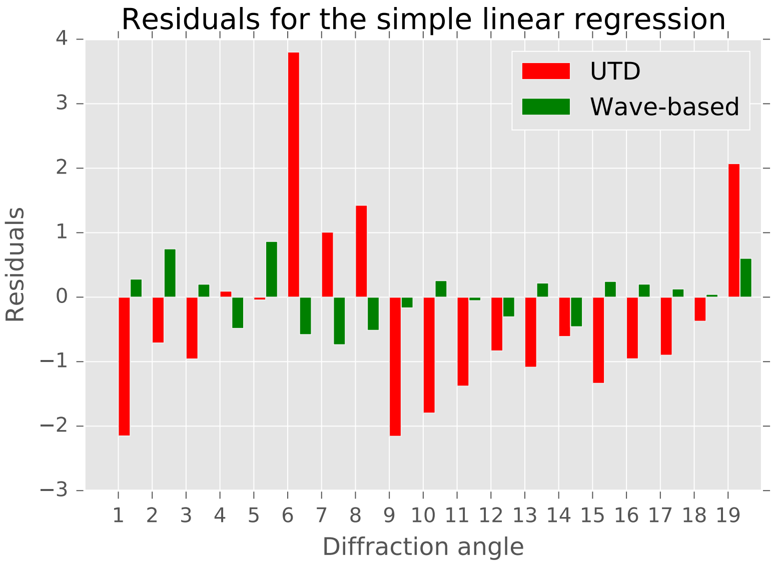

Figure 2.7: The residual error from the simple linear regression for both diffraction methods. As is clearly evident, UTD-based diffraction shows high error when trying to fit its response to a straight line as compared to the wave-based method. The X-axis represents the 19 equi-spaced diffraction angles ranging from 0° to 180°

substantially more deviation from linear. In contrast, the wave-based diffraction is highly linear in its trend. Fig. 2.7 shows how amenable each of the methods is to a linear model fitting.

Based on these results, it can be concluded that UTD-based diffraction is a reasonable diffraction model perceptually and for applications where some discontinuity in the sound field is acceptable (e.g., games). In addition, UTD serves as a relatively inexpensive means to get reasonably good diffracted sound. On the other hand, applications where presence should not be affected by abruptly changing sound fields should use wave-based methods.

2.6.2 Reverberation Study

Study I It is well-known that volume estimation is a complex phenomenon, and the trends in the data for size scaling from reverberant cues reflect this, in showing that people are relatively insensitive to these cues. We would expect to see a monotonically increasing trend in both cases. The full impulse response shows this characteristic, in that the subjective magnitude for the full impulse response shows an increasing trend over the smaller room sizes, but this trend quickly saturates. However, the subjects seem to learn to scale with this technique over the course of the experiment. The scaling by reverberation filter, on the other hand tends to vary in a non-systematic manner with increasing room size and shows high inter-subject variability as well.

Based on log-linear regression model, there is consistent poor performance in the case of the reverberation filter while the physically accurate reverberation improves with repetition. Fig. 2.8 shows how the value of R^2 varies with the block: reverberation filter performs poorly across the three blocks, indicating it doesn't lend itself to any meaningful relation between the perceived reverberation intensity and the volume of the environment. Accurate reverberation, on the other hand, starts showing improvement with $R^2_{(IR)_{B3}} = 0.92$ for the third block, indicating the logarithmic relation between volume and reverberation intensity becomes more prominent as subjects start to learn judging the volume based only on the sound cues alone.

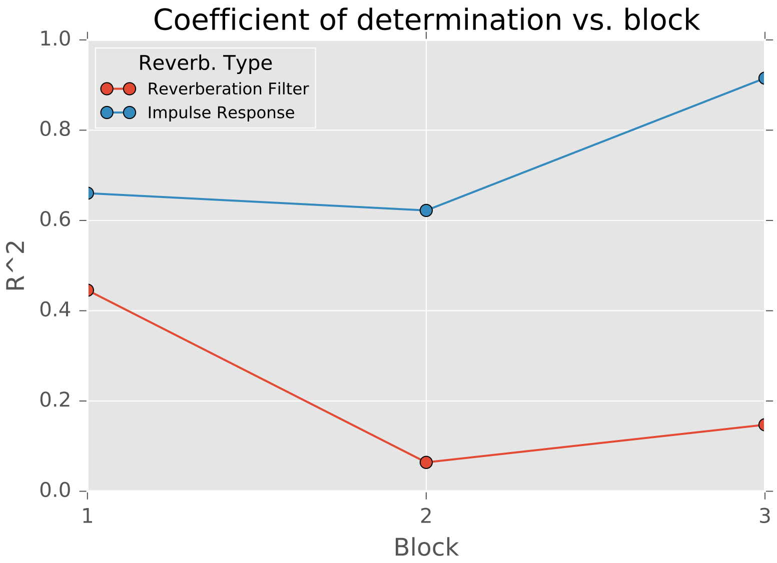

Figure 2.8: The value of R^2 for the reverberation types across all the three trial blocks.

Although, the data indicate that untrained listeners are not particularly good at judging the size of an environment, it appears using full impulse response offers some benefit in the way of consistency and seems to offer much superior discriminatory ability in room size perception as the subjects become trained. (Hameed et al., 2004) pointed out that subjects tend to only use the RT_{60} in room size perception and if fidelity with respect to this phenomenon is desired, the absorption coefficient of the room can be adjusted to get the proper relationship to the volume of the space.

2.7 Conclusions, Limitations, and Future Work

In this chapter we have presented two user studies to evaluate the perceptual merits of accurate and approximate interactive sound propagation algorithms. To the best of our knowledge, the diffraction study is the first of its kind in evaluating the perceived smoothness of the diffracted field for different methods; while the reverberation study evaluates the ability to provide accurate space cues for different methods. The study results show that accurate sound propagation algorithms offer an increased perceptual differentiation capability, exemplified by the reverberation study. Although, the subjects were not able to differentiate the volumes of the rooms very well, the accurate reverberation method resulted in a significantly better discriminatory capability compared to the reverberation filter.

The regression fits offer a concise and numerical value that encapsulate the perceptual differentiation capabilities of the algorithms and serve as simple metrics to evaluate the performance of these methods. We expect these metrics will serve as a means to let users decide which algorithms to use specific to their requirements. Currently, the R^2 values offer a very high level view of the algorithms and may not present an accurate picture for all scenarios. This can easily be the case if the wave-based diffraction had an unacceptably high peak at particular listener position(s), but still lent itself to a better linear fit than UTD. We believe that our choice of these metrics can be made more robust to such anomalies by using more sophisticated statistical models, as part of future work.

We would also like to point out that the environments in case of both experiments were kept very simple so that the approximate algorithms could work to their full potential. Normally, reverberation filters are tuned with respect to simple cubical environments, whereas UTD's performance starts degrading substantially with increasing number of diffraction edges. So, in a way, these experiments also represent the 'best case scenario' for the performance of the approximate algorithms. Since most VEs are going to be significantly more complicated, these approximate algorithms should be expected to perform worse than what they did in our experiments.

The studies, too, can be extended in multiple ways: It would be interesting to vary the spectral content of the source in the diffraction experiment, since diffraction is a frequency dependent phenomenon and evaluate the subjects' responses. The subject's distance from the source can also be varied and evaluated. The reverberation experiment could be verified by constructing real world rooms and using an actual sound source to verify the logarithmic relation of subjects' responses to changing room size.

CHAPTER 3: Mode-Aware Sound Propagation

3.1 Introduction

Realistic sound simulation can increase the sense of presence for users in games and VR applications (Durlach and Mavor, 1995; Shilling and Shinn-Cunningham, 2002). Sound augments both the visual rendering and tactile feedback, provides spatial cues about the environment, and improves the overall immersion in a virtual environment, e.g., playing virtual instruments (Ren et al., 2012; Serafin, 2004; Rocchesso et al., 2008; Young and Serafin, 2003) or walking interaction (Franinovic and Serafin, 2013; Nordahl et al., 2010; Steinicke et al., 2015; Visell et al., 2009; Turchet, 2015). Current game engines and VR systems tend to use pre-recorded sounds or reverberation filters, which are typically manipulated using digital audio workstations or MIDI sequencer software packages, to generate the desired audio effects. However, these approaches are time consuming and unable to generate appropriate auditory cues or sound effects that are needed for virtual reality. Further, many sound sources have a very pronounced directivity patterns which get propagated into the environment. And as these sources move, so do their directivities. Thus, it is important to model these time-varying, dynamic directivities propagating in the environment to make sure the audio-visual correlation is maintained and the presence not disrupted.

Recent trend has been on development of physically-based sound simulation algorithms to generate realistic effects. At a broad level, they can be classified into sound synthesis and sound propagation algorithms. Sound synthesis techniques (van den Doel et al., 2001; O'Brien et al., 2002b; Zheng and James, 2009; Chadwick et al., 2009; Zheng and James, 2011; Turchet et al., 2015) model the generation of sound based on vibration analysis of the object resulting in modes of vibration that vary with frequency. However, these techniques only model sound propagation in free-space and do not account for the acoustics effects caused by interaction of sound waves with the objects in the environment. On the other hand, sound propagation techniques (**??**Mehra et al., 2014b, 2015) model the interaction of sound waves with the objects in environment, but assume pre-recorded or pre-synthesized audio clips as input. Therefore, current sound simulation algorithms ignore the dynamic interaction between the processes of sound synthesis, emission (radiation), and propagation, resulting in inaccurate (or non-plausible) solutions for the underlying physical processes. For example, consider the case of a kitchen bowl falling from a countertop; the change in the directivity of the bowl with different hit positions and the effect of this time-varying, mode-dependent directivity on the propagated sound in the environment is mostly ignored by the current sound simulation techniques. Similarly, for a barrel rolling down the alley, the sound consists of multiple modes, where each mode has a time-varying radiation and propagation characteristic that depends on the hit positions on the barrel along with the instantaneous position and orientation of the barrel. Moreover, the interaction of the resulting sound waves with the walls of the alley cause resonances at certain frequencies and damping at others. Current sound simulation techniques model the barrel as a sound source with either static, mode-independent directivity, and model the resulting propagation in the environment with a mode-independent acoustic response or model the time-varying directivity of the barrel but propagate those in free-space only (Chadwick et al., 2009). Due to these limitations, artists and game audio-designers have to manually design sound effects corresponding to these different scenarios, which can be very tedious and time-consuming (Raghuvanshi and Snyder, 2014).

Main Results: In this chapter, we present the first coupled synthesis-propagation algorithm which models the entire process of sound simulation starting from the surface vibration of rigid objects, radiation of sound waves from these surface vibrations, and interaction of the resulting sound waves with the virtual environment for interactive applications. The key insights of our work is the use of a single-point multipole expansion (SPME) to couple the radiation and propagation characteristics of a source for each vibration mode. Mathematically, a single-point multipole corresponds to a single radiating source placed inside the object; this expansion significantly reduces the computational cost of the propagation stage compared to a multi-point multipole expansion. Moreover, we present a novel interactive mode-adaptive sound propagation technique that uses ray tracing to compute the per-mode impulse responses for a source-listener position. We also describe a novel perceptually-driven Hankel function approximation scheme that reduces the computational cost of this mode-adaptive propagation to enable interactive performance for virtual environments. The main benefits of our approach include:

1. Per-mode coupling of synthesis and propagation through the use of single-point multipole expansion.

2. Interactive mode-adaptive propagation technique based on perceptually-driven Hankel function approximation.

3. High degree of dynamism to model dynamic surface vibrations, sound radiation and propagation for moving sources and listeners.

Our technique performs end-to-end sound simulation from first principles and enables automatic sound effect generation for interactive applications, thereby reducing the manual effort and the time-spent by artists and game-audio designers. Our system can automatically model the complex acoustic effects generated in various dynamic scenarios such as (a) swinging church bell inside a reverberant cathedral, (b) swaying wind chimes on the balcony of Tuscany countryside house, (c) a metal barrel falling downstairs in an indoor game scene and (d) orchestra playing music in a concert hall, at 10fps or faster on a multi-core desktop PC. We have integrated our technique with the Unity^{TM} game engine and demonstrated complex sound effects enabled by our coupled synthesis-propagation technique in different scenarios

Furthermore, we evaluated the effectiveness of our perceptual Hankel approximation algorithm by performing a preliminary user-study. The study was an online one where the subjects where shown snippets of three benchmarks (Cathedral, Tuscany, and Game) with audio delivered through headphones/earphones and rendered using our perceptual Hankel approximation and using no approximation. The subjects were asked to judge the similarity between the two sounds for the three benchmarks. Initial results show that the subjects were unable to distinguish between the two sounds indicating that our Hankel approximation doesn't compromise on the audio quality in a perceptible way.

3.2 Related Work and Background

In this section, we give an overview of sound synthesis, radiation, and propagation and survey some relevant work.

3.2.1 Sound Synthesis for rigid-bodies

Given a rigid body, sound synthesis techniques solve the modal displacement equation

$$\mathbf{Kd} + \mathbf{C\dot{d}} + \mathbf{M\ddot{d}} = \mathbf{f}, \tag{3.1}$$

where $\mathbf{K}$, $\mathbf{C}$, and $\mathbf{M}$ are the stiffness, damping, and mass matrices, respectively and $\mathbf{f}$ represents the (external) force vector. This gives a discrete set of mode shapes $\hat{\mathrm{d}}_{\mathrm{i}}$, their modal frequencies ω_i, and the amplitudes $q_i(t)$.

The vibration's displacement vector is given by:

$$\mathbf{d(t)} = \mathbf{Uq(t)} \equiv [\hat{d_1}, ..., \hat{d_M}]\mathbf{q(t)}, \tag{3.2}$$

where M is total number of modes and $\mathbf{q(t)} \in \Re^M$ is the vector of modal amplitude coefficients $q_i(t)$ expressed as a bank of sinusoids:

$$q_i(t) = a_i e^{-d_i t} sin(2\pi f_i t + \theta_i), \tag{3.3}$$

where f_i is the modal frequency (in Hz.), d_i is the damping coefficient, a_i is amplitude, and θ_i is the initial phase.

(Adrien, 1991) introduced modal analysis approach to synthesizing sounds. (van den Doel et al., 2001) introduced a measurement-driven method to determine the modes of vibration and their dependence on the point of impact for a given shape. Later, (O'Brien et al., 2002b) were able to model arbitrarily shaped objects and simulate realistic sounds for a few of these objects at interactive rates. This approach is called the *modal analysis* and requires an expensive precomputation, but achieves interactive runtime performance. The number of modes generated tend to increase with the geometric complexity of the objects. (Raghuvanshi and Lin, 2006) used a system of spring-mass along with perceptually motivated acceleration techniques to generate realistic sound effects for hundreds of objects in real time. (Ren et al., 2010) developed a contact model to capture multi-level surface characteristics based on (Raghuvanshi and Lin, 2006). Recent work on modal synthesis also uses the single point multipole expansion (Zheng and James, 2011).

3.2.2 Sound Radiation and Propagation

Sound propagation in frequency domain is described using the Helmholtz equation

$$\nabla^2 p + \frac{\omega^2}{c^2} p = 0, \quad \mathbf{x} \in \Omega, \tag{3.4}$$

where $p = p(\mathbf{x}, \omega)$ is the complex-valued pressure field, ω is the angular frequency, c is the speed of sound in the medium, and ∇^2 is the Laplacian operator. To simplify the notation, we hide the dependence on angular frequency and represent the pressure field as $p(\mathbf{x})$. Boundary conditions are specified on the boundary of the domain $\partial\Omega$ by either the Dirichlet boundary condition that specifies the pressure on the boundary $p = f(\mathbf{x})$ on $\partial\Omega$, the Nuemann boundary condition that specifies the velocity of the medium $\frac{\partial p(\mathbf{x})}{\partial n} = f(\mathbf{x})$ on $\partial\Omega$, or a mixed boundary condition that specifies $Z \in \mathbb{C}$, so that $Z\frac{\partial p(\mathbf{x})}{\partial n} = f(\mathbf{x})\ on\ \partial\Omega$. The boundary condition at infinity is also specified using the *Sommerfeld radiation condition* (Pierce et al., 1981)

$$\lim_{r\to\infty}[\frac{\partial p}{\partial r} + i\frac{\omega}{c}p] = 0, \tag{3.5}$$

where $r = ||x||$ is the distance of point **x** from the origin.

Equivalent Sources: The uniqueness of the acoustic boundary value problem guarantees that the solution of the free-space Helmholtz equation along with the specified boundary conditions is unique inside Ω (Ochmann, 1999). The unique solution $p(\mathbf{x})$ can be found by expressing the solution as a linear combination of *fundamental solutions*. One choice of fundamental solutions is based on *equivalent sources*. An equivalent source $q(\mathbf{x}, \mathbf{y_i})$ is the solution of the Helmholtz equation subject to the Sommerfeld radiation condition. Here **x** is the point of evaluation, y_i is the source position and $\mathbf{x_i} \neq \mathbf{y_i}$. The equivalent source can be expressed as:

$$q(\mathbf{x}, \mathbf{y_i}) = \sum_{l=0}^{L-1}\sum_{m=-l}^{l} c_{ilm}\varphi_{lm}(\mathbf{x}, \mathbf{y_i}) = \sum_{k=1}^{L^2} e_{ik}\varphi_k(\mathbf{x}, \mathbf{y_i}), \tag{3.6}$$

where k is a generalized index for (l, m), φ_k are multipole functions, and c_{ilm} is the strength of multipoles. Multipoles are given as a product of two functions:

$$\varphi_{lm}(\mathbf{x}, \mathbf{y_i}) = \Gamma_{lm}h_l^{(2)}(kd_i)\psi_{lm}(\theta_i, \phi_i), \tag{3.7}$$

where (d_i, θ_i, ϕ_i) is the vector $(\mathbf{x} - \mathbf{y_i})$ expressed in spherical coordinates, h_l^2 is the spherical *Hankel* function of the second kind, k is the wavenumber given by $\frac{\omega}{c}$, $\psi_{lm}(\theta_i, \phi_i)$ are the complex-valued spherical harmonics functions, and Γ_{lm} is the normalizing factor for the spherical harmonics.

3.2.2.1 Sound Radiation

The Helmholtz equation is the mathematical way to model sound radiation from vibrating rigid bodies. Boundary element method is a widely used method for solving acoustic radiation problems (Ciskowski and Brebbia, 1991) but has a major drawback in terms of high memory requirements. An efficient technique known as the Equivalent source method (ESM) (Ochmann, 1999) exploits the uniqueness of the solutions to the acoustic boundary value problem. ESM expresses the solution field as a linear combination of equivalent sources of various orders (monopoles, dipoles, etc.) by placing these simple sources at variable locations inside the object and matching the boundary conditions on the object's surface, guaranteeing the correctness of solution. The pressure at any point in Ω due to N equivalent sources located at $\{y_i\}_{i=1}^{N}$ can be expressed as a linear combination:

$$p(\mathbf{x}) = \sum_{i=1}^{N} \sum_{l=0}^{L-1} \sum_{m=-l}^{m=l} c_{ilm} \varphi_{lm}(\mathbf{x}, \mathbf{y_i}). \tag{3.8}$$

This compact representation of the pressure $p(\mathbf{x})$ makes it possible to evaluate the pressure at any point of the domain in an efficient manner. This is also known as the *multi-point multipole expansion*. Typically, this expansion uses a large number of low-order multipoles ($L = 1$ or 2) placed at different locations inside the object to represent the pressure field. (James et al., 2006b) use this multi-point expansion to represent the radiated pressure field generated by a vibrating object. Another variant of this, is the *single-point multipole expansion* represented as

$$p(\mathbf{x}) = \sum_{l=0}^{L-1} \sum_{m=-l}^{m=l} c_{lm} \varphi_{lm}(\mathbf{x}, \mathbf{y}). \tag{3.9}$$

discussed in (Ochmann, 1999). In this expansion, only a single multipole of high order is placed inside the object to match outgoing radiation field.

3.2.2.2 Geometric Sound Propagation

Geometric sound propagation techniques use the simplifying assumption that the wavelength of sound is much smaller than the features on the objects in the scene. As a result, these methods are most accurate for high frequencies and approximately model low-frequency effects like diffraction and scattering as separate phenomena. Commonly used techniques are based on image source methods and ray tracing. Recently, there has been a focus on computing realistic acoustics in real-time using algorithms designed for fast simulation. These include beam tracing (Funkhouser et al., 1998b) and ray-based algorithms (Lentz et al., 2007; Taylor et al., 2009a) to compute specular an diffuse reflections and can be extended to approximate edge diffraction. Diffuse reflections can also be modeled using acoustic rendering equation (Siltanen et al., 2007; Antani et al., 2012a). In addition, frame-to-frame coherence of the sound field can be utilized to achieve a significant speedup (Schissler et al., 2014b).

3.2.3 Coupled Synthesis-Propagation

Ren et al. (Ren et al., 2012) presented an interactive virtual percussion instrument system that used modal synthesis as well as numerical sound propagation for modeling a small instrument cavity. However, the coupling proposed in this system did not incorporate a time-varying, mode-dependent radiation and propagation characteristic of the musical instruments. Additionally, this system only modeled propagation inside the acoustic space of the instrument and not the full 3D environment. Furthermore, the volume of the underlying acoustic spaces (instruments) in (Ren et al., 2012) was rather small in comparison to the typical scenes shown in this chapter.

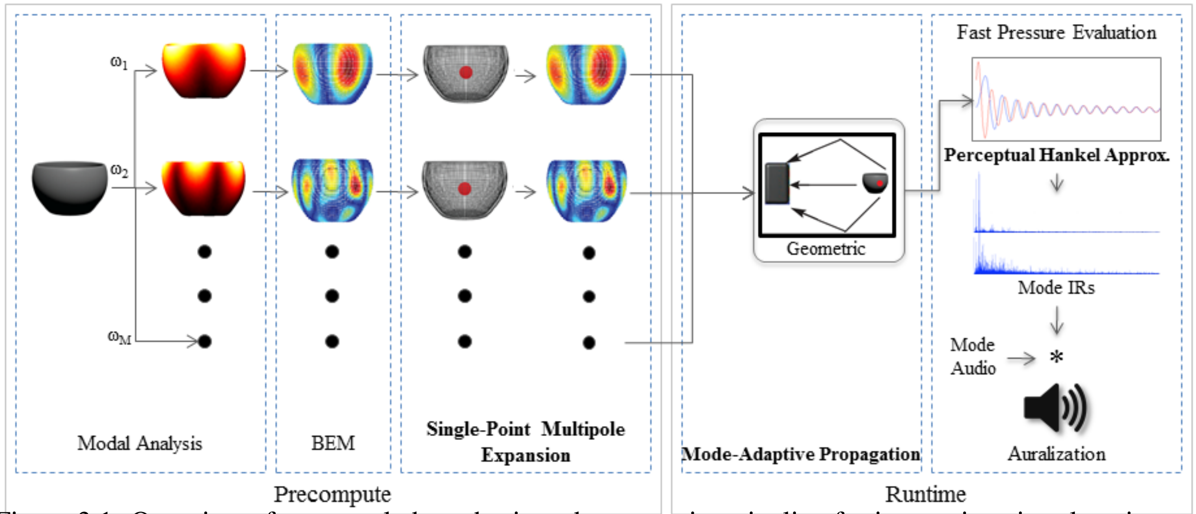

Figure 3.1: Overview of our coupled synthesis and propagation pipeline for interactive virtual environments. The first stage of precomputation comprises the modal analysis. The figures in red show the first two sounding modes of the bowl. We then calculate the radiating pressure field for each of the modes using BEM, place a single multipole at the center of the object, and approximate the BEM evaluated pressure. In the runtime part of the pipeline, we use the multipole to couple with an interactive propagation system and generate the final sound at the listener. We present a new perceptual Hankel approximation algorithm to enable interactive performance. The stages labeled in bold are the main contributions of our approach.

3.3 Overview

In this section, we provide an overview of our mode-adaptive, coupled synthesis-propagation technique (see Figure 3.1).

The overall technique can be split into two main stages: *preprocessing* and *runtime*. In the preprocessing stage, we start with the vibration analysis of each rigid object to compute its modes of vibrations. This step is performed using the finite element analysis of the object mesh to compute displacements (or shapes), frequencies, and amplitudes of all the modes of vibration. The next step is to compute the sound radiation field corresponding to each mode. This is done by using the mode shapes as the boundary condition for the free-space Helmholtz equation and solving it using the state-of-the-art boundary element method (BEM). This step computes the outgoing radiation field corresponding to each vibration mode. To enable interactive evaluation at runtime, the outgoing radiation fields are represented compactly using the *single-point multipole expansion* (Ochmann, 1999). This representation significantly reduces the runtime computational cost for sound propagation by limiting the number of multipole sources to one per mode instead of hundreds or even thousands per mode in the case of multi-point multipole expansion (Ochmann, 1999; James et al., 2006b). This completes our preprocessing step. The coefficients of the single-point multipole expansion are stored for runtime use.

At runtime, we use a mode-adaptive sound propagation technique that uses the single-point multipole expansion as the sound source for computing sound propagation corresponding to each vibration mode. In order to achieve interactive performance, we use a novel perceptually-driven Hankel function approximation. The sound propagation technique computes the impulse response corresponding to the instantaneous position for source-listener pair for each vibration mode. High modal frequencies are propagated using the geometric sound propagation techniques. Low modal frequencies can be propagated using the wave-based techniques. Hybrid techniques combine geometric and wave-based techniques to perform sound propagation in the entire frequency range. The final stage of the pipeline takes the impulse response for each mode, convolves it with that mode's amplitude, and sums it for all the modes to give the final audio at the listener.

We now describe each stage of the pipeline in detail.

Modal Analysis: We adopt a finite element method (O'Brien et al., 2002b) to precompute the modes of vibration of an object. In this step, we first discretize the object into a tetrahedral mesh and solve the modal displacement equation (Eq. 3.1) analytically under the Raleigh-damping assumption (i.e. damping matrix $\mathbf{C}$ can be written as a linear combination of stiffness $\mathbf{K}$ and mass matrix $\mathbf{M}$). This facilitates the diagonalization of the modal displacement equation, which can then be represented as a generalized eigenvalue problem and solved analytically as system of decoupled oscillators. The output of this step is the vibration modes of the object along with the modal displacements, frequencies, and amplitudes. (Ren et al., 2013) showed that the Raleigh damping model is a suitable geometry-invariant sound model and is therefore a suitable choice for our damping model.

Sound Radiation: This step computes the sound radiation characteristic of the vibration modes of each object by solving the free-space Helmholtz equation (James et al., 2006b). The modal displacements of each mode serves as the boundary condition for the Helmholtz equation. The boundary element method (BEM) is then used to solve the Helmholtz equation and resulting outgoing radiation field is computed on an offset surface around the object. This outgoing pressure field can be efficiently represented by using either the single-point or multi-point multipole expansion.

Single-point Multipole fitting A key aspect of our approach is to represent the radiating sound fields for each vibrating mode in a compact basis by fitting the single-point multipole expansion, instead of a multi-point expansion. This representation makes it possible to use just one point source position for *all the vibration modes*. This formulation makes it possible to perform interactive modal sound propagation (Eq. 3.9).

Mode-Adaptive Sound Propagation: The main idea of this step is to perform sound propagation for each vibration mode of the object independently. The single-point multipole representation calculated in the previous step is used as the sound source in this step. By performing a mode-adaptive propagation, our technique models the mode-dependent radiation and propagation characteristic of sound simulation. The modal frequencies generated for the objects in our scenes tend to be high (i.e., more than 1000Hz). Ideally, we would like to use wave-based propagation algorithms (Mehra et al., 2014b, 2015), as they are regarded more accurate. However, the complexity of wave-based methods increase as a fourth power of the frequency, and therefore they can very high time and storage complexity. We use a mode-adaptive sound propagation based on geometric methods.

Geometric Propagation: Given single-point multipole expansions of the radiation fields of a vibrating object, we use a geometric acoustic algorithm based on ray-tracing to propagate the field in the environment. In particular, we extend the interactive ray-tracing based sound propagation algorithm (Schissler et al., 2014b; Schissler and Manocha, 2015) to perform mode-aware propagation. As discussed above, we use a single source for all the modes and trace rays from this source into the scene. Then, at each listener position, the acoustic response is computed for each mode by using the pressure field induced by the rays and scaled by the mode-dependent radiation filter corresponding to the the single-point multipole expansion for that mode. In order to handle low-frequency effects, current geometric propagation algorithm use techniques based on uniform theory of diffraction. While they are not as accurate as wave-based methods, they can be used to generate plausible sound effects for virtual environments.

Auralization: The last stage of the pipeline involves computing the final audio corresponding to all the modes. We compute this by convolving the impulse response of each mode with the mode's amplitude and summing the result:

$$q(\mathbf{x}, t) = \sum_{i=1}^{M} q_i(t) * p^{\omega_i}(\mathbf{x}, t), \tag{3.10}$$

where $p^{\omega_i}(\mathbf{x}, t)$ is the acoustic response of the ith mode with angular frequency ω_i computed using sound propagation, $q_i(t)$ is the amplitude of the ith mode computed using modal analysis, $\mathbf{x}$ is the listener position, M is the number of modes, and $*$ is the convolution operator.

3.4 Coupled Synthesis-Propagation

In this section, we discuss in detail the single-point multipole expansion and the mode-adaptive sound propagation.

3.4.1 Single-Point Multipole Expansion

There are two types of multipole expansions that can be used to represent radiating sound fields: *single-point* and *multi-point*. In a single-point multipole expansion (SPME), a single multipole source of high order is placed inside the object to represent the sound field radiated by the object. On the other hand, multi-point multipole expansion places a large number of low order multipoles at different points inside the object to represent the sound field. Both SPME and MPME are two different representations of the outgoing pressure field and do not restrict the capabilities of our approach in terms of handling near-field and far-field computations.

To perform sound propagation using a multipole expansion, the number of sound sources that need to be created depend on the number of modes and the number of multipoles in each mode. In case of a single-point expansion, the number of sound sources is equal to M where M is the number of modes since the number of multipoles in each expansion is 1. In case of multi-point multipole expansion, the number of sound sources is equal to $\sum_i^M N_i$ where N_i is the number of multipoles in ith mode. The number of multipoles at each mode vary with the square of the mode frequency. This results in thousands of sound sources for multi-multipole expansion. The computational complexity of a sound propagation technique (wave-based or geometric) varies with the number of sound sources. As a result, we selected SPME in our approach. However, it is possible that there are some cases where low-order MPME could be more efficient than a single and very high-order SPME. However, in the benchmarks used in the chapter, SPME results in efficient runtime performance.

Previous sound propagation approaches have proposed the use of source clustering to reduce the computation required for scenes with many sources (Tsingos et al., 2004). However, these techniques cannot be used to cluster multipoles as the clustering disrupts the phase of the multipoles, producing error in the sound radiation. Therefore, we chose to use a single-point multipole expansion to enable interactive sound propagation at runtime.

The output of this stage is the set of coefficients of the single-point multipole expansion (Eq. 3.9) for each mode (for example, coefficients c^{ω}_{lm} for mode ω).

3.4.2 Mode-adaptive Sound Propagation

We now propose a *position invariant* method of computing the sound propagation for each mode of the vibrating object. This approach brings down the number of sound sources to be propagated from M to just one. This is achieved by placing the SPME for all the modes at exactly the same position. Given a ray-tracing based geometric technique, this implies that instead of tracing rays for each mode separately, we trace rays from only a single source position. These rays are emitted from the source in different directions, get reflected/diffracted/scattered/absorbed in the scene, and reach the listener with different pressure values. Mode-dependent impulse response is computed for each mode by multiplying the pressure values produced by the traced rays with the corresponding SPME weights for each ray. We describe this approach in detail as follows:

Sound propagation is split into two computations: *mode-independent* and *mode-dependent* computations.

Mode-independent: We make use of the ray-based geometric technique of (Schissler et al., 2014b) to compute sound propagation paths in the scene. This system combines path tracing with a cache of diffuse sound paths to reduce the number of rays required for an interactive simulation. The approach begins by tracing a small number (e.g., 500) of rays uniformly in all directions from each sound source. These rays strike the surfaces and are reflected recursively up to a specified maximum reflection depth (e.g., 50). The reflected rays are computed using vector-based scattering (Christensen and Koutsouris, 2013), where the resulting rays are a linear combination of the specularly reflected rays and random Lambertian-distributed rays. The listener is modeled as a sphere the same size as a human head. At each ray-triangle intersection, the visibility of the listener sphere is sampled by tracing a few additional rays towards the listener. If some fraction of the rays are not occluded, a path to the listener is produced. A path contains the following output data: The total distance the ray traveled d, along with the attenuation factor α due to reflection and diffraction interactions. Diffracted sound is computed separately using the UTD diffraction model (Tsingos et al., 2001b). The frequency dependent effects are computed using a vector of frequency attenuation coefficients given the mode's frequency for both diffraction and reflection. This step remains the same for all the modes since the position of the source remains the same (across all the modes) as described above.

Mode-dependent: Given the output of the geometric propagation system, we can evaluate the mode-dependent acoustic response for a mode with angular frequency ω as:

$$p^{\omega}(\mathbf{x},t) = \sum_{r\in R} |p^{\omega}_r(\mathbf{x})| \;\; w_r \;\; \delta(t - d_r/c), \tag{3.11}$$

where w_r is the contribution from a ray r in a set of rays R, d_r is the distance traveled by the ray r, c is the speed of sound, δ is the delta function, and $p^{\omega}_r(\mathbf{x})$ is the pressure contribution generated by the ray r for mode ω computed using the single-point multipole expansion:

$$p^{\omega}_r(\mathbf{x}) = \alpha_r \sum_{l=0}^{L-1} \sum_{m=-l}^{m=l} c^{\omega}_{lm} \varphi^{\omega}_{lm}(d_r, \theta_r, \phi_r), \tag{3.12}$$

where φ^{ω}_{lm} is the multipole, k is wavenumber of the mode ($k = \omega/c$), (θ_r, ϕ_r) is the direction of emission of ray r from the source, and α_r is the attenuation factor. We switch between $h^{(2)}_l(kd_r)$ and its approximate variant $\tilde{h}^{(2)}_l(kd_r)$ based on the distance d_r in a mode-dependent manner as described next.

These mode-dependent acoustic responses are used in the auralization step as described in Section 3.

3.4.3 Hankel Approximation

The spherical Hankel function of the second kind, $h^{(2)}_l(kd)$, describes the radially-varying component of the radiation field of a multipole of order l. It is a complex-valued function of the distance d from the multipole position and the wave number $k = \omega/c$. This function itself is a linear combination of the spherical Bessel functions of the first and second kind, $j_l(kd)$ and $y_l(kd)$: $h^{(2)}_l(kd) = j_l(kd) - iy_l(kd)$. (Abramowitz and Stegun, 1972). These Bessel functions are often evaluated to machine precision using a truncated infinite power series.

While this computation of the Bessel functions is accurate, it is also slow when the functions need to be evaluated many times. Within sound propagation algorithm, both Bessel functions need to be evaluated for each mode and each sound path through the scene. The number of paths in a reflective scene (e.g. cathedral) can easily exceed 10^5, and the number of modes for the sounding objects is around 20 to 40, resulting in millions of Bessel function evaluations per frame. The Hankel function is also amenable to computation using recurrence relation(s). One such relation is given as:

$$h_{l+1}^{(2)}(kd) = \frac{2l+1}{kd} h_l^{(2)}(kd) - h_{l-1}^{(2)}(kd) \tag{3.13}$$

Unfortunately, computing the Hankel function using this recurrence relation has similar runtime costs as evaluating the Bessel functions, and can become a bottleneck for interactive applications. If the Hankel function is used directly, its evaluation for all modes and paths can take seconds.

Another possibility is to precompute a table for different values, and perform table lookup at runtime. However, such an approach is not practical, since Hankel is a 2D function (l, kd). For a table, the granularity of the arguments would have to be extremely fine, given the high numeric sensitivity of the function. Although, it would be easy to store the values of l and k as they're known beforehand, the value of d can have a large range, even for a small scene. This is because the value of d depends on the distance a ray travels as it reaches the listener position which could include multiple bounces in the environment.

Perceptual Hankel Approximation: We present an approximation technique for evaluation of the Hankel function for interactive applications. Our approach uses a perceptually-driven error threshold to switch between the full function evaluation and the approximation. We use the approximation function given by (Mehra et al., 2014b):

$$h_l^{(2)}(kd) \approx \tilde{h}_l^{(2)}(kd) = i^{l+1} \frac{e^{-ikd}}{kd}. \tag{3.14}$$

This approximation converges to $h_l^{(2)}(kd)$ for large values of kd, but does not match well near the multipole. For this reason, we apply this approximation only in the *far field*, where the value of the distance d is greater than a threshold distance $d_{\tilde{h}}$. Overall, the approximation works well even for small scenes since the reflected rays can take a long path before they reach the listener and be in the *far field*.

We determine this distance threshold independently for each mode frequency ω and its corresponding wave number k so that a perceptual error threshold is satisfied. We derive the error threshold for each mode from the absolute threshold of hearing at the mode's frequency. If the pressure error from the approximation is less than the threshold of hearing, the difference in pressure is unable to be perceived by a human listener (Painter and Spanias, 2000). The threshold of hearing can be well-approximated by the analytic function (Terhardt, 1979):

$$\begin{aligned} T_q(f) =& 3.64(f/1000)^{-0.8} - 6.5e^{-0.6(f/1000-3.3)^2)} + \\ & 10^{-3}(f/1000)^4. \text{ (dB SPL)}, \end{aligned} \tag{3.15}$$

SPL stands for Sound Pressure Level and is measured in decibels (dB).

In a preprocessing step, we evaluate this function at each mode's frequency to determine a per-mode error threshold, and then determine the distance threshold $d_{\tilde{h}}$ where the approximation is perceptually valid for the mode. This information is computed and stored for each sounding object. At runtime, when the pressure contribution for each path i is computed, we use the original Hankel $h_l^{(2)}(k_i d_i)$ when $d_i < d_{\tilde{h}}$ and the approximation $\tilde{h}_l^{(2)}(k_i d_i)$ when $d_i \geq d_{\tilde{h}}$.

We would like to note that although the approximation to Hankel function specified in Eq 3.14 is standard, the novelty of our approach lies in the way we use it. As described above, we use perceptually-driven thresholds to decide when to automatically switch to the approximate version. We also did a user-evaluation to make sure the perceptually-motivated approximation doesn't cause any loss of quality in our context. The details of the evaluation are presented in the Section 5.

Error Threshold Preprocessing: Given a perceptual error threshold such as $\epsilon = 5$ dB SPL, we use a brute-force approach to determine the smallest value of $d_{\tilde{h}}$ for which the error of the approximation is less than ϵ for all distances $d > d_{\tilde{h}}$. We have included Figure 3.2 that shows an example of how the error shrinks at increasing values of d. Our approach starts at the multipole position and samples the error value at $\lambda/10$ to avoid aliasing. The method stops when d reaches a point past the end of the longest expected impulse response (e.g., 1000m). The final value for $d_{\tilde{h}}$ is chosen to be the last d sample where the error dropped below ϵ.

The result of applying this approximation is that our sound propagation system is able to handle pressure computation for interactive scenes that are much more complex and with many more sound paths than with the original Hankel formulation. In addition, the error due to our approach is small and not perceptible by a human listener.

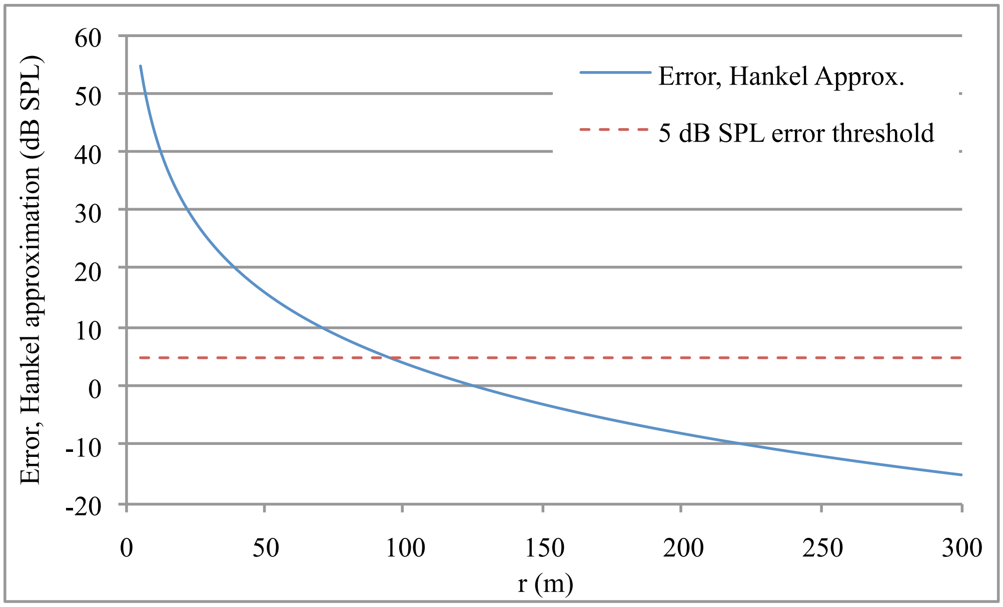

Figure 3.2: The error between the Hankel function approximation $\tilde{h}_l^{(2)}(kd)$ and the original function $h_l^{(2)}(kd)$ decreases at increasing values of d for order $l = 6$ and mode frequency 1000Hz. An error threshold of $\epsilon = 5$ dB SPL is overlaid. For this case, the approximation threshold distance is chosen to be $d_{\tilde{h}} = 93$m. All sound paths for this mode frequency with $d > 93$m use this approximation.

Near-field vs. far-field: As mentioned in Sec 4.1, Equivalent Source theory states that if the pressure on the offset surface is matched by matching the appropriate boundary condition, the pressure field is valid in the near-field as well as far-field. We use the perceptual Hankel approximation for far-field computation, but we don't truncate the order of the multipole anywhere. In particular, we use the exact multipole formulation everywhere with the following difference: the Hankel function part of multipole is approximated in the far-field but the expansion is never truncated anywhere in the domain. So the only difference in the computation of near and far-fields is in terms of Hankel computation.

3.5 User-Evaluation of Hankel Approximation

In order to evaluate the accuracy of our chosen thresholds, we performed an online user-study with 3 benchmark: the Cathredal, Tuscany, and the Game benchmark. Given the scope of our experiments, an online study was the best choice as it offered the subjects convenience of taking the study as per their convenience and at a pace they were comfortable with. This also eased the process of keeping their identities confidential. We generated the audio for these scenes using the perceptual Hankel approximation and the full Hankel computation. The Tuscany benchmark has the Unity in-game, static soundscape playing and was left that

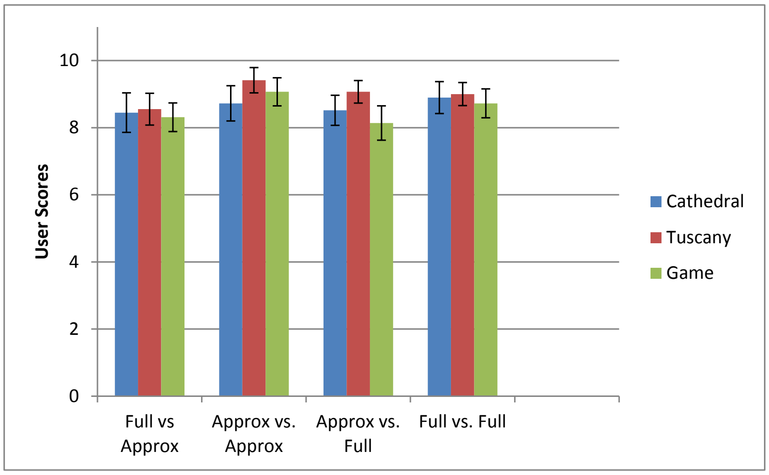

Figure 3.3: Mean and standard errors of the subjects' scores on the user-study. Full refers to sound computed using Full Hankel, while Approx refers to sound computed using our perceptual approximation. The response is to the question,*"Compared to the audio in the left video, how similar is the audio in the right video?"*

way to make scene appear more natural and have a better audio-visual correlation. For the study, we consider the full Hankel computation to be the *base* method while the approximated-Hankel was considered as *our* method.

Participants The study was taken by 29 subjects all within the age group of 18 and 50 with 18 males and 11 females. The mean age of all the participants was 27.3 and all of them reported normal hearing. The subjects were recruited by sending out emails to the departments, colleagues, and friends. The subjects were not paid for their participation.

Procedure The participants were given instructions on the study and asked to fill out a questionnaire on their background. The subjects were required to have a headphone/earphone before they could take part in the study. There was one test scene to help them calibrate their headphones/earphones and make sure they're oriented correctly (right channel on right ear, left channel on left). We designed four cases: *base vs. base, our vs. base, base vs. our, and our vs. our* for each of the three scenes. In total, twelve video pairs were generated for the benchmarks (4 cases x 3 benchmarks). We performed an online survey where subjects were presented the four cases in a random order and asked to answer a single question, "*Compared to the audio in the left video, how similar is the audio in the right video ?*". The choice of the question was motivated by (Polk et al., 2002; Aldrich et al., 2009) where the authors use a similar question and a similar scale to measure similarity between two stimuli. Our hypothesis was: Sound produced by *our* method would be indistinguishable from the *base* method. If our hypothesis is validated, it would *indicate* that our Hankel approximation is perceptually equivalent to full Hankel computation. The subjects were then presented the 12 benchmarks in a random order and asked to rate the similarity on a scale on 1 to 11 with 1 being the audio in the two videos is very different and 11 being the audio in the two videos is virtually the same. There was no repetition of stimuli to make sure there was no learning between subsequent iterations given the low number of stimuli present. The study had no time constraints and the participants were free to take breaks in-between the benchmarks as long as the web-session did not expire. After presenting the 12 benchmarks, the subjects were given the opportunity to leave open (optional) comments. Although, it is difficult to ascertain the average time it took the subjects to finish the study, in our experience, the study took around 15-20 minutes on average.

Results and Discussion The questions posed to participants of the study include mixed cases between audio generated using the full Hankel and approximate Hankel functions as well as cases where either the full or approximate Hankel function was used to generate both audio samples in a pair. Our hypothesis is thus that the subjects are going to rate the full vs. approximate similar to what they rate full vs. full, which would indicate that users are unable to perceive a difference between results generated using the full functions and those generated using their approximation. The mean values and the standard errors are shown in the Fig 3.3. The figure shows how close the mean scores are for the full vs. approximate test as compared to the full vs. full test.

The responses were analyzed using the non-parametric Wilcoxon signed-rank test on the full vs. full and approximate vs. approximate data to ascertain whether their population mean ranks differ. The Wilcoxon signed-rank test failed to show significance for all the three benchmarks: Cathedral (Z = -0.035, p = 0.972), Tuscany (Z = -1.142, p = 0.254), and Game (Z = 0.690, p = 0.49) indicating that the population means do not differ for all the three benchmarks. The responses were also analyzed using the non-parametric Friedman test. The Friedman test, too, failed to show significance for the benchmarks: Cathedral ($\chi^2(1)$ = 0.048, p = 0.827), Tuscany ($\chi^2(1)$ = 2.33, p = 0.127), Game ($\chi^2(1)$ = 0.053, p = 0.819).

The responses were further analyzed using confidence interval approach to show equivalence between the groups. The equivalence interval was chosen to be the $\pm 20\%$ of our 11-point rating scale, i.e., ± 2.2. The confidence level was chosen to be 95%. Table 3.1 shows that the lower and upper values of the confidence intervals lie within our equivalence intervals indicating that the groups are equivalent.

Scene	Full vs. Approx		Approx vs. Approx		Approx vs. Full		Full vs. Full	
	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
Cathedral	-0.4021	1.2054	-0.3587	1.0754	-0.3064	0.9184	-0.3259	0.9769
Tuscany	-0.3246	0.9729	-0.2572	0.7710	-0.2298	0.6890	-0.2350	0.7043
Game	-0.2919	0.8751	-0.2856	0.8562	-0.3504	1.0502	-0.2935	0.8798

Table 3.1: Equivalence test results for the three scenes. The equivalence interval was ±2.2 while the confidence level was 95%

3.6 Implementation and Results

In this section, we describe the implementation details of our system. All the runtime code was written in C++ and timed on a 16-core workstation with Intel Xeon E5 CPUs with 64 GB of RAM running Windows 7 64-bit. In the preprocessing stage, the eigen decomposition code was written in C++, while the single-point multipole expansion was written in MATLAB.

Preprocessing: We used finite element technique to compute the stiffness matrix $\mathbf{K}$ which takes the tetrahedralized model, Young's modulus, and the Poisson's ratio of the sounding object and compute the stiffness matrix for the object. Next, we compute the eigenvalue decomposition of the system using Intel's MKL library (DSYEV) and calculate the modal displacements, frequencies, and amplitudes in C++. The code to find the multipole strengths was written in MATLAB, the Helmholtz equation was solved using the FMM-BEM (Fast-multipole BEM) method implemented in FastBEM software package. Our current implementation is not optimized. It takes about 1-15 hours on our current benchmarks.

Sound Propagation: We use a fast, state-of-the-art geometric ray tracer (Schissler et al., 2014b) to get the paths for our pressure computation. This technique is capable of handling very high orders of diffuse and specular reflections (e.g., 10 orders of specular reflections and 50 orders of diffuse reflections) and still

Scene	#Tri.	#Paths	#S	#M	Time		
					Prop.	Pres.	Tot
Sibenik	77083	30850	1	15	52.2	57.9	110.1
Game	100619	58363	1	5	69.5	22.7	92.2
Tuscany	98274	9232	3	14	62.2	16.8	79
Auditor.	12373	13742	3	17	82.5	12.5	95

Table 3.2: We show the performance of our runtime system (mode-adaptive propagation). The number of modes for Tuscany and Auditorium is the sum over all sources used. The number of modes and number of paths were chosen to give a trade-off for speed vs. quality. All timings are in milliseconds. We show the breakdown between ray-tracing based propagation (Prop.) and pressure (Pres.) computation and the total (Tot) time per frame on a multi-core PC. #S is the number of sources and #M is the number of modes.

maintain interactive performance. The ray tracing system scales linearly with the number of cores keeping the propagation time low enough for the entire frame to be interactive (see Table 3.2).

Spherical Harmonic computation: The number of spherical harmonics computed per ray varies as $O(L^2)$, making naive evaluation too slow for an interactive runtime. We used a modified version of available fast spherical harmonic code (Sloan, 2013) to compute the pressure contribution of each ray. The available code computes only the real spherical harmonics by making extensive use of SSE (Streaming SIMD Extension). We find the complex spherical harmonics from the real ones following a simple observation:

$$Y_l^m = \frac{1}{\sqrt{2}}(Y_l^m + \iota Y_l^{-m}) \quad m > 0, \tag{3.16}$$

$$Y_l^m = \frac{1}{\sqrt{2}}(Y_l^m - \iota Y_l^{-m})(-1)^m \quad m < 0. \tag{3.17}$$

Since our implementation uses the recurrence relation to compute the associated Legendre polynomials along with extensive SIMD usage, it makes it faster than the GSL implementation and significantly faster other implementation such as BOOST.

Approximate Hankel Function: As mentioned in Section 4, the Hankel function is approximated when the listener is sufficiently far away from the listener. The approximate Hankel function $\tilde{h}_l^{(2)}(kd) = i^{l+1}\frac{e^{-ikd}}{kd}$ reduces to computing $sin(kd)$ and $cos(kd)$. In order to accelerate this computation further, we use a lookup table for computing sines and cosines, improving the approximate Hankel computation by a factor of about four, while introducing minimal error as seen in Section 7.3. The lookup table for the sines and cosines make no noticeable perceptual difference in the quality of sound.

Parallel computation of mode pressure: In order to make the system scalable, we parallelize over the number of paths in the scene rather than the number of modes. Parallelizing over the number of modes would not be beneficial if *number of cores* $>$ *number of modes*. Since the pressure computation for each ray is done independent of the other, the system parallelizes easily over the paths in the scene. We use OpenMP for the parallelization on a multi-core machine. Further, the system is configured to make extensive use of SIMD allowing it to process 4 rays at once. Refer to Table 3.2 for a breakdown of time spent on pressure computation and propagation for the different scenes.

Real-Time Auralization: The final audio for the simulations is rendered using a streaming convolution technique (Egelmeers and Sommen, 1996). Once the audio is rendered, it can be played on the usual output devices such as headphones or multi-channel stereo. Although, headphones would give the best results in terms of localization. All audio rendering is performed at a sampling rate of 44.1 kHz.

3.6.1 Results

Objeect	#Tris	Dim. (m)	#Modes	Freq. Range (Hz)	Order
Bell	14600	0.32	20	480 - 2148	13-36
Barrel (Auditorium)	7410	0.6	20	397 - 2147	13-37
Barrel (Game)	7410	1.03	9	370 - 2334	8-40
Chime - Long	3220	0.5	4	780 - 2314	7-19
Chime - Medium	3220	0.4	6	1135 - 3958	10-24
Chime - Short	3220	0.33	4	1564 - 3495	10-15
Bowl	20992	0.35	20	870 - 5945	8-36
Drum	7600	0.72	13	477 - 1959	8-28
Drum stick	4284	0.23	7	1249 - 3402	7-15
Trash can	7936	0.60	5	480 - 1995	11-17

Table 3.3: We show the characteristics of SPME for different geometries and materials.

We now describe the different scenarios we used to test our system.

Cathedral: This scene serves as a way to test the effectiveness of our method in a complex indoor environment. We show a modal object (Bell) that has impulses applied to it. As the listener moves about in the scene the intensity of sound varies depending on the distance of the listener from the bell. Further, since the cathedral corresponds to an indoor environment, effects such as reflections and late reverberation coupled with modal sounds become apparent.

Tuscany: The Tuscany scene provides a means to test the indoor/outdoor capabilities of our system. The modal object (Three bamboo chimes) is placed on the balcony with the wind providing the impulses. As the listener goes around the house and moves inside, the propagated sound of the chimes changes depending on the position of the listener in the environment. The sound is much less in intensity outside owing to most of the propagated sound being lost in the environment and increases dramatically when the listener goes in.

Game Scene: This demo showcases the effectiveness of our system in a game like environment containing, both, an indoor and a semi-outdoor environment. We use a metal barrel as our sounding object and let the listener interact with it. Initially, the barrel rolls down a flight of stairs in the indoor part of the scene. The collisions with the stairs serve as input impulses and generate sound in an enclosed environment, with effects similar to that in the Cathedral scene. The listener then picks up the barrel and rolls it out of the

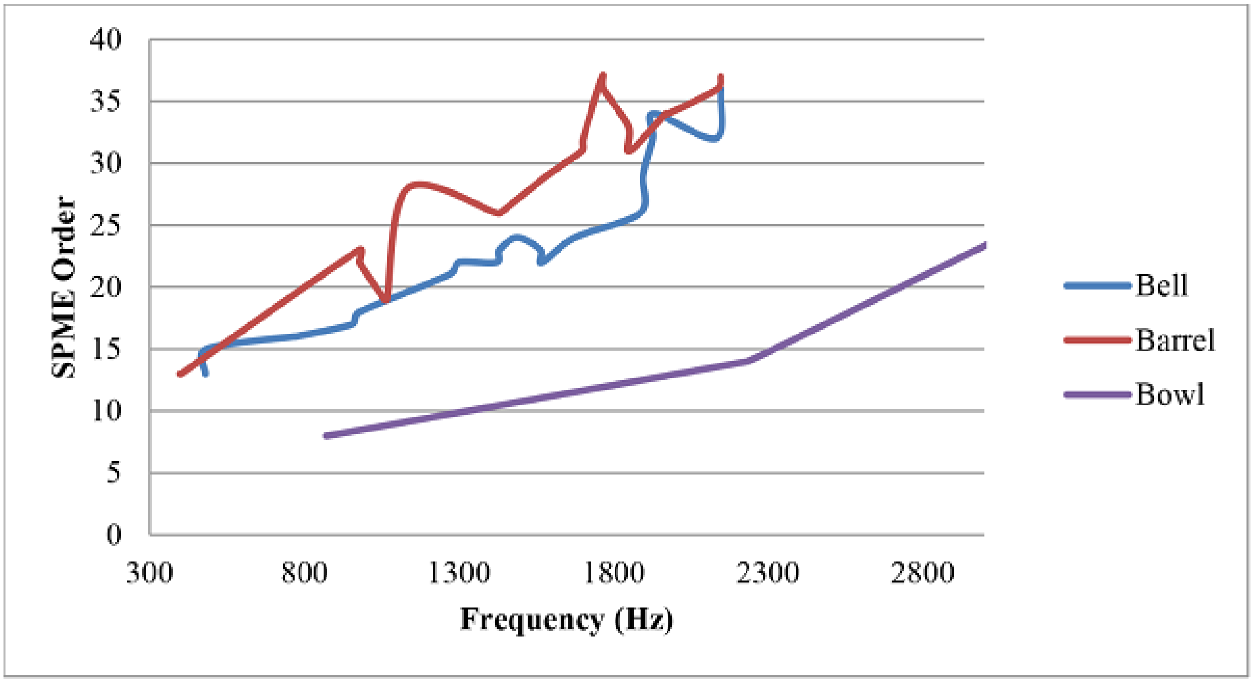

Figure 3.4: The order required by the Single-Point multipole generally increases with increasing modal frequency. We show the results for the objects used in our simulations. It is possible for the same modal frequency (for different objects) to have different order multipole owing to difference in geometries of these objects. The plot shows the SPME order required for approximating the radiation pattern of different objects as a function of their increasing modal frequencies.

door and follows it. As soon as the barrel exits the door, the environment outside is a semi-outdoor one, the reverberation characteristics change, demonstrating the ability of our system to handle modal sounds with different environments in a complex game scene.

Auditorium: This scene showcases the ability of our system to support multiple sound sources and propagate them inside an environment. We use a metal barrel, bell (from the Cathedral), a toy wooden drum, a drum stick, and a trash can lid to form a garage band. The *instruments* play a joyful percussive piece and provide the listener with the sound from a particular seat in the auditorium. (Fig. 3.5)

3.6.2 Analysis

Fig 3.4 shows the different orders of Single-Point Multipoles needed for the different objects as function of their modal frequencies. We choose an error threshold based on (Mehra et al., 2014b) as our error threshold ϵ when computing the co-efficients of SPME for a particular mode. The order of the SPME is iterated till the error drops below ϵ. We used $\epsilon = 0.15$ for each mode. (Fig. 3.6)

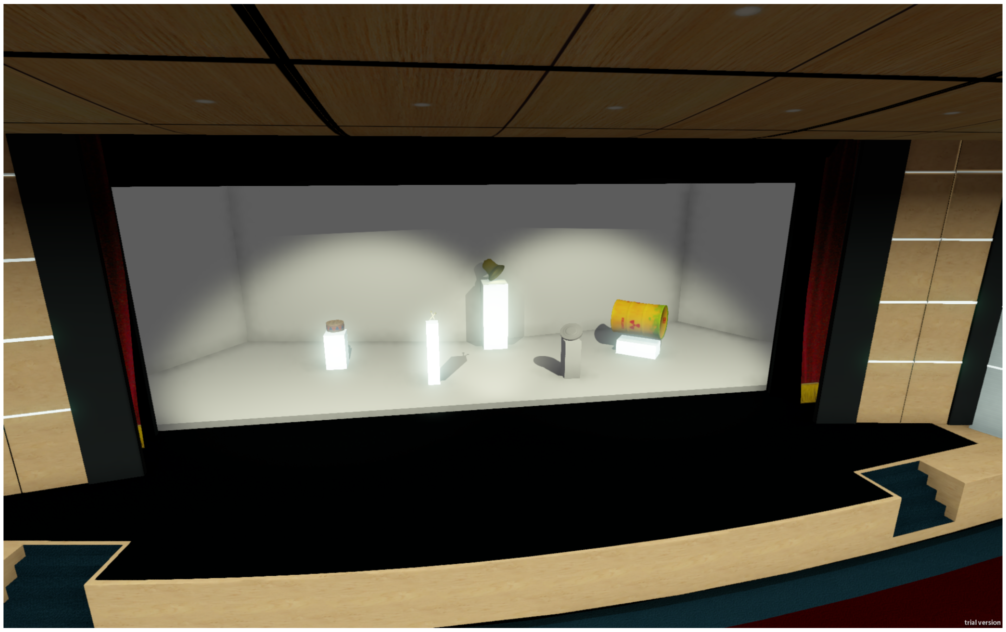

Figure 3.5: The Auditorium Music Scene. This scene includes multiple sources playing a musical composition.

We have included a table (Table 3.4) that shows the performance improvement we get in various scenes with our Perceptual-Hankel approximation. The results were computed on a single thread. The first three scenes had the listener moving around in the scene and being at different distances from the sounding object. This indicates that the listener moves in and out of the near-field of the object (Refer to the supplemental video). And as the table indicates, approximation is still at least 3x faster than full Hankel computation without loss in quality.

Scenario	#Paths	F-Hankel(ms)	P-Hankel(ms)	Speed-up
Sibenik	42336	7837.72	1794.5	4.37
Game	55488	5391.6	754.5	7.14
Tuscany	6575	225.73	69.75	3.23
Auditorium	11889	1395	284.75	4.9

Table 3.4: The speed-up obtained using the Perceptual-Hankel approximation. We achieve at least $3 - 7$x speed-up with no loss in the perceptual quality of sound. Here, F-Hankel stands for Full-Hankel while P-Hankel stands for Perceptual-Hankel. The results for Tuscany and Auditorium are averaged over all the sources.

Table 3.2 shows that we can achieve interactive performance (10 fps) using our system. The number of modes and number of rays in the scene can be controlled in order to get the best performance vs. quality balance. Table 3.5 shows the case for the Cathedral scene. The bell has 20 computed modes with about 44k rays on the one end and 13k rays with 1 mode on the other. The framework can be customized to suit the

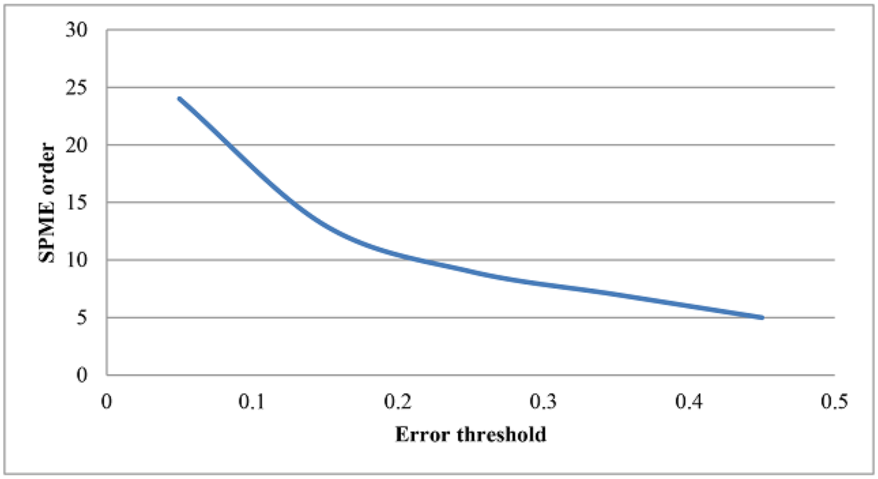

Figure 3.6: For an increasing error threshold ϵ, the order of the multipole decreases almost quadratically. This demonstrates our SPME algorithm provides a very good approximation.

needs of a particular scenario to offer the best quality/cost ratio. Further, owing to the scalable nature of our system, more number of cores scales the performance almost linearly.

#Paths	Prop. Time	1 mode	5 modes	10 modes	15 modes	20 modes
44148	84.23	3.23	15.46	31.9	60.8	152.5
30850	52.27	2.2	11.2	29.9	57.9	144.1
22037	37.8	2	10.5	31.3	61	127.9
13224	25	1.6	9.4	27.8	53.7	102.7

Table 3.5: The table shows how controlling the number of rays and the number of modes can influence the timing in the Cathedral scene with a bell. This can help one customize the system to provide the best quality/performance ratio for a particular scenario. The total time taken is propagation time + time for chosen number of modes. All times are reported in milliseconds.

3.7 Limitations, Conclusion and Future Work

We present the first coupled sound synthesis-propagation algorithm that can generate realistic sound effects for computer games and virtual reality, by combining modal sound synthesis, sound radiation, and sound propagation. The radiating sound fields are represented in a compact basis using a single-point multiple expansion. We perform sound propagation using this source basis via a fast ray-tracing technique to compute the impulse responses using perceptual Hankel approximation.

The resulting system has been integrated and we highlight the performance in many indoor and outdoor scenes. Our user-study demonstrates that perceptual Hankel approximations doesn't degrade sound quality and results in interactive performance. To the best of our knowledge, ours is the first system that successfully combines these methods and can handle a high degree of dynamism in term of source radiation and propagation in complex scenes.

Our approach has some limitations. Our current implementation is limited to rigid objects and modal sounds. Moreover, the time complexity tends to increase with the mode frequency. Our single-point multipole expansion approach can result in high orders of multipoles. The geometric sound propagation algorithm may not be able to compute the low frequency effects (e.g. diffraction) accurately in all scenes. Moreover, the wave-based sound propagation algorithm involves high pre-computation overhead and is limited to static scenes.

There are several avenues for future work. In addition to overcoming these limitations, we can further integrate other acceleration techniques, such as mode compression, mode culling etc (Raghuvanshi and Lin, 2006) for use in more complex indoor and outdoor environments and generate other sound effects in large virtual environments (e.g. outdoor valley). It would also be useful to consider the radiation efficiency of each mode and use more advanced compression techniques (Raghuvanshi and Snyder, 2014). It would be useful to accelerate the computations using iterative algorithms like Arnoldi's (Arnoldi, 1951). Integrating non-rigid synthesized sounds, e.g., liquid sounds (Moss et al., 2010) into our framework would be an interesting direction of future research. Our system is fully compatible with binaural rendering techniques such as HRTF-based (Head Related Transfer Function) rendering and it is our strong belief that using such techniques would improve the degree of presence that our system currently provides. (Begault et al., 1994; Larsson et al., 2002a). To this end, we would like to incorporate fast HRTF extraction methods such as (Meshram et al., 2014) and evaluate the benefits. Our current user-evaluation can be expanded in multiple ways that might reveal interesting perceptual metrics which might further help optimize the system. Finally, we would like to use these approaches in VR applications and evaluate their benefits.

CHAPTER 4: Diffraction Kernels for Interactive Sound Propagation

4.1 Introduction

Research in virtual environments over the last few decades has demonstrated that improved sound simulation and rendering can significantly augment a user's sense of presence (Larsson et al., 2002b). Sound can induce a sense of "object presence" and "spatial presence" at the same time, raising the fidelity of VR and AR simulations (Dubois et al., 2010). As the characteristics of the environment or source locations vary in real time, it is important to perform interactive auralization that accurately captures any changes caused by the user or the environment, and to generate smoothly rendered audio.

The most accurate algorithms for sound simulation are based on directly solving the acoustic wave equation using numerical methods and compute the pressure field. Recently, different precomputation-based solvers have been proposed to compute an acoustic kernel, which is used at runtime for interactive propagation for dynamic sources or listeners (James et al., 2006a; Raghuvanshi et al., 2016; Yeh et al., 2013). However, these technique have two major limitations: (a) the precomputation time and the memory overhead can be very high requiring large compute clusters; (b) they are limited to static scenes and cannot handle dynamic objects, a common scenario in virtual environments.

Most interactive algorithms for sound simulation and rendering for dynamic scenes are based on geometric acoustics and ray tracing (Krokstad et al., 1968b; Lentz et al., 2007; Taylor et al., 2012; Schissler et al., 2014b). Recent ray tracing algorithms can handle a high number of sources and compute higher order reflections at interactive rates on commodity desktop processors (Cao et al., 2016). It is well-known that pure geometric-acoustics techniques work well for high frequencies, and can't model low-frequency wave effects such as diffraction or occlusion. In practice, it is important to model these wave effects to correct the spectral content of reflected sound from finite surfaces, such as overhead reflectors or wall edges. Diffraction becomes very important for listeners located inside the shadow zones of obstacles and the inaccurate modeling of these effects can lead to a loss of realism in VR (Rungta et al., 2016).

There is significant literature on augmenting the geometric acoustic techniques with diffraction approximation. Most prior techniques are designed to model edge diffraction (Tsingos et al., 2001c; Taylor et al., 2012; Schissler et al., 2014b), which includes propagating sound around the corners as well as scattering sound in all directions from wedges of any angle.

However, current interactive diffraction algorithms have some limitations. First, they are less accurate for highly tessellated objects or smooth surfaces and can result in discontinuous sound field in occlusion scenarios. Second, it is computationally challenging to handle highly tessellated objects because the computational complexity increases exponentially with the number of diffracting edges.

Main Results: We present a novel approach based on object-based diffraction kernels to model sound propagation in dynamic environments. Our hybrid formulation combines the accuracy benefits of wave-based computation with the efficiency and flexibility of geometric ray-tracing methods. The resulting approach can handle virtual environments composed of highly tessellated, dynamic objects at interactive rates and offers these benefits:

- Efficient source-placement algorithm that significantly reduces the precomputation time by reducing the required number of wave-based simulations needed to compute the diffraction kernel of an object.
- Handles highly tessellated or smooth objects while modeling diffraction and occlusion effects.
- Efficient runtime based on ray tracing with minimal overhead enabling interactive performance for dynamic scenes.

In the preprocessing stage, we efficiently compute the *diffraction kernels* that encapsulate the sound interaction behavior of individual objects in *free field*. These kernels capture all the interactions of sound waves with the objects, including reflections, diffraction, scattering and interference, and we present a novel source placement algorithm for efficient computation. Our algorithm exploits the symmetric properties of the scattering field and the object shape to compute diffraction kernels at only a few incoming directions and accelerates the precomputation by $1 - 2$ orders of magnitude (Section 3) for efficient desktop computation.

We present a new coupling algorithm that integrates these diffraction kernels with interactive ray tracing at runtime. Our modified ray tracing algorithm uses the object-based diffraction kernels to approximate the wave effects such as diffraction and combines them with standard geometric ray tracing techniques to compute reflections at interactive rates (Section 4).

We demonstrate the interactive performance on many dynamic scenes with smooth, highly tessellated objects undergoing rigid motion (Section 5). We highlight improved accuracy as compared to prior interactive, geometric methods for capturing diffraction effects and also compare the performance with wave-based solvers (Section 6). We also perform a perceptual evaluation using a user study to compare the auditory perception of our algorithm with a wave-based propagation algorithm (Section 7).

4.2 Related Work

In this section, we give a brief overview of prior work on sound propagation.

4.2.1 Wave-Based Methods

These methods are the most accurate way of simulating sound propagation as they solve the acoustic wave equation directly. Some of the frequency domain solvers include methods based on the finite-element method (FEM), boundary-element method (BEM), and the time domain solvers include methods such as finite-difference time domain (FDTD) and adaptive rectangular decomposition (ARD). However, their space and time complexity increases as a third or fourth power of frequencies. Many interactive propagation techniques have been proposed for static scenes that precompute an acoustic kernels and use them to compute the impulse responses at runtime as a function of the source or listener positions. Equivalent source method based techniques have been used to precompute the acoustic radiation characteristics of rigid objects (James et al., 2006a) or the per-object and inter-object transfer functions for sound propagation (Mehra et al., 2013, 2015) and can also be combined with ray tracing algorithms (Yeh et al., 2013). However, the computational overhead of these methods is very high and they require large compute clusters for pre-computations. Furthermore, none of these methods can handle dynamic objects in the scene owing to need to recompute the total field if the objects in the scene move (i.e., the inter-object transfer function computation would change), thereby limiting their application to static scenes. (Mehra et al., 2015) can handle moving sources and listeners but not moving objects while (Mehra et al., 2013) can handle either a moving source or a moving listener. In contrast, our method has a lower computational overhead and can handle dynamic source, listeners, and objects, though our accuracy is slightly lower. Raghuvanshi et al. (Raghuvanshi et al., 2010, 2016) use the adaptive rectangular decomposition method to precompute acoustic responses on a sampled spatial grid.

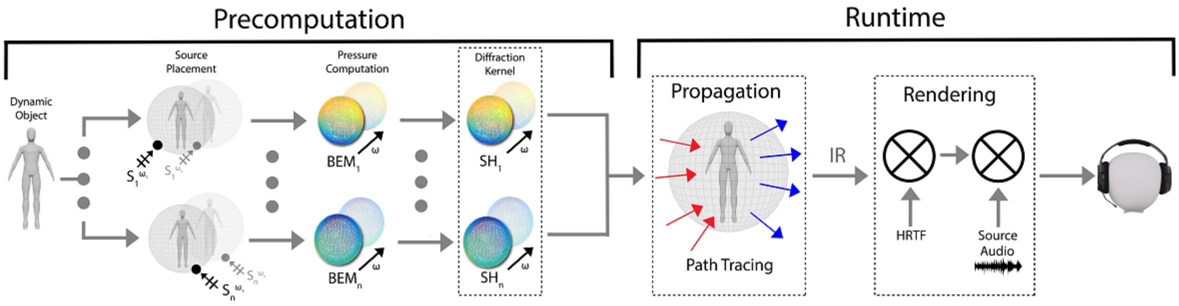

Figure 4.1: **Interactive Sound Propagation and Rendering:** We highlight different stages of our novel sound propagation and rendering pipeline, which uses per-object diffraction kernels. In the precomputation stage, we adaptively perform BEM simulations for certain directions (computed using our novel source placement algorithm) and measure the outgoing pressure fields produced by the scattering of plane waves at various frequencies. These pressure fields encode the scattering as a function of frequency, the input and output directions and converted into an efficient spherical harmonic representation called the diffraction kernel. At runtime, the diffraction kernel is coupled with an interactive path tracing algorithm to simulate sound propagation and auralization in dynamic scenes.

4.2.2 Geometric Acoustics and Diffraction

Geometric techniques model the acoustic effects based on ray theory and typically work well for high-frequency sounds to model specular and diffuse reflections (Savioja and Svensson, 2015). Wave phenomena such as diffraction must be modeled explicitly or separately and prior methods are limited to edge diffraction. The Biot-Tolstoy-Medwin (BTM) model is an accurate time-domain diffraction formulation that evaluates an integral of diffracted sound along finite rigid edges, can be extended to higher-order diffraction, and can be combined with wave-based methods (Svensson et al., 1999; Román et al., 2016). However, it is expensive to evaluate for complex scenes and limited to offline computations. An alternative approach, the uniform theory of diffraction (UTD), is a less accurate frequency-domain model of diffraction for infinite edges that can generate plausible results for interactive simulation in certain scenarios (Tsingos et al., 2001c; Taylor et al., 2012). The complexity of these edge-based diffraction techniques can increase exponentially with the maximum diffraction order, since each edge in the scene can interact with every other edge. To reduce the cost of visibility testing for high-order UTD diffraction, a precomputed edge-to-edge visibility graph can be used for static scenes, but current interactive systems are limited to low orders of edge-diffraction (Schissler et al., 2014b). However, it is not clear whether techniques based on UTD can handle complex (highly tesselated) models that are frequently used in gaming and VR due to the high number of potential diffraction edges (Tsingos et al., 2007). An accurate sound particle model of edge diffraction based on the Heisenberg uncertainty principle has been proposed for high-order diffraction (Stephenson, 2010), but is not robust for complex objects.

4.2.3 Hybrid Methods

Given the relative benefits of wave-based and geometric methods, hybrid techniques have been proposed to combine them. These include methods based on spectral decomposition of the low frequencies (i.e., less than 1kHz or 2kHz) are modeled using wave-based solvers, such as FDTD or FEM, and the high frequencies are modeled using ray tracing or beam tracing (Lokki et al., 2011; Southern et al., 2011; Granier et al., 1996). The computational complexity of these hybrid approaches is dominated by the wave-based methods that are performed over the entire acoustic domain. Another set of hybrid algorithms performs a spatial decomposition of the simulation domain into near-object regions and far-field regions for precomputation (Yeh et al., 2013). It uses an equivalent source formulation to compute the per-object and inter-object transfer functions, and combines that with a geometric ray tracing method to handle higher frequencies. However, the computation of per-object and inter-object transfer is expensive and also requires a large compute cluster for precomputation. As with (Mehra et al., 2013, 2015), moving objects in the scene would require recomputing the inter-object transfer functions making this method limited to static scenes with either a moving source or a moving listener. Furthermore, different coupling techniques have been proposed to combine the results at the interfaces, based on BEM (Hampel et al., 2008), FDTD (Wang et al., 2000), FEM (Barbone et al., 1998), and ESM (Yeh et al., 2013). However, none of these approaches can handle dynamic scenes at interactive rates.

4.3 Acoustic Field & Diffraction Kernel

In this section, we present a high-level overview of our sound propagation algorithm based on diffraction kernels. Figure 4.1 shows the overall pipeline of our approach divided into two distinct stages: precomputation to compute the diffraction kernels and runtime based on interactive ray tracing.

Table 4.1 gives a list of all the symbols used.

Symbols	Meaning
$\mathbf{x}$	Incident direction
$\mathbf{y}$	Outgoing direction
ω	Frequency
$d(\mathbf{y}, \omega)$	Scattered pressure field
$\tilde{d}(\mathbf{y}, \omega)$	SH representation of $d(\mathbf{y})$
$D(\mathbf{x}, \mathbf{y}, \omega)$	Diffraction kernel
$P(\mathbf{y})$	Probability density function
I_i	Incident sound intensity
I_o	Outgoing sound intensity
A^{proj}	Projected area
H	Visible-curvature histogram
SS	Shape signature

Table 4.1: A table of important mathematical symbols used in the text.

4.3.1 Acoustic wave equation

The acoustic wave-equation models the scattering behavior of objects. In spherical coordinates the wave-equation can be expressed as:

$$\nabla^2 p = \underbrace{\frac{\partial^2 p}{\partial r^2} + \frac{2}{r}\frac{\partial p}{\partial r}}_{\text{radial part}} + \underbrace{\frac{1}{r^2 sin\theta}\frac{\partial}{\partial \theta}(sin\theta \frac{\partial p}{\partial \theta}) + \frac{1}{r^2 sin^2\theta}\frac{\partial^2 p}{\partial \phi^2}}_{\text{angular part}} \tag{4.1}$$

where p is the pressure and (r, θ, ϕ) correspond to $\mathbf{x}$ in spherical coordinates. The complete solution to the equation above can be expressed in terms of the radial and angular parts:

$$\psi_{lm}(\mathbf{x}, \mathbf{y}) = \Gamma_{lm} \underbrace{h_l^2(kr)}_{\text{radial}} \underbrace{Y_l^m(\mathbf{x} - \mathbf{y})}_{\text{angular}} \tag{4.2}$$

The angular part of the solution is described using spherical harmonics Y_l^m while variation in the pressure because of distance is controlled by the Hankel function h_l^2.

4.3.2 Diffraction Kernels

We use a diffraction kernel representation to capture the angular portion of the solution while the radial variation of pressure is approximated by a geometric sound propagation technique. We consider a spherical grid of incoming directions and generate plane-waves from each direction of this grid. For each plane

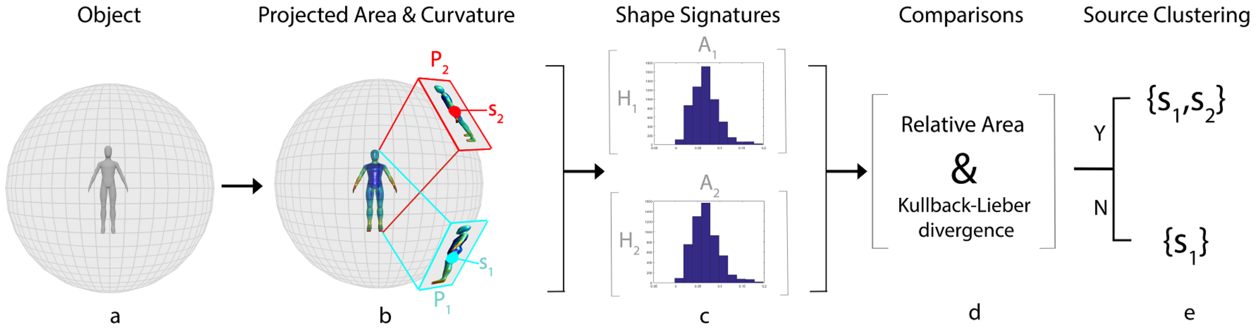

Figure 4.2: **Overview of our source placement algorithm:** We use a novel source placement algorithm to compute the representative source positions for each object: (a) Given a scatterer (human), we consider a densely sampled sphere around it; (b) For each point s_i on the sphere the projected area and viewed curvatures are computed; (c) The curvature values are binned into histograms H_i and together with the projected area A_i give a shape signature SS_i at s_i. (d) These shape signatures are used to compute geometric similarity between different viewpoints. (e) Points are grouped together if their shape signatures are within error thresholds ϵ_A and ϵ_H. The overall algorithm results in $1-2$ order of magnitude improvement in the precomputation stage.

wave, we compute the scattered field for the object on an offset surface of the object using a wave-based method. The angular portion of this scattered field is expressed using the *diffraction kernel* in a compact spherical harmonic basis. With the angular scattering behavior of an object computed for all the plane wave directions and frequencies, we use a geometric sound propagation method to handle the radial portion thus approximating the solution to the wave-equation.

Our diffraction kernel encapsulates the sound field interactions of the object and maps the incoming sound field reaching the object to outgoing, diffracted field emanating from the object. In contrast to the per-object transfer function (Mehra et al., 2013), the diffraction kernel formulation is defined in the far-field of the object. This can significantly reduce the precomputation overhead and makes it easier to integrate with interactive ray tracing. Mathematically, the incoming pressure field in the far-field can be expressed in the plane-wave basis whereas the outgoing sound field in the far-field is expressed using spherical harmonic basis, as shown in Equation 4.1.

Scene Classification Our approach is based on computing the diffraction kernel for each objects in the scene. As a preprocess, we classify the scene in terms of the object type. The scene is first classified into *static* and *dynamic (moving)* objects. Static objects typically include walls, buildings, and other typically large, immovable objects in the scene. Dynamic objects can include cars, humans, chairs, and doors, all of which can potentially undergo rigid motion in the environment. Our approach is designed to capture the scattering behavior of dynamic objects, while the static environment is handled by other sound propagation techniques.

4.3.3 Source Placement

Diffraction is a direction dependent phenomenon and in order to capture the variations in the diffracted field, we need to capture the sound interaction behavior of an object from all possible directions. This can be naïvely computed by constructing a densely sampled sphere around the object and evaluating the diffracted field for each vertex on the sphere. However, such a method would incur a large precomputation cost because the wave-based solvers typically used to compute the scattering are slow and the complexity increases as a function of the geometric tessellation and maximum frequency. In order to reduce the precomputation overhead, we present a novel source placement algorithm that exploits the *acoustic scattering invariance* of a dynamic object to reduce the number of sources we need to place on the sphere to capture its scattering behavior from all incident angles. Our source placement algorithm is used in the first stage of our precomputation pipeline and computes the representative source positions, as shown in Figure 4.1.

4.3.3.1 Visual Symmetry vs. Scattering Field Symmetry

The goal of our source placement tends to exploit the symmetry in the acoustic scattering field of each object, and thereby compute a few representative source positions. One possibility is to exploit the visual or shape symmetry of each object. There is extensive work on symmetry detection in computer vision and geometry processing (Mitra et al., 2006), which are used to compute a representation of their Euclidean symmetries. However, the criteria used in these methods are not sufficient for detecting the symmetry in the acoustic scattering field of an object. For example, there are objects that exhibit little or no shape symmetry, but still exhibit symmetry in their acoustic scattering field. As a result, our goal is to develop an approach that generalizes the notion of shape-similarity and is not sensitive to the small variations in the viewed-geometry. One of the metrics in our source placement algorithm is to use projected areas to overcome these issues.

4.3.3.2 Multi-stage Algorithm

Next, we describe various stages of our source placement algorithm including mesh simplification, computing the projected area for each incident direction and identifying the shape and diffraction field from each direction. We compare the view-dependent shape information to compute the geometric invariance among various incident directions and clump them together.

Projected Area: The diffraction field is a strong function of the shape and orientation of the object. In particular, for convex objects it has been shown that diffraction is a function of the projected area of the object (Chinnery et al., 1997; Vickers, 1996). Formally,

$$P_{sc}(\mathbf{x}) = KA^{proj}(\mathbf{x})P_{in}, \tag{4.3}$$

where $A^{proj}(\mathbf{x})$ is the projected area at $\mathbf{x}$, P_{in} is the incident field, P_{sc} is the scattered field (or diffraction field), and K is a constant. We exploit this dependence of the scattered field on the projected area and extend it to arbitrary or non-convex objects by augmenting the projected area with curvature histograms (described below) to uniquely identify the *shape signature*.

Shape Signature: Our source placement algorithm initially considers a densely-sampled list of possible source positions S on a sphere. For a point $s_i \in S, \forall i \in (1..|S|)$, we compute an orthographic projection matrix $P(\mathbf{x}_i)$ and compute the projection of the vertices of the object ($v \in V$), whose normals N_j^v satisfy the $N_j^v \cdot \mathbf{s_i} > 0$. Next, we construct a boundary B_i using the $alpha$ or $\alpha-$shape of v and compute the area enclosed by the boundary A_i^{proj}. $\alpha-$ shape is the generalization of the notion of a convex-hull of a point set M, with $\alpha \rightarrow 0$ gives us M, while $\alpha \rightarrow \infty$ giving us the convex-hull of M. At the end of this step, we have computed the projected area of the object for each point on S.

In practice, the projected area alone cannot be used as a unique signature of the viewed shape and may result in false positives, in terms of classifying rather different shapes as similar. Therefore, we augment our metric by using the curvature of the object to define the *shape signature* of the object for each s_i. This view-dependent shape signature encapsulates the intrinsic characteristics of the shape when viewed from different source points (or incident angles). We use well-known techniques (Cohen-Steiner and Morvan, 2003) to compute the principal curvatures κ_1 and κ_2 for the scatterer, and compute them for each v. Instead of using κ_1 & κ_2 separately, we consider them as $\kappa_{v_j} = |\kappa_{1_{v_j}}| + |\kappa_{2_{v_j}}|$ and bin them in a histogram H_i that uses N bins. The bin values range between the minimum and maximum values of $|\kappa_1| + |\kappa_2|$. Using the projected area and curvature, we get a shape signature ($\mathbf{SS_i}$) for each s_i:

$$\mathbf{SS_i} = \begin{pmatrix} A_i^{proj} \\ \\ H_i \end{pmatrix} \tag{4.4}$$

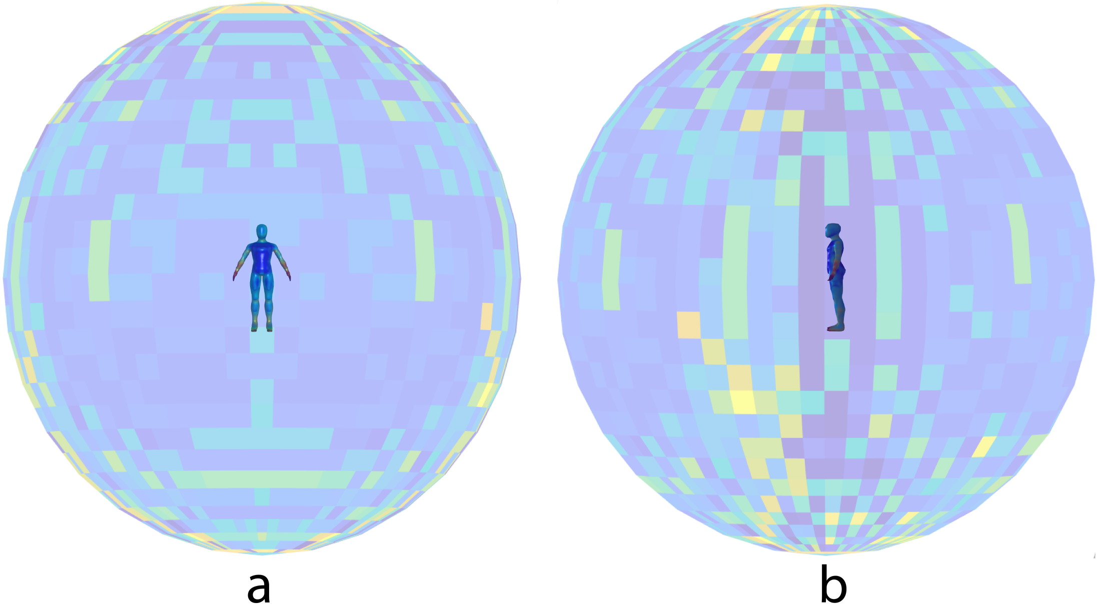

Figure 4.3: **Similar Shape Signatures:** We highlight different points that have similar shape signatures. Each set of points with the same color on the sphere corresponds to a set that is computed as geometrically invariant and will be represented using a single sound source. (a) The source placement automatically detects symmetry in the model which is bilateral in this case; (b) shows another viewpoint and the since no symmetry exists in that plane, the two hemispheres have different colors.

Rotational symmetry: After computing the shape signature, we iterate over the points in S. Starting with a point s_i, we compare its shape signature SS_i with every other point's shape signature SS_j by computing the relative difference in the projected area A_i^{proj} and A_j^{proj}; and we also compute the difference in histogram H_i and H_j using the Kullback-Lieber divergence:

$$D_{KL}(H_i||H_j) = \sum_{k}^{N} H_i(k) log(\frac{H_i(k)}{H_j(k)}) \tag{4.5}$$

This metric gives us a measure of the mutual information contained in two shapes. A D_{KL} value of zero indicates that the shapes are similar and would likely have similar scattering properties. On the other hand, a value of one would indicate that the shapes are very dis-similar. Using appropriate thresholds for the relative projected areas and D_{KL}, we cluster the points that fall within the threshold bounds with respect to s_i. The threshold values are used to strike a balance between the number of sources that are selected from S and the error in computing the total scattering function of an object. Finally, we choose one representative point in each cluster and use that point as the source position for which the scattering function is computed

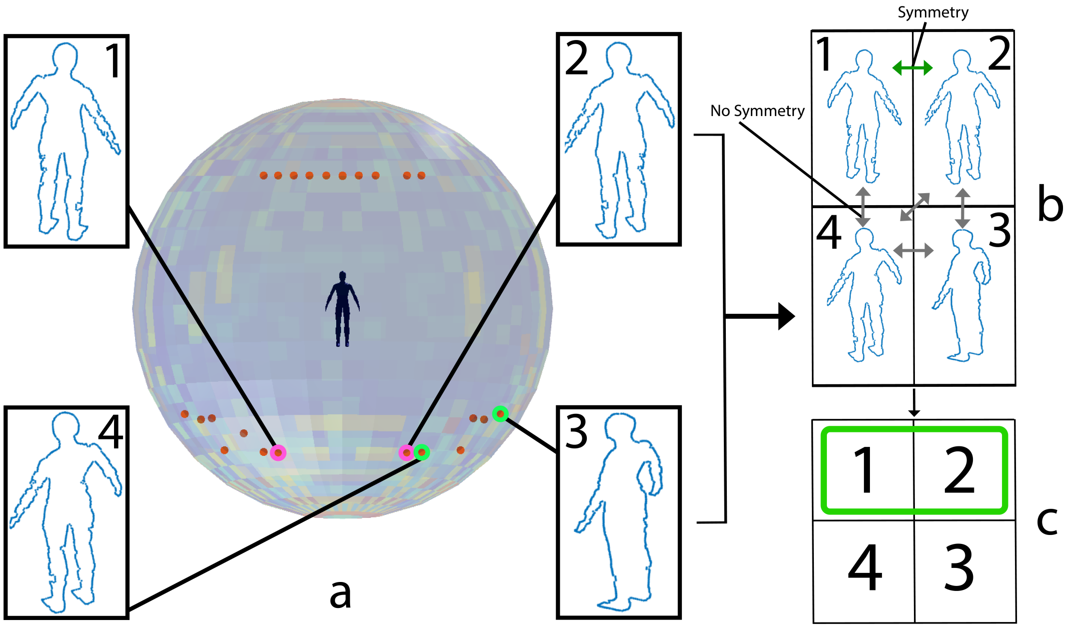

Figure 4.4: **Reflection symmetry detection**: Given a cluster of points with similar shape signatures (a) we perform a pair-wise comparison of the boundaries and compute Hausdorff distance; (b) Boundaries B_1 and B_2 that nearly overlap after being reflected result in a large drop of their Hausdorff distances, while B_3 and B_4 do not exhibit reflection symmetry with each other or with B_1 or B_2' (c) A relative change in the distance indicates reflection symmetry between (B_1 and B_2).

using a wave-based solver. The scattering functions for other points in the cluster are extrapolated from this representative point.

Reflection Symmetry: Many objects used in the real world exhibit reflective symmetry (e.g., a pillar). Although our algorithm can recognize some sort of symmetry in the object, it cannot identify the nature of that symmetry. The previous steps detect the invariance in geometry which includes rotational symmetry along with insignificant changes in shape with the change in incident angle.

In order to explicitly identify the reflection symmetries in a cluster, we perform a pair-wise comparison between the points in the cluster (Fig. 4.4). For each such pair of points, we compute the silhouette of the object from these points and compute the Hausdorff distance between these boundaries. The Hausdorff distance (d_H) between two non-empty sets (X, Y) that are subsets of a metric space (M, d) is given by:

$$d_H(X, Y) = \underset{x \in X}{max}\{\underset{y \in Y}{min}\{d(X, Y)\}\}, \tag{4.6}$$

where $d(X, Y)$ is some measure of distance in M (L_2 metric in our case). We reflect these 2D boundaries either along X or Y axis depending on the source positions and compare their Hausdorff distances. If the relative Hausdorff distance is below our threshold, we consider these boundaries as reflections of each other. (Fig. 4.4(b,c)). In case an object exhibits both rotational and reflection symmetry at the same point, our method automatically considers them to be rotationally symmetric.

4.3.4 Diffraction Kernel Computation

After the reflection symmetry test, we compute the set $RP(S) = \{\{A_1^{Proj}, H_1\}, \{A_2^{Proj}, H_2\}, \ldots, \{A_n^{Proj}, H_n\}\}$, where each element of RP is the set of points with a similar projected area (A_i^{proj}) and curvatures (H_i). In terms of diffraction kernel computation, we perform a single wave-based simulation for such a set, as explained below. Overall, our algorithm performs $O(n)$ wave simulations, where $n = |RP|$ with $n \leq |S|$, to capture the diffracted field from all incident directions. In practice, n is orders of magnitude smaller than $|S|$ (Table 2). After these n simulations, we extrapolate the field within a particular set by rotating and/or reflecting the computed field, thereby giving us the complete diffracted field for all $s_i \in S$.

For an incoming plane wave coming from direction **x**, the outgoing sound field $d(\mathbf{y}, \omega)$ is computed using state-of-the-art wave-based methods (e.g. BEM) on a spherical offset surface in far-field. This outgoing field $d(\mathbf{y}, \omega)$ can be expressed in the spherical harmonic basis using least-squares fitting:

$$d(\mathbf{y}, \omega) \approx \tilde{d}(\mathbf{y}, \omega) = \sum_{l=0}^{l_{max}} \sum_{m=-l}^{l} Y_l^m(\mathbf{y}) c_l^m(\omega) \tag{4.7}$$

where $d(\mathbf{y}, \omega)$ is the outgoing sound field computed using the wave-based solver, l_{max} is the spherical harmonic order, and $c_l^m(\omega)$ are the basis function coefficients as a function of frequency. This process is repeated for all the incoming plane wave directions for all the frequencies. We use a rectangular subdivision in spherical coordinates to compute the possible incoming plane wave directions. This enables efficient bi-linear interpolation of the outgoing field for any arbitrary incoming direction during the runtime stage of our pipeline.

4.4 Interactive Ray Tracing with Diffraction Kernels

In this section, we present our diffraction kernel-based technique for object-based sound propagation in dynamic scenes. We utilize a precomputed *diffraction kernel* to model sound interactions for complex objects and couple it with a Monte Carlo path tracing framework to compute sound propagation for the entire scene.

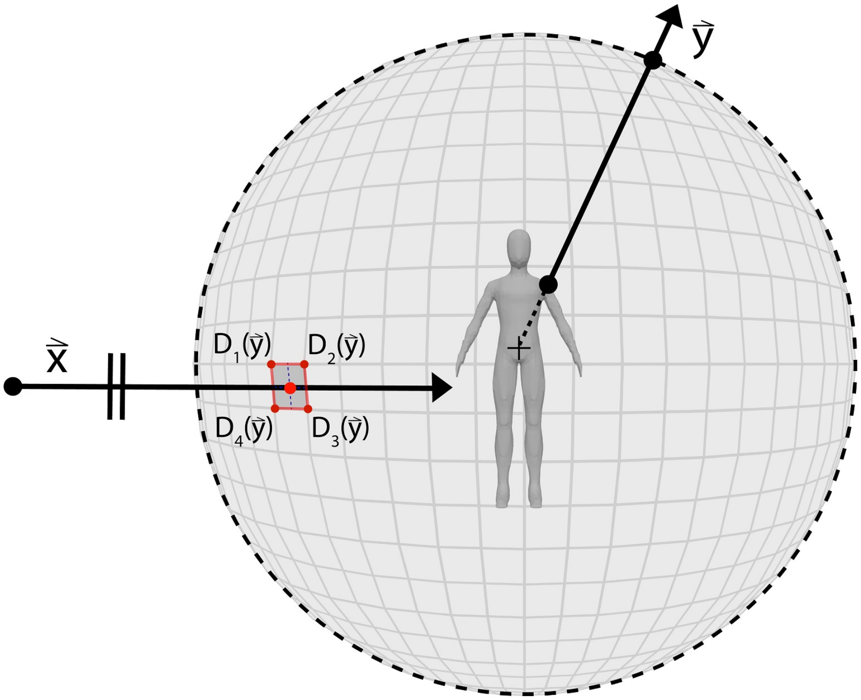

Figure 4.5: We highlight how the diffraction kernel $D(\mathbf{x}, \mathbf{y}, \omega)$ can be integrated into Monte Carlo path tracing using two-way coupling. When an incoming ray with direction **x** strikes a diffracting object, the ray is scattered in a randomly chosen direction **y** with probability density function $P(\mathbf{y})$. The diffraction kernel in the direction **y** is evaluated at the four corners of the quad intersected by **x**, and the resulting pressures $D(\mathbf{x}_1, \mathbf{y}, \omega)$, $D(\mathbf{x}_2, \mathbf{y}, \omega)$, $D(\mathbf{x}_3, \mathbf{y}, \omega)$, and $D(\mathbf{x}_4, \mathbf{y}, \omega)$ are bilinearly interpolated according to **x** to yield the pressure transfer function $D(\mathbf{x}, \mathbf{y}, \omega)$. The energy carried by the ray is then multiplied by $\frac{D(\mathbf{x},\mathbf{y},\omega)^2}{P(\mathbf{y})}$ to get the output ray energy.

For the simulation of diffuse reflections, many variants of Monte Carlo path tracing have been proposed that simulate the propagation of sound energy by rays in frequency bands (Krokstad et al., 1968b). These include backward ray tracing for multisource scenes (Schissler and Manocha, 2016), and bidirectional path tracing (Cao et al., 2016), which can also be accelerated by exploiting temporal coherence. Our approach extends these methods by developing new interactive techniques for two coupling between the rays and diffraction kernels.

The interactive ray tracing uses a bounding volume hierarchy to accelerate ray intersections. These hierarchies are updated using refitting algorithms, as the dynamic objects undergo rigid motion.

After the diffraction kernel D of a particular object is computed according to Section 4, it can be used within any Monte Carlo path tracing sound propagation algorithm to efficiently compute diffracted sound for the object. This kernel information is stored in the bounding volume hierarchy nodes associated with those dynamic objects. Our formulation treats the diffraction kernel using a mathematical framework similar to surface scattering modeled using bidirectional scattering distribution functions (BSDF), which is widely used in visual rendering. BSDFs describe the distribution of sound energy as a function of frequency and the input and output direction of sound transport (Dindart et al., 1999). We use the diffraction kernel in a similar way to model the wave scattering induced by objects in all directions. Our modified path tracing algorithm uses the diffraction kernel information to compute the new paths using D for each ray, after it hits a dynamic object.

4.4.1 Coupling between ray and diffraction kernels

Our propagation algorithm exploits a two-way coupling between D that are computed using BEM (i.e., wave-based method) and path tracing (i.e., geometric acoustics). For the case of a single ray with input sound intensity I_i and direction **x**, the outgoing sound intensity I_o is given by a spherical integral of the diffraction kernel over the outgoing direction **y**:

$$I_o(\mathbf{x}, \omega) = \int_S I_i D(\mathbf{x}, \mathbf{y}, \omega)^2 dS. \tag{4.8}$$

Monte Carlo techniques are a simple way to numerically evaluate integrals of this form as a weighted sum of many random samples (Krokstad et al., 1968b). As the number of samples approaches ∞, the expected value of the integral converges to the exact value. The outgoing scattered intensity can be approximated by a Monte Carlo estimator:

$$I_o(\mathbf{x}, \omega) \approx \frac{1}{N} \sum_{j=1}^{N} I_i \frac{D(\mathbf{x}, \mathbf{y}_j, \omega)^2}{P(\mathbf{y}_j)} \tag{4.9}$$

where N is the number of samples, $\mathbf{y}_j$ are the samples, and $P(\mathbf{y}_j)$ is the probability of generating sample $\mathbf{y}_j$. If a uniform sampling strategy is used, $P(\mathbf{y}_j) = \frac{1}{4\pi}$. This formulation can be easily integrated with any Monte Carlo ray tracer to compute object-based scattering. We utilize this formulation to model the diffraction effects and approximate the sound field in the regions that are occluded from each source.

In traditional forward path tracing, N random rays are emitted from the surface of a sound source in the scene with energy $\frac{1}{N}$. These rays are then propagated through the environment until they strike a surface, where the rays are scattered and attenuated according to the sound material BSDF. The rays may undergo many interactions with the geometry before either exiting the scene, reaching a maximum interaction order or propagation time (Schissler and Manocha, 2016), or being eliminated via Russian Roulette (Kapralos et al., 2005). If a ray hits the listener, the ray's intensity at various frequency bands is accumulated to the impulse response at the appropriate delay time. At each interaction, a shadow ray can also be traced to the listener's position to find additional propagation paths. This is known as next-event estimation or *diffuse rain* (Schröder, 2011). This procedure can also be conducted in reverse by emitting rays from the listener (Schissler and Manocha, 2016), or by emitting rays from both source and listener (Cao et al., 2016). Figure 4.5 demonstrates how the diffraction kernel can be integrated into this path tracing framework. When a ray hits an object in the scene that has an associated precomputed diffraction kernel, we scatter the ray using the precomputed scattering function rather than the usual BSDF. This is performed by randomly sampling the outgoing ray direction **y** according to probability density function $P(\mathbf{y})$. The diffraction kernel is evaluated at **y** for the four precomputed scattering functions that are closest to the incident ray direction **x**, then the evaluated pressure is bilinearly interpolated. The energy carried by the outgoing ray is then given by:

$$I_o(\mathbf{x}, \omega) \approx I_i \frac{D(\mathbf{x}, \mathbf{y}, \omega)^2}{P(\mathbf{y})}. \tag{4.10}$$

When many rays hit the scattering object, the integral of the outgoing energy over all rays converges to the exact solution.

4.5 Implementation & Results

In this section, we discuss our implementation and highlight the results on complex benchmarks.

4.5.1 Performance and Comparisons

The preprocessing algorithm has been implemented using MATLAB. We used available MATLAB code for computing the curvatures of our objects. FastBEM is used as the boundary element method solver. The runtime interactive ray-tracer is based on the geometric sound propagation algorithm described in (Schissler et al., 2014b) and written in C++. We do not use the original UTD-based method proposed in (Schissler et al., 2014b), and rather use the coupled algorithm described in Section 4.2.1 for path tracing with diffraction kernels.

Precomputation: Table 4.2 gives the geometric details of the objects used in our scenes and the performance of our source placement algorithm. Since BEM computation can be expensive and increases as cubic function of the frequency, our novel source placement algorithm makes it possible to handle complex, smooth objects with thousands of triangles. We observe $8 - 137$X speedups due to our source placement algorithm. That enables us to perform the diffraction kernel precomputation on a desktop PC, as opposed to using a large compute cluster. Most prior wave-based methods (Mehra et al., 2013; Yeh et al., 2013; Raghuvanshi et al., 2010) have significantly higher memory and computational requirements.

Runtime System: Our interactive sound propagation algorithm has been integrated with the Unreal Engine and used to evaluate the performance of complex, dynamic benchmarks shown in Fig. 1. All the timings were generated on a multi-core desktop PC CPU. The overall system with integrated visual and sound rendering runs at 60Hz or more, as shown in the video. The additional overhead of handling diffraction kernels is very small and the overall performance is comparable to UTD-based interactive propagation algorithms (Tsingos et al., 2001c; Schissler et al., 2014b).

4.5.2 Benchmarks

We have evaluated our approach on various scenarios to highlight the performance of our diffraction kernels in challenging environments (Table 4.3). To accentuate the effect of diffraction kernels, we turn off reflections in each of our benchmarks.

Concert: The concert scene shows the effectiveness of our diffraction kernels in handling complex diffracting objects such as humans in an open environment. Each human model is represented using $11K$ triangles and prior interactive UTD-based methods can't handle such scenes for plausible diffraction effects. The listener moves among a crowd of people attending a concert and ducks to pick up a dropped phone. The complex interactions of the sound source and human bodies are efficiently and plausibly calculated using diffraction kernels.

City Block: This scene shows the listener moving through a modern metropolis with various high-rise buildings. A helicopter flying over the city goes behind one of the high-rises (cylindrical) causing the sound to diffract around highly-tessellated objects. This scene demonstrates the ability of our method to handle highly-tessellated, curved objects and generate a smooth diffraction field around them. This results in smooth audio rendering.

Parking Garage: This benchmark consists of a typical parking garage with multiple pillars and cars. We use the pillars in the garage and a moving ambulance as the diffracting objects. The listener moves through the garage experiencing diffraction effects as the pillars obstruct the line-of-sight between the listener and various sources. Then an ambulance comes into the garage to park and acts as a dynamic diffraction object.

Object	**#Vert.**	**Size(m)**	**freq(Hz)**	**|RP|**	**Speedup**
Ellipsoid	10242	2	1000	50	38X
Ambulance	21746	3.9	1000	150	13X
Human	11250	1.8	2000	254	8X
Column	29954	4.7	1000	118	16X
Tower	44168	15	500	14	137X
Ball	2562	0.5	2000	1	1922X
Monitor	3650	0.46	2000	99	19X
Robot	23971	0.44	1000	229	8X
Pillar	25746	3	500	221	8X
Planter	11114	2.78	500	323	6X

Table 4.2: **Diffraction Kernel Computations**: The table highlights the geometric complexity, size of objects (meter), maximum frequency, running times for computing the diffraction kernel of different objects. The value of $|\mathbf{S}|$ is 1922 in all the benchmarks. The speedups obtained using our source placement algorithm are highlighted in the last column.

Oculus® First Contact: This benchmark is a modified version of the famous First Contact demonstration that is being shipped by Oculus ®, along with their HMD. In this scenario, a playful robot acts as the object and comes in between the sound source and the listener and creates diffraction effects dynamically due to no line-of-sight. The 3D printer in the scene generates an interactive object that also results in diffraction effects along with a static monitor. Our approach can model the diffraction effects due to these dynamic objects and generate smooth audio rendering effects. We highlight these benefits in the video, by only playing the diffracted sound with no reflections.

Multi-player Game: We showcase the efficiency of our approach in this multi-player networked game. In this scenario, two players play against each other in a networked environment and are trying to shoot at each other. As the players move around, the sound gets diffracted around different objects in the scene. As a result, simulating object-based diffraction is important to simulate a continuous sound field. We highlight these benefits in the video, by only playing the diffracted sound with no reflections.

4.6 Analysis

In this section, we analyze the various steps of our pipeline and highlight the approximations and possible sources of error in the computations. We also compare the accuracy of our precomputation algorithm with a wave-based solver (BEM) to evaluate the numeric accuracy of the computed sound pressure field. There are

Scene	#Vert.	#D	PreC(Hr)	Runtime(ms)
Concert	10242	11	4	53
City-Block	21746	2	1	101
Parking-Garage	11250	4	7	115
First-Contact	29954	1	2	43
Game	44168	2	6	84

Table 4.3: **Runtime Performance Analysis:** We highlight the performance of our interactive sound propagation algorithm on a desktop multi-core PC. We highlight the number of diffraction objects (D-objects), precomputation time (PreC) in hours and the average frame time (ms) on a multi-core desktop CPU. Our algorithm can perform interactive sound propagation in dynamic scenes with specular and diffuse reflections and diffraction effects.

three main sources of error in our pipeline: Error in source placement, error introduced by the band-limited diffraction kernel, and error incurred as a result of Monte-Carlo ray-tracing at runtime.

4.6.1 Source Placement

The source placement algorithm introduces errors due to simplification and metrics used to detect rotational and reflection symmetry. This error is also governed by the underlying mesh representation and the initial choice of source positions on the sphere. This could result in changes or errors in the final pressure field that is computed using BEM using those source cluster positions. This error is more at the higher frequencies, because the diffracted component of the sound field is a lot more ”focused”, as compared to that at the lower frequencies, and is sensitive to spatial variation. In our benchmarks, we limit the maximum frequency to 2kHz.

Fig. 4.6 shows the error introduced by our source placement algorithm expressed as mean absolute error (MAE) in dB. MAE is computed as:

$$MAE = \frac{\sum_{i=1}^{n} |P(i)_{computed} - P(i)_{ref}|}{n} \tag{4.11}$$

where $P_{computed}$ is the interpolated field computed by our algorithm at an incoming source direction and P_{ref} is the reference pressure at the source direction. n is the number of the points on which the scattered pressure is computed.

As can be seen in Fig. 4.6, the error introduced for complex objects such as human and robot is below 2 dB even at frequencies as high as 1 kHz.

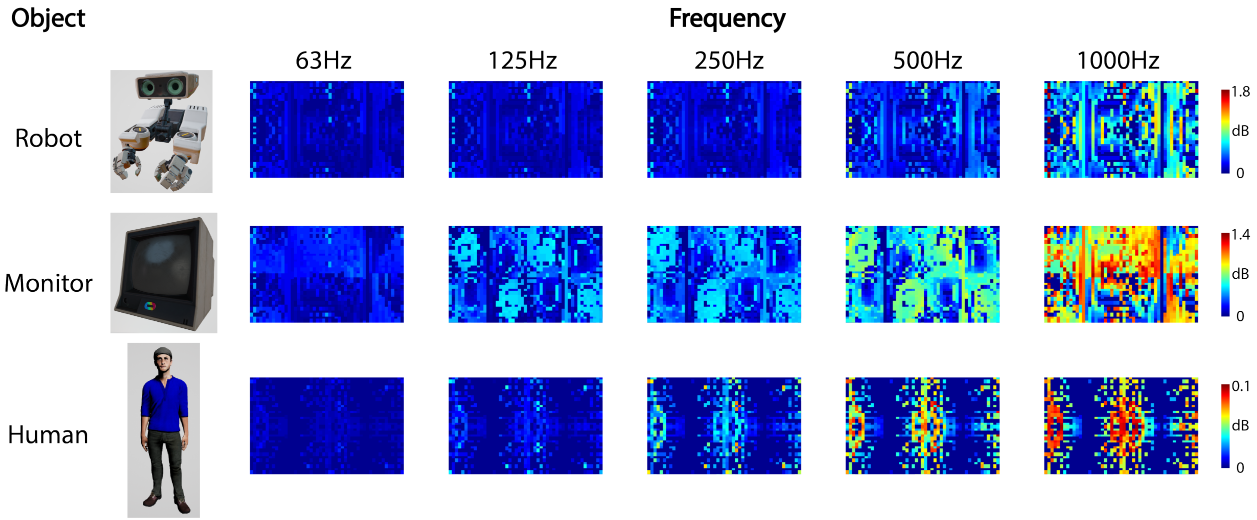

Figure 4.6: The plot shows the heat map error introduced by our source placement algorithm for three different objects. The error is computed on a sphere representing all the incoming directions for the diffraction (S). Given the source positions (RP) computed by our source placement algorithm, we run BEM at these points and interpolate the field for the rest of the points in S using reflection and/or rotation. The plots here show the MAE at each point on S by unwrapping it on to a 2D plane. The horizontal axis represents the latitude while the vertical represents the longitude. As can been seen, even for complex objects at high frequencies, the error introduced by our source placement algorithm is < 2 dB.

4.6.2 Diffraction Kernel

Diffraction Kernels represent the BEM pressure computed on a sphere based on a spherical harmonic basis (Eq. 4.7). Theoretically, spherical harmonics can fully represent a spherical function with $L_{max} \to \infty$, but in practice they have to be band-limited for practical reasons. This introduces an error given be $\epsilon_d = d(\mathbf{y}, \omega) - \tilde{d}(\mathbf{y}, \omega)$ in our diffraction kernels (Fig. 4.7). We highlight how this error increases in the diffraction kernel with increasing spherical harmonic order. In our current implementation, we use 9th order spherical harmonics and they generate plausible sound effects in our benchmarks.

4.6.3 Monte-Carlo Sampling

We show the plot (Fig. 4.8) of the pressure field generated by a densely sampling of the diffraction kernel and compare to the pressure field generated by BEM for the human models. This comparison highlights the numerical accuracy of the sound pressure that is approximated using the diffraction kernels.

We use dense ray sampling to generate this plot for the diffraction field of an object. In this case, each point on the grid is used to trace a ray backwards from that position towards the object. This ray is filtered by the diffraction kernel, depending on the angle of incidence $\mathbf{y}$ and used to compute the pressure at that

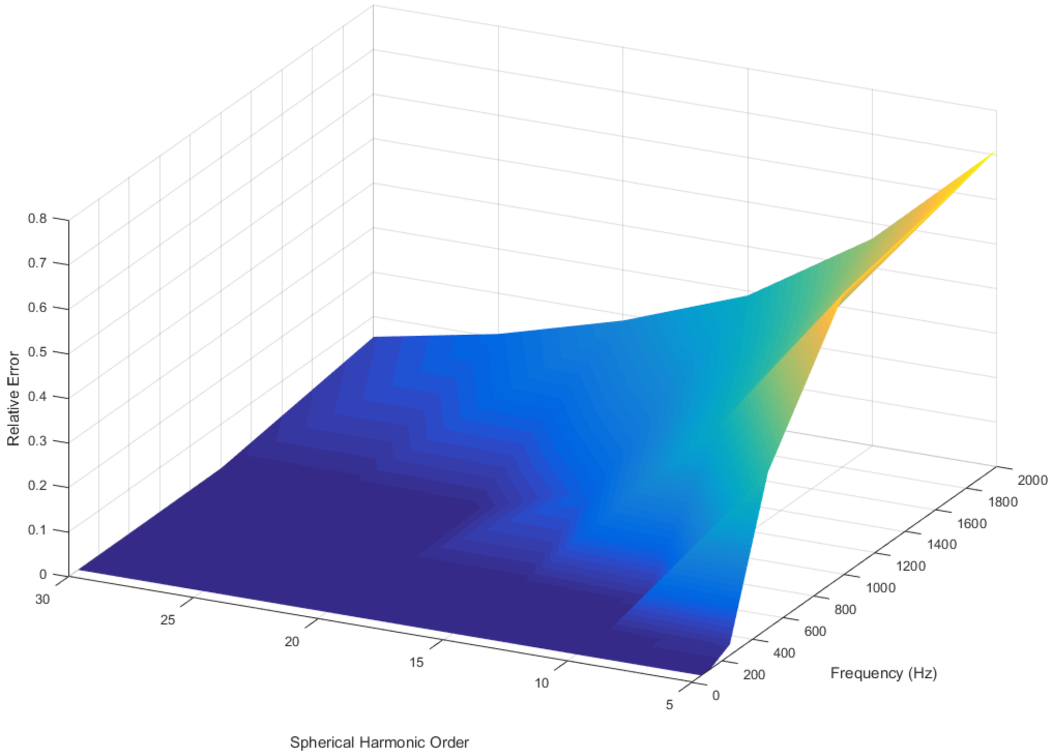

Figure 4.7: The plot shows the variation of the relative error when trying to represent a diffracted field of a human in a spherical harmonic basis. As can be seen, the error increases sharply with frequency and low SH-order but stays close to zero with high order spherical harmonic. High-order spherical harmonics are more expensive to evaluate and tend to be numerically unstable

position. This process is akin to Monte-Carlo sampling in the limit with a very high sampling density. As mentioned in Section 4.2, Monte-Carlo path tracing methods converge to the value of the sampled function as the number of samples approach infinity. As shown in the figure, the diffraction kernels converge to the BEM computed pressure field for different frequencies. This indicates the accuracy of our diffraction kernel based method is governed by the underlying sampling criterion used in path tracing.

4.7 Perceptual Evaluation

We performed a user study to evaluate the perceptual efficacy of diffraction-kernel-based sound propagation algorithm. Our study is based on the psycho-acoustic evaluation of numeric and geometric sound propagation algorithms (Rungta et al., 2016, 2017). In particular, that study compared UTD-based interactive sound propagation algorithm with a wave-based sound propagation algorithm by evaluating the diffracted

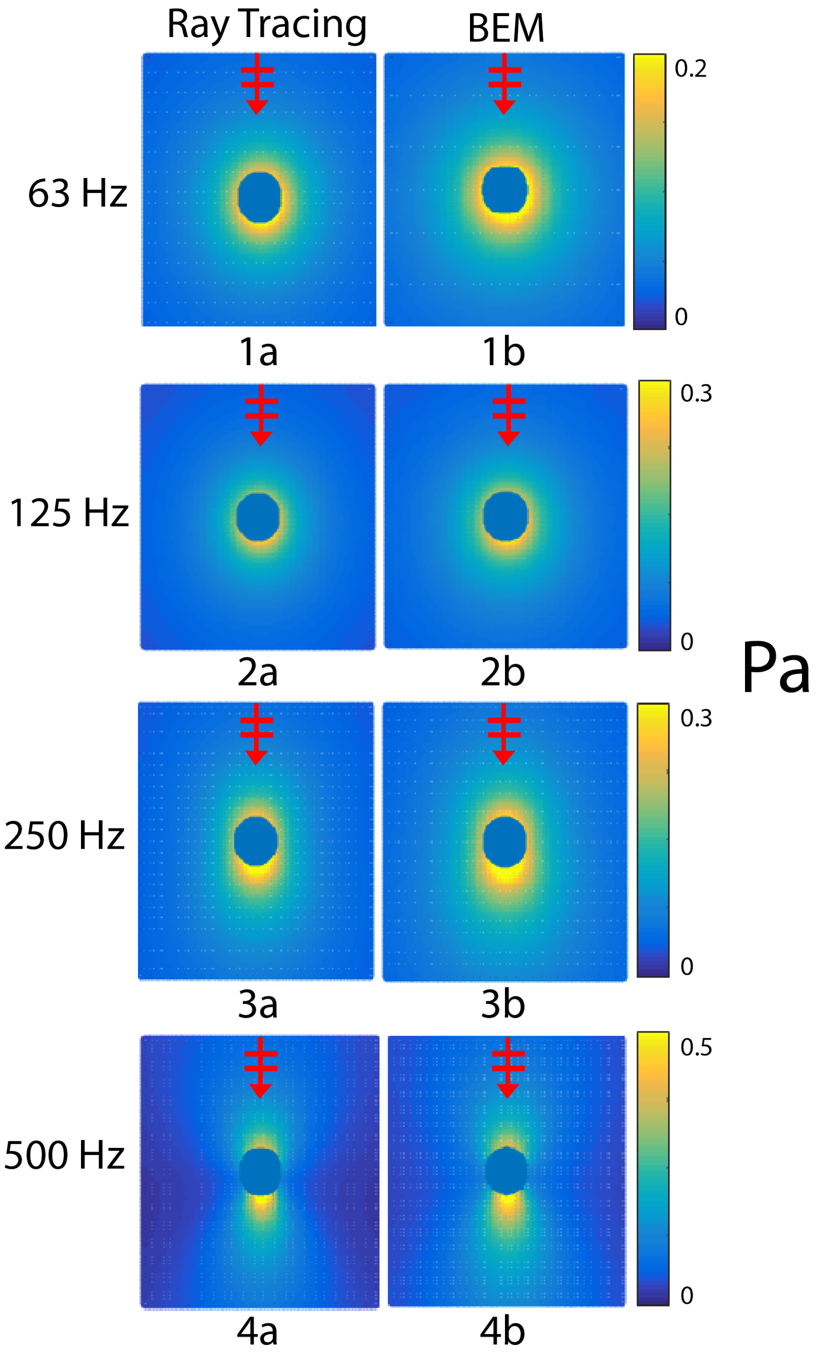

Figure 4.8: We compare the sound pressure field for an object (human) computed using our modified ray tracing algorithm (Section 4) vs. BEM (wave-based solver). We perform a dense ray sampling using the diffraction kernel to compute the pressure field in the left figures. The red arrows indicate the incident direction of the plane-wave. We use 100×100 grid to sample at each point and filter them through the diffraction kernel to compute the angular variation in the diffracted sound field. The pressure values are in Pascals and we demonstrate the results for two different frequencies, where diffraction effects are prominent. These benchmarks show a close match between the sound fields computed using our method vs. BEM. In practice, our approach can perform these computations at interactive rates, where BEM solver can take minutes.

sound field around an obstacle by placing the subjects along a semi-circle. The study (Rungta et al., 2016) demonstrated that auditory perception improves due to wave-based sound propagation and the computed diffracted field decays nearly linearly with an increasing diffraction angle. On the other hand, the diffracted field computed using UTD-based diffraction exhibited an erratic behavior. Given the known benefits of wave-based sound propagation algorithms, we perform a 2-way comparison between the diffracted sound fields computed using diffraction kernels and BEM based sound propagation.

4.7.1 Participants

Fourteen subjects participated in this study with informed consent. The ages ranged from 23 to 28 (Mean = 25.7 with SD = 3.22). The participants were recruited at a university campus. All participants reported normal hearing.

4.7.2 Apparatus

The setup consisted of a Dell T7600 workstation with the sound delivered through a pair of Beyerdynamic DT990 PRO headphones. The subjects wore a blindfold.

4.7.3 Stimuli

As in (Rungta et al., 2016), the source was a ringing bell that was low-pass filtered with a cut-off frequency of 300 Hz, so that the diffraction effects are prominent. The sound source was placed $2m$ from the origin. The subjects were placed at 5 equispaced positions along an arc with a radius $3.5m$ of from the origin as shown in Fig. 4.9. The resulting sound was prerecorded at each of these 5 positions and two diffraction methods (diffraction kernel and BEM) and stored. On each trial the subject was randomly placed at one these 5 positions with the diffraction method randomized, too.

4.7.4 Design and Procedure

This was a within-subject study with the subjects wearing a blindfold. The audio was delivered through headphones and rendered monaurally. Before starting the experiments, the source sound clip was played to familiarize the subjects with it. A $1.2m \times 1m \times 4m$ column served as the diffraction object for the experiment.

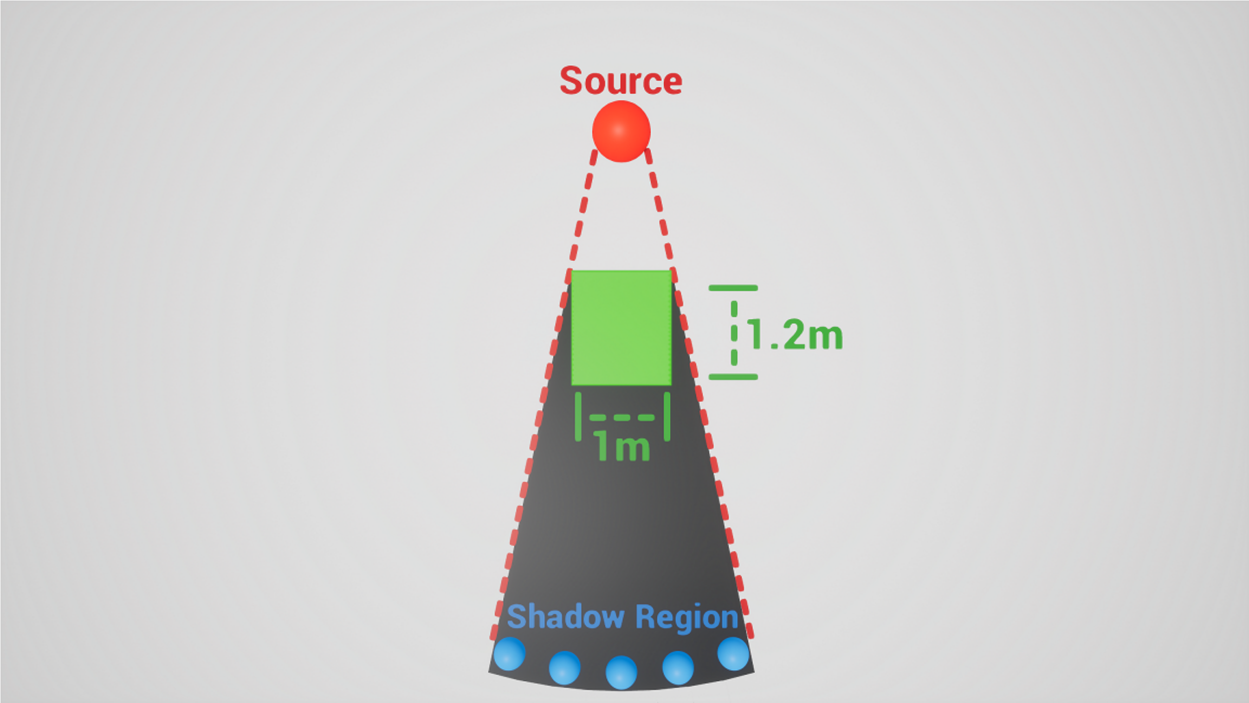

Figure 4.9: The figure shows the setup used for the user study that compared the psychoacoustic characteristic of our diffraction kernel based algorithm with BEM-based wave propagation algorithm. We considered 5 equi-spaced points in the shadow region (black) of the obstacle(green). The obstacle is a column and only the diffracted sounds are audible in the shadow region. We evaluated the auditory perception using diffraction kernel and BEM-based sound propagation.

The scene was open to make sure no reflections interfere with the experiments. The subjects were placed in the ‘shadow-zone’ of the diffracting object (Fig. 4.9) which is a region where the source is not in the line-of-sight and only the diffracted sound can reach the listener.

A total of 14 participants took part in each group. For each of the 5 positions, the subjects were asked to rate the loudness of the sound heard. The loudness was rated on an arbitrary, non-physical scale ranging from $1 - 20$. The scale was explained to the subjects before the start of the experiment: the extrema of our scale was relative to a verbal standard with 1 corresponding to a very quiet sound such as that of a falling leaf, while 20 was a loud sound akin to someone shouting in one’s ears. It should be noted that loudness perception was not the focus of our experiment; rather, the smoothness of change in perceived loudness across spatial variations as measure of the quality of each diffraction method. The loudness of the sounds for the two diffraction methods was level-matched by matching the root mean square (rms) of the sounds generated by the two methods at a reference position in the line-of-sight of the source.

A block consisted of 10 (5 positions × 2 diffraction methods) trials with three blocks per subject giving a total of 30 (5 positions × 2 diffraction methods × 3 blocks) readings. The subjects were placed randomly at one of the 5 positions with the sound played through two diffraction methods which were chosen randomly, as well. The subjects were allowed to take any many breaks as needed. Subjects took an average of $10 - 15$ minutes for the entire experiment.

4.7.5 Results

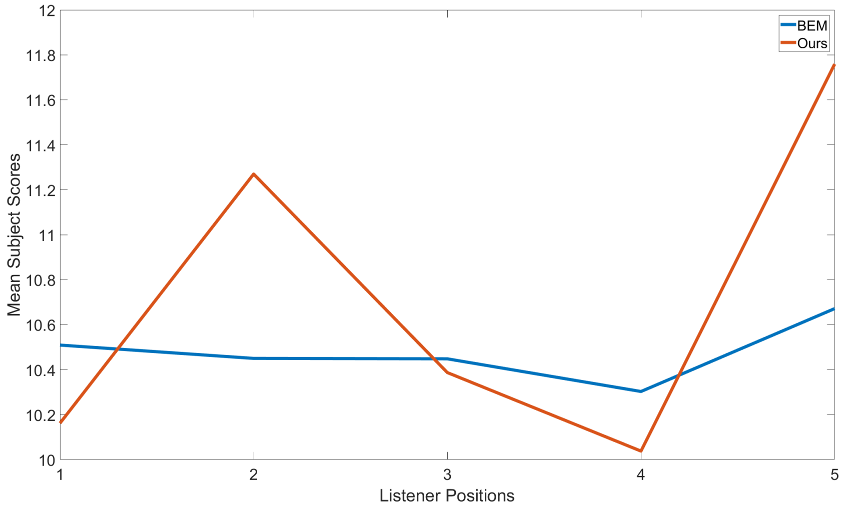

Figure 4.10: Mean subject scores for different positions for the two methods of diffraction.

A two-way, repeated measures ANOVA (factors: diffraction method and listener positions) was performed on the subject's ratings which were averaged over the three blocks, normalized by the subject's mean score for all listener positions and diffraction methods, and scaled by the grand-mean. The test failed to show significance for position and diffraction method. Fig. 4.10 shows the mean values of the subject ratings for the two methods. The results show that our diffraction kernel algorithm performs comparably to the BEM-based wave propagation algorithm.

4.8 Conclusions, Limitations and Future Work

We present a novel approach to model diffraction effects for ray tracing based plausible sound generation algorithms. We introduce the notion of diffraction kernels that can capture many wave effects like diffraction, reflections, scattering, intra-object interference and other interactions using wave-based precomputation.

These kernels are computed independently for each dynamic object in the scene in a few minutes based on a novel source placement algorithm. Moreover, we can easily integrate these kernels with ray tracing based interactive geometric propagation algorithms and have small runtime overhead. We demonstrate the benefits over prior sound propagation algorithms on complex dynamic scenes. We also performed a user study to evaluate the perceived smoothness of the diffracted field and observed that the auditory perception using our approach is comparable to that of a wave-based sound propagation method. To the best of our knowledge, this is the first practical method to generate diffraction effects from a smooth object in dynamic scenes for VR applications.

Our approach has some limitations. While our hybrid approach offers many benefits over geometric acoustic methods, it is less accurate than wave-based propagation methods. Our approach is mainly designed for scenes with well-separated rigid objects, whose scattering behavior does not change at runtime. The diffraction kernels only encapsulate the sound interaction behavior of individual objects in the free field and do not account for phase or inter-object interactions. As a result, they may not work well in certain scenarios. Our formulation of diffraction kernels only take into account the magnitude and the direction, and not the phase.

There are many avenues for future work. It would be useful to model other interactions such as first order surface scattering based on Kirchoff approximation (Tsingos et al., 2007) or wave-based geometric acoustics (Lam, 2005) to model other wave interactions. It would be useful to design approximate schemes that can also model phase, as that is needed for certain applications, such as seat-dip effects in concert halls. The main goal is to estimate the propagation delays for all possible paths. Finally, we would like to perceptually evaluate our approach in other applications such as social VR and telepresence, where it is important to simulate diffraction effects and generate smooth sound fields.

CHAPTER 5: Perceptual Characterization of Early and Late Reflections for Auditory Displays

5.1 Introduction

Sound rendering uses auditory displays to communicate information to a user. Harnessing a user's sense of hearing enhances the user's experience and provides a natural and intuitive human-computer interface. Studies have shown a positive correlation between the accuracy or fidelity of sound effects and the sense of presence or immersion in virtual reality (Larsson et al., 2002a; Dubois et al., 2009; Rungta et al., 2016). Sound is also an important cue for perceiving distance (Zahorik et al., 2005) and orientating oneself in an environment (Wilson et al., 2007).

The sound emitted from a source and reaching the listener can be broken down into three components, described in more detail below: direct sound, early reflections, and late reflections or reverberation (Fig. 5.1). All three components of the sound field have perceptual relevance and have been extensively studied in psychoacoustics. Direct sound gives us an estimate of the loudness and the distance to the sound source (Zahorik and Wightman, 2001). Early reflections (ERs) arrive later than the direct sound, often in a range from 5 to 80 milliseconds. Late reflections or reverberation (LRs) are generated when the sound signal undergoes a large number of reflections and then decays as it is absorbed by the objects in the scene.

Because of the importance of different components of sound fields, there has been considerable work on simulating these effects and incorporating them into auditory displays. Some of the commonly used methods approximate the sound field using artificial reverberation filters, which use reverberation time (RT_{60}) to tune parametric digital filters (Valimaki et al., 2012). These filters tend to have low computational requirements and are widely

used for interactive auditory displays (Kleiner et al., 1993). However, finding the right parameters for reverberation filters can be time-consuming and current methods do not provide sufficient fidelity. Geometric sound propagation methods work under the assumption that sound travels in straight lines and can be modeled using ray tracing (Krokstad et al., 1968a).

This allows resulting algorithms to model sound's interaction with the environment as it undergoes reflection and scattering. Many techniques have been proposed to accelerate ray tracing, and current methods can generate early reflections (ERs) and late reflections (LRs) at interactive rates in dynamic scenes using high-order ray tracing (e.g., more than 100 orders of reflections) (Schissler and Manocha, 2017). In practice, high-order ray tracing can be expensive and current interactive systems use multiple CPU cores on desktop workstations. The most accurate methods for sound rendering are based on wave-based acoustics, which directly solve the acoustic wave equation using numerical methods. However, their precomputation and storage overheads are very high and current methods are only practical for lower frequencies (Mehra et al., 2012, 2013; Raghuvanshi et al., 2010).

Many applications, including games, virtual environments, and multi-modal interfaces require an interactive sound rendering capability, i.e., 20fps or more. Furthermore, these systems are increasingly used on game consoles or mobile platforms where computational resources are limited. As a result, we need faster techniques to generate ERs and LRs in dynamic scenes for high-fidelity sound rendering. In particular, LR computation can be a major bottleneck.

Main Results: We present a novel, perceptually derived metric called $P - Reverb$ that relates the ERs to the LRs in the scene. Our approach is based on the relationship between the mean-free path (μ) and reverberation (Eq. 5.3), and we use early reflections to numerically estimate the mean-free path of the environment. We conduct two extensive user evaluations that establish the just-noticeable difference (JND) of sound rendered using early reflections and late reflections in terms of the mean-free path. We derive our perceptually-based $P - Reverb$ metric by expressing the JNDs of early and late reflections in terms of the mean-free path. Moreover, our metric is used to efficiently estimate the late reverberation parameter (RT_{60}).

We have evaluated the accuracy of our perceptual metrics in terms of computing the mean-free paths and reverberation time and comparing their performance with prior algorithms based on analytic or high-order ray tracing formulations. The mean-free path is within 3% and reverberation time is within 4.6%, which are within the JND values specified by ISO 3382-1 (ISO, 2009).

Overall, we observe significant benefits using our $P - Reverb$ metric for fast evaluation of mean-free path and reverberation parameters for sound rendering and auditory displays. We have used for sound propagation and rendering in complex indoor scenes.

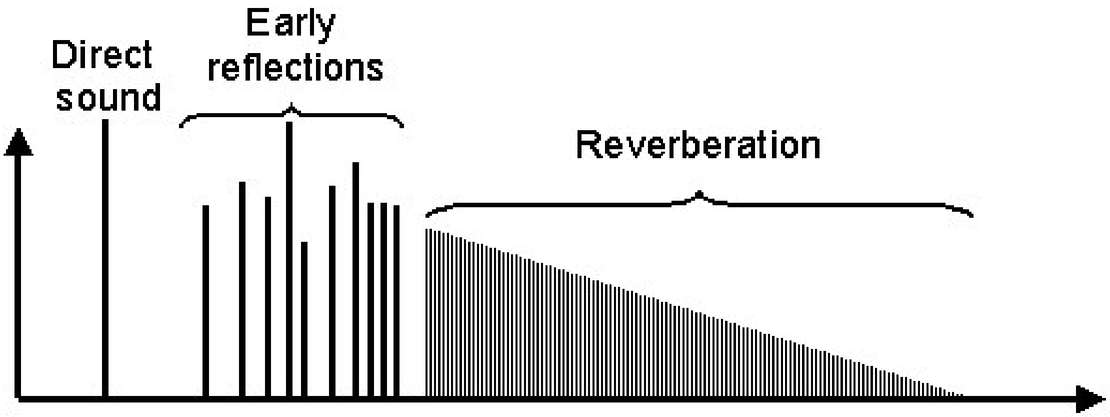

Figure 5.1: We highlight the different components of the sound field. The sound directly reaching the listener is called the direct sound, the reflections that reach in the first 80 ms are called early reflections (ERs), while the reflections following the early reflections that show a decaying exponential trend are called late reflections or reverberation (LRs). Our $P - Reverb$ metric presents a new perceptual relationship between ERs and LRs and we use it for fast sound rendering.

5.2 Related Work

In this section, we give an overview of prior work in sound propagation, psychoacoustic characteristics, and related areas.

5.2.1 Reverberation

Reverberation forms the late sound field and is generated by successive reflections as they diminish in intensity. Reverberation is regarded as a critical component of the sound field. Many acoustic parameters such as the reverberation time (RT60) and clarity index (C50 and C80) are used to characterize reverberation (Kuttruff, 2016).

5.2.1.1 Reverberation Time (RT_{60}):

RT_{60} is defined as the time for the sound field to decay by 60dB. A well-known expression used to compute the reverberation time is given by Sabine's formula, which gives the relationship between the RT60 of a room in terms of its volume, surface area, and the total absorption coefficients of the materials used:

$$RT_{60} \approx 0.1611 sm^{-1} \frac{V}{Sa}, \tag{5.1}$$

where V is the total volume of the room in m^3, S is total surface area in m^2, a is the average absorption coefficient of the room surfaces, and Sa is the total absorption in sabins. In this paper, we use RT_{60} as the main reverberation parameter and use our $P - Reverb$ metric for fast computation in complex scenes.

5.2.1.2 Mean-Free Path

The mean-free path (MFP) of a point in the environment is defined as the average distance a sound ray travels in between collisions with the environment and is directly related to the RT_{60} (Kuttruff, 2016):

$$RT_{60} = k\frac{\mu}{log(1-\alpha)}, \tag{5.2}$$

where k is the constant of proportionality, μ is the mean-free path, and α is the average surface absorption coefficient. A closed form expression (Bate and Pillow, 1947) for computing the mean-free path is given by:

$$\mu = \frac{4V}{S}, \tag{5.3}$$

where V is the volume of the environment and S is the surface area. The mean-free path can be computed to a reasonable degree of accuracy by only considering the specular reflection paths in the scene (Vorländer, 2000). We use our $P - Reverb$ metric for fast computation of μ using only ER in complex scenes.

5.2.2 Sound Propagation and Acoustic Modeling

Artificial reverberators provide a simple mechanism to add reverberation to "dry" audio, which has led to their widespread adoption in the music industry, virtual acoustics, computer games, and user interfaces. One widely used artificial reverberator was introduced by Schroeder (Schroeder and Logan, 1961) and it uses digital nested all-pass filters in combination with a parallel bank of comb filters to produce a series of decaying echoes. These filters require parameters such as reverberation time (RT_{60}) to tune the all-pass and comb filters. Geometric methods work on the underlying assumption of the rectilinear propagation of sound and use ray tracing to model the acoustics of the environment (Krokstad et al., 1968a). Other geometric methods include beam tracing (Funkhouser et al., 1998b) and frustum tracing (Chandak et al., 2008). In practice, ray tracing remains the most popular because of its relative simplicity and generality and because it can be accelerated on current multi-core processors. Over the years, research in ray tracing-based sound propagation has led to efficient methods to compute specular and diffuse reflections (Schissler et al., 2014a) for a large number of sound sources (Schissler and Manocha, 2017).

5.2.3 Early & Late Reflections: Psychoacoustics

Early reflections (ERs) have been shown to have a positive correlation with the perception of auditory spaciousness and are very important in concert halls. (Barron, 1971; Blauert and Lindemann, 1986) showed that adding early reflections generated the effect of "spatial impression" in subjects. Early reflections are also known to improve speech clarity in rooms. (Bradley et al., 2003) showed that adding early reflections increased the signal-to-noise ratio and speech intelligibility scores for both impaired and non-impaired listeners. (Hartmann, 1983) showed that early reflections that come from the same direction as the direct sound reinforce localization, while those coming from the lateral directions tend to de-localize the sources.

Late reflections or reverberation (LRs) provide many perceptual cues. Source localization ability deteriorates in reverberant conditions, with localization accuracy decreasing in a reflecting room compared to the same absorbing room (Hartmann, 1983). Reverberation has a negative impact on speech clarity and (Knudsen, 1932) showed the reduction in the number of sounds heard correctly in the presence of reverberation.

Although reverberation decreases localization accuracy and speech intelligibility, it is known to have positive effects with respect to the perceived distance to a sound source in the absence of vision (Zahorik et al., 2005).

While there is considerable work on separately characterizing the perceptual effects of ERs or LRs, we are not aware of any work that establishes any relationship between ERs and LRs. $P - Reverb$ is a metric that establishes the relationship between the respective JNDs and uses them for interactive sound rendering.

5.2.4 Estimating Reverberation Parameters

Given the importance of reverberation to the overall sound field, multiple methods have been established to measure the reverberation parameters over the years. (RT_{60}), in particular, is considered to be the most important parameter in estimating reverberation and has been referred as the 'The mother of all room acoustic parameters' (Skålevik, 2010). The most commonly used method to estimate reverberation time was given by Schroeder, and uses a backward time integration approach. (Ratnam et al., 2003) presents a method for blind estimation of RT_{60} that does not require previous knowledge of sound sources or room geometry by modeling reverberation as an exponentially damped Gaussian white noise process. (Löllmann and Vary, 2008) describes a method to estimate reverberation time using maximum likelihood estimator from noisy observations. (Vorländer and Bietz, 1994) presents a comparison of different methods for estimating RT_{60}.

5.3 Perceptual Evaluations and P-Reverb

In this section, we describe two user evaluations that establish the just-noticeable difference (JND) for early and late reflections in terms of the mean-free path. Further, we show the relationship between the two JND values, thereby establishing our $P - Reverb$ metric.

5.3.1 Experiment I - Just-noticeable difference of ERs

In this experiment, we seek to establish the just-noticeable difference (JND_{er}) of sound rendered using only direct and early reflections. In Experiment II, we establish the relationship between JND_{er} and sound rendered using the full simulation (direct + early + late reverberation) JND_{lr}.

Participants: 106 participants took part in this web-based, online study. The subjects were recruited using a crowd-sourcing service. All subjects were either native English speakers or had professional proficiency in the language.

Apparatus: The online survey was set up in Qualtrics. The impulse responses were generated using an in-house, realtime, geometric sound propagation engine written in C++, while the convolutions to generate the final sounds were computed using MATLAB.

Stimuli: The stimuli were sound clips derived from 7 cube-shaped rooms with increasing edge lengths such that their MFPs (Eq. 5.3) varied from $2 - 2.2$m in increments of 0.033m. The range of lengths was chosen with the experimental goal in mind, namely, to extract a psychophysical function showing a gradient in perceived sound similarity relative to edge-length difference. The walls of the rooms had reflectivity similar to that of an everyday room. The source was a sound of clapping, which was chosen because it represents a broadband signal. The clips were filtered in 4 logarithmically spaced frequency bands ($0-176$Hz, $176 - 775$Hz, $775 - 3408$Hz, and $3408 - 22050$Hz) to evaluate the effects of frequency on JND_{er}. Each of these 4 filtered clapping sounds was convolved with the early impulse responses of the 7 rooms. The final sound clip was around 4 seconds long and contained 3 distinct parts: 1.5 seconds of the clapping sound in Room 1 ($\mu = 2m$), 1 second of silence, and another 1.5 seconds of clapping in a second room drawn from

$1 - 7$ ($\mu = [2to2.2]m$). All sounds were recorded assuming that the listener and the source were located at the origin $(0, 0, 0)$. Given this symmetry, the sounds were rendered in mono with both speakers playing the same sound.

Design & Procedure: To estimate the JND, our experiment used the method of constant stimuli (Gescheider, 2013) with a within-subject design. A stimulus comprised a sound clip containing Room 1 and one of the 7 possible comparison rooms (including Room 1). For each clip the subjects heard, they were asked to identify if the first clapping sound seemed to be different from the second clapping sound by selecting yes or no. Note this is a similarity judgment, not a discrimination. A block of judgments consisted of 28 clips (4 frequencies x 7 comparison rooms paired with Room 1). A block was repeated 5 times, giving a total of 140 clips (4 frequencies x 7 possible rooms paired with Room 1 x 5 blocks). The ordering of the clips was randomized within a block. Each subject judged all 140 stimuli. Before starting the experiments, subjects listened to a sample clip for familiarization. The subjects were required to have a pair of ear-buds/headphones to take part in the study, which took an average of 25 minutes to complete.

Results & Analysis: Fig. 5.2 shows the proportion of responses in which rooms were judged as sounding different, over all participants, as a function of the comparison level of μ. The first data point corresponds to rooms that were objectively identical, providing a baseline. The data essentially increases linearly with a larger μ, showing greater discrimination up to Room 6 ($\mu = 2.17$), after which ($\mu = 2.2$) the discriminatory ability seems to taper off. The standard errors are low and consistent, indicating the robustness of the results.

An interesting observation is the near-invariance of subjects' ability to discriminate across the frequency bands. This was verified by an ANOVA analysis with factors of edge length (or μ) and frequency. The analysis showed significant main effects for both factors of edge length $F(6, 180) = 61.78, p < 0.05, \eta_p^2 = 0.673$ and frequency $F(3, 90) = 2.95, p = 0.037, \eta_p^2 = 0.09$. The interaction between edge length and frequency was also significant $F(18, 540) = 1.66, p = 0.04, \eta_p^2 = 0.052$, reflecting that the performance decrement at the largest edge length is slightly greater for frequency band 4. However, the η_p^2 values are very low for effects involving frequency. Thus, while the effects of frequency show statistical significance, they are small in effect size and do not reflect consistent variation in frequency across edge length (or μ). Therefore, for the purposes of constructing an overall rule, using data averaged over the frequencies is a valid simplification, particularly if the largest value of edge length is excluded. Fig. 5.3 shows the results averaged over the frequency for

Rooms $1-6$. As shown in the figure, the data fits a linear function well, with $R^2 = 0.98$. Given our linear fit:

$$\delta = 3.89\mu - 7.5, \tag{5.4}$$

we can easily estimate the JND_{er} by considering the MFP values (μ_{JND}) where the subjects successfully discriminated the sounds 50% of the time given by $\mu_{JND} = \mu_{50\%} - \mu_{room1} = 2.06 - 2 = 0.06m$. This tells us that a change in μ greater than 0.06m would result in perceptually differentiable sounds when using *early impulse responses*, but it doesn't necessarily indicate if the relationship holds if the sounds were rendered using the *full impulse response* (LR). This led us to conduct the next experiment to establish the relationship between the JND of early reflections (JND_{er}) and the JND of full impulse response or late reverberation (JND_{lr}).

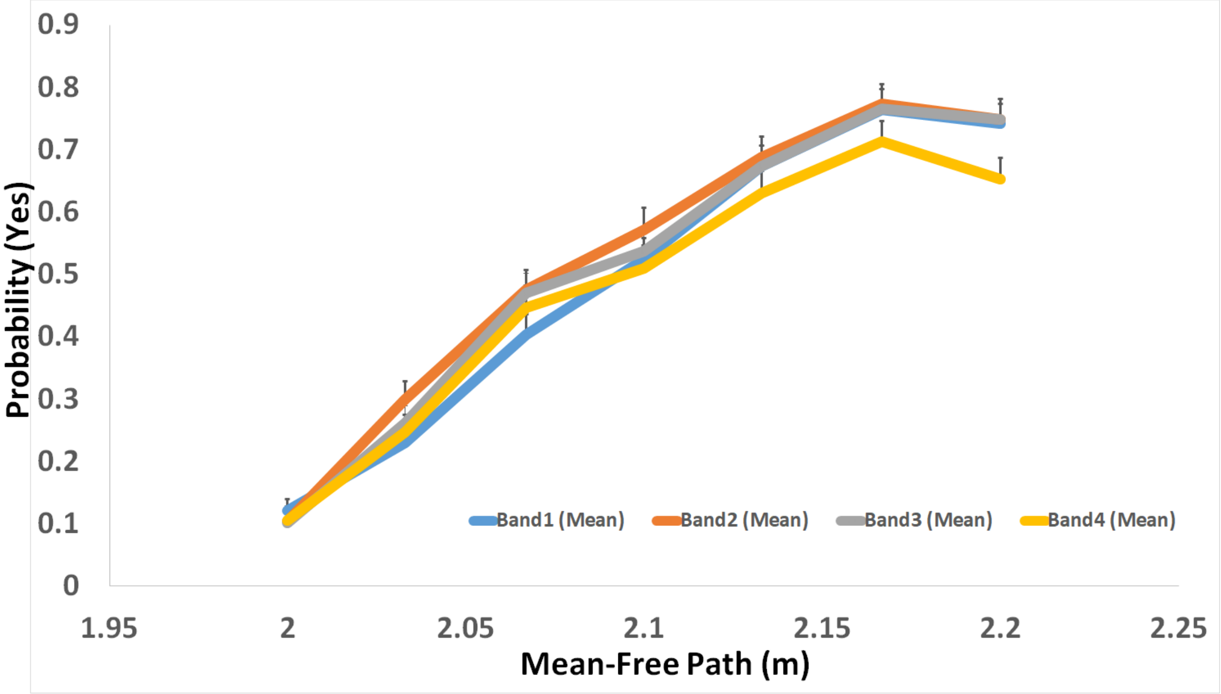

Figure 5.2: The psychometric function for sound rendered using the early reflections for the 4 frequency bands. The Y-axis shows the proportion of responses indicating the sounds were different. We see a clear, linear trend between increasing μ and the probability of responding different, until the last room $\mu = 2.2$, where the responses seem to flatten out.

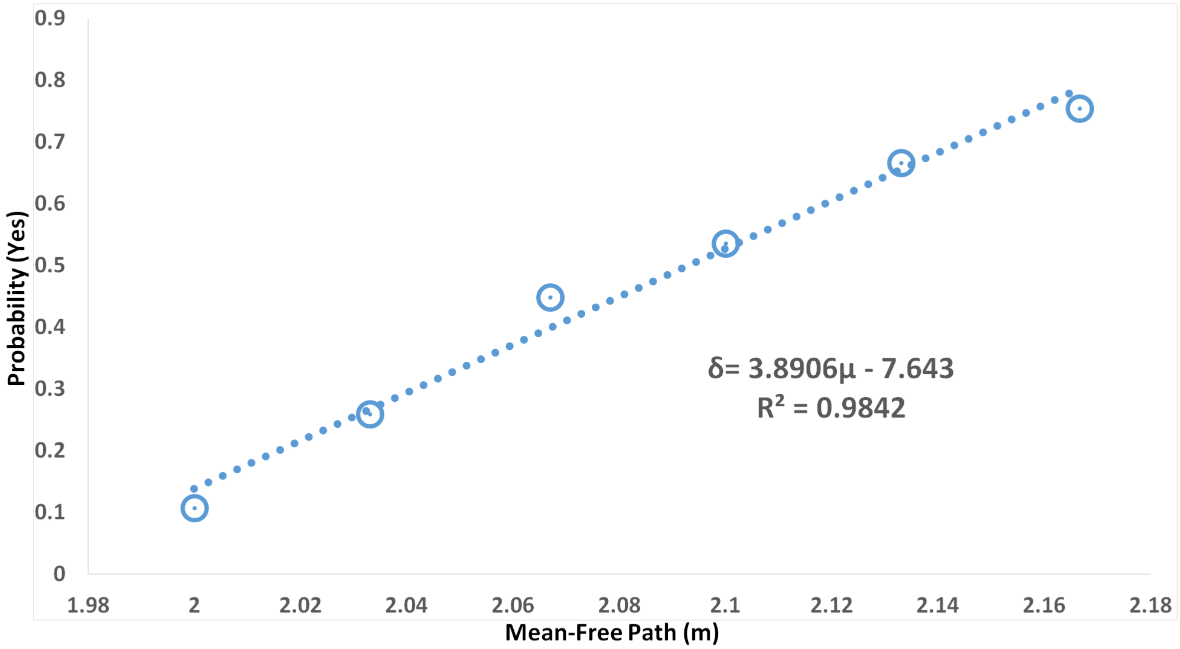

Figure 5.3: The average JND over the frequency bands. The Y-axis shows the proportion of responses indicating that a difference was judged. The psychophysical function is essentially linear, showing that the probability of judging the sounds as different increases linearly with the increasing mean-free paths of the rooms.

5.3.2 Experiment II - Relationship between JND_{er} & JND_{lr}

Once we have established the perceptibility threshold or JND_{er} of ERs, we need to relate this to the JND_{lr} of LRs. Our goal is to use these relationships to cluster points $p \in P$ with similar reverberation characteristics. We conducted another user study based on the results of the first study, described above.

Participants 31 participants took part in this online, web-based study. The subjects were recruited using the same crowd-sourcing service as in the previous experiment. All subjects were either native English speakers or had professional proficiency in the language.

Apparatus The apparatus was the same as in Experiment I. The full impulse responses were generated using our in-house, realtime, geometric sound propagation engine written in C++, with the convolutions being computed using MATLAB.

Stimuli The sound source used was the same as in the previous experiment, filtered for the same logarithmically-spaced frequency bands. Given our goal of establishing the relationship between JND_{er} and JND_{lr}, we use our previously computed psychometric function (Eq. 5.4), to compute the 6μ *values corresponding to* *detection rates ranging from* 0.2 *to* 0.7. This gives us 6 μ values that can then be used to compute the cube rooms' edge lengths using Eq. 5.3. These 6 rooms and Room 1 from the previous experiment serve as the environments in which the full impulse responses are computed. The material properties of the rooms were the same as in the previous experiment. Each sound clip in this case was about 6 seconds because of the increased length of full impulse response, with 2.5 seconds of clapping in Room 1, followed by a second of silence, followed by 2.5 seconds of clapping in Rooms $2-7$. The total number of sound clips was 140, as before (4 frequency bands x 7 rooms x 5 blocks). The ordering of the sound clips was randomized within each block.

Design & Procedure The study design was the same as in the ER study. Before starting the study, the subjects were asked to listen to a sample sound clip from the 28 clips computed above for familiarization. The source and listener locations in the rooms were located at $(0, 0, 0)$. The sound was rendered in mono. The subjects took an average of 30 minutes to complete the study.

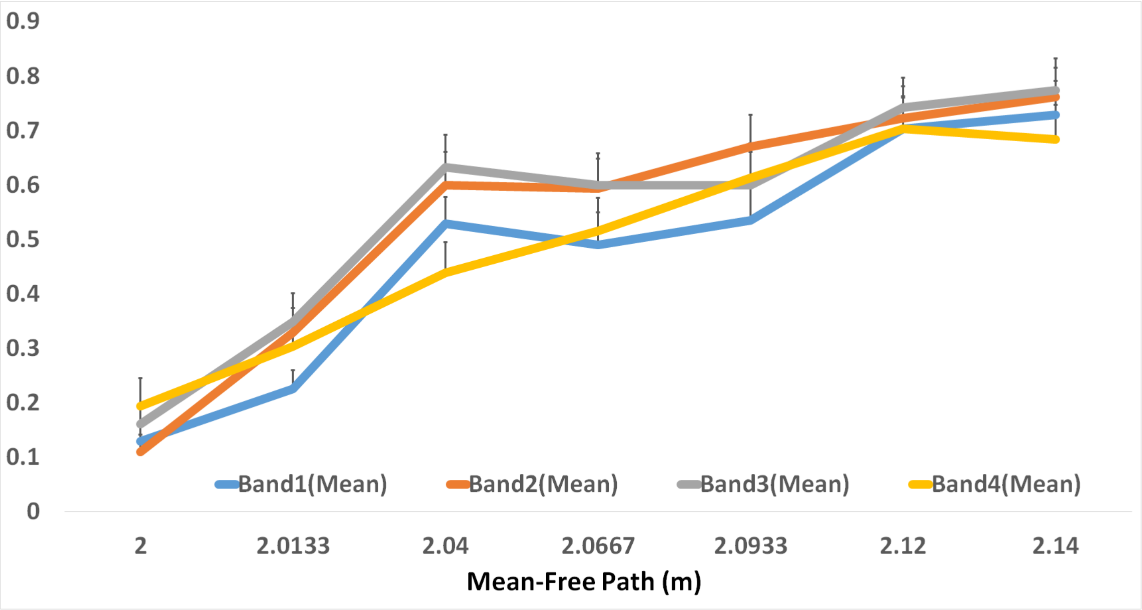

Figure 5.4: The psychometric function for sound rendered using the full impulse response (LR) for the 4 frequency bands. The Y-axis shows the proportion of responses indicating sounds were judged to be different. In this case, we observe more variability for the different frequency bands, which could be attributed to the greater sensitivity of human hearing to a more accurate signal (compared to the less accurate ER signal). Overall, however, the responses can be modeled as a linear function with reasonable accuracy.

Results & Analysis Fig. 5.4 shows the proportion of responses judging the sounds as different, as a function of increasing μ or edge-length. As before, we performed an ANOVA to assess the effect of edge length and frequency. The analysis showed significant main effects for edge length $F(6, 180) = 61.78, p < 0.05, \eta_p^2 =$

0.673 and frequency $F(3, 90) = 2.95, p = 0.037, \eta_p^2 = 0.09$. The interaction between edge length and frequency was also significant $F(18, 540) = 1.66, p = 0.04, \eta_p^2 = 0.052$. Again, the effect size for terms involving frequency was low, allowing us to average the responses for the frequency bands. Fig. 5.5 shows the values averaged for the 4 frequency bands.

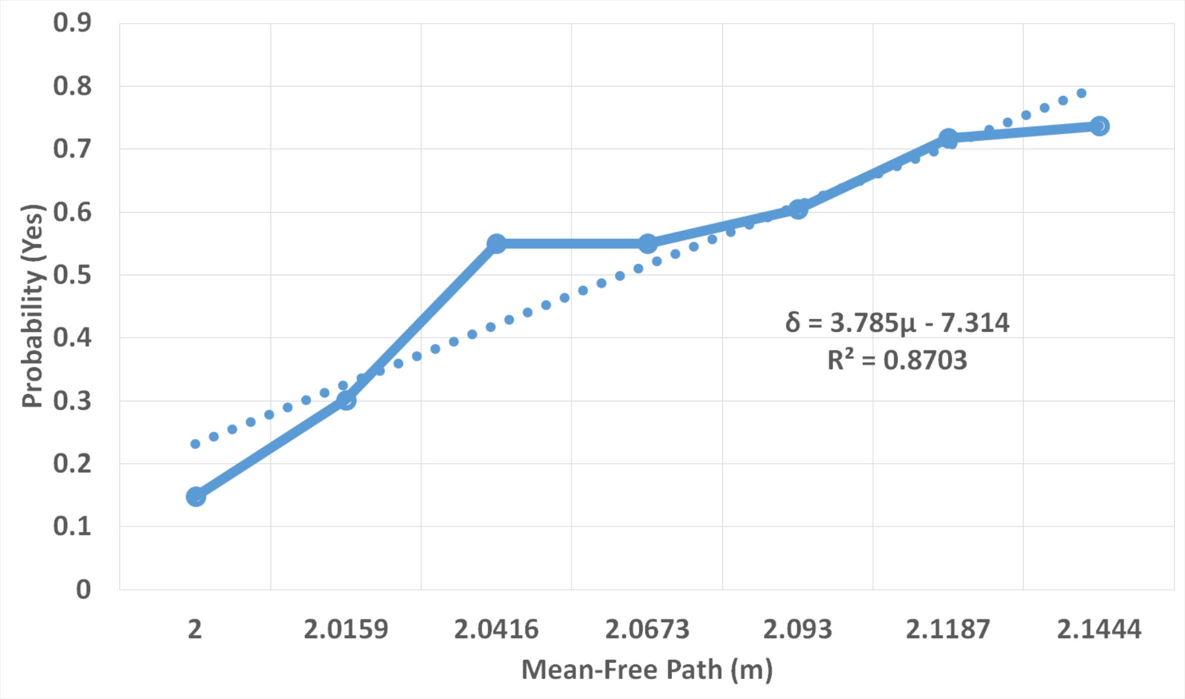

Figure 5.5: The average JND over the frequency bands for the full impulse response signal.The Y-axis shows the proportion of responses indicating a judgment of difference. The psychophysical function is not as linear as the early reflection signal, but a linear function approximates the subject responses reasonably well ($R^2 = 0.87$), accounting for most of the variability.

5.3.3 $P - Reverb$ Metric

Fig. 5.6 shows the relationship between the sounds rendered using only the early responses and the sounds rendered using the full impulse response. Note that the first point for both functions corresponds to two identical stimuli, and no difference is expected. However, beginning at the smallest edge lengths where objectively different stimuli were presented, the figure shows that the subjects were more likely to differentiate between sounds rendered with the full impulse response than they were with sounds rendered using only the early reflections. A difference in difference judgments is expected, because the full impulse response conveys more information about the space and is supposed to enable better perceptual differentiation than the early impulse response, thus giving a lower JND for the full impulse response, i.e., $JND_{lr} < JND_{er}$.

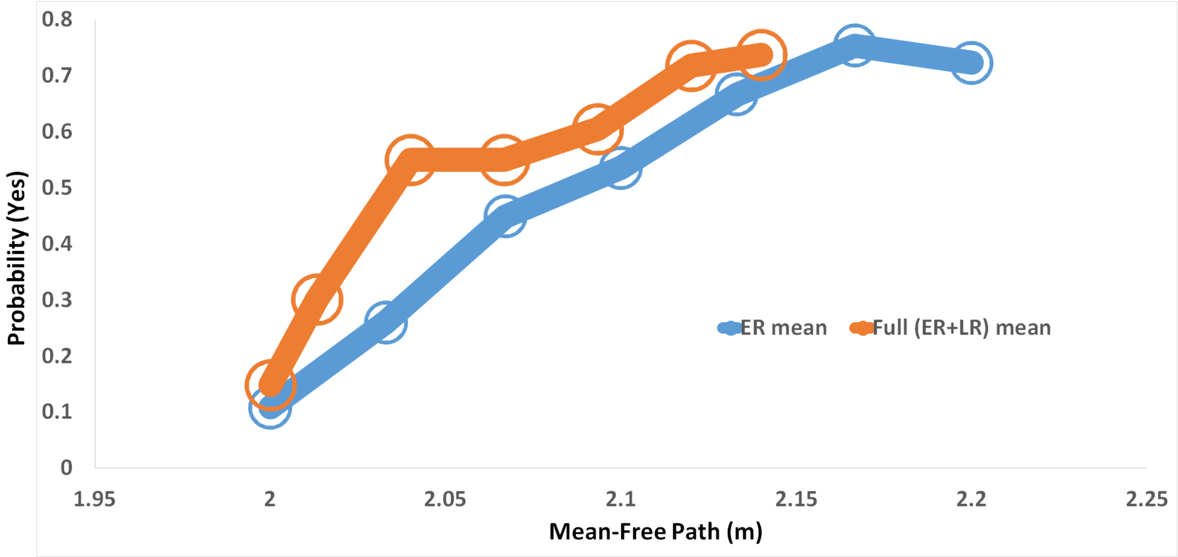

Figure 5.6: This plot shows the overlaid psychometric functions for signals rendered using the early reflections (blue) and full impulse response (orange). Note that the first data point corresponds to differences being reported when the stimuli are objectively identical. Although the full impulse data shows a greater departure from a linear relationship beyond that point, the results are similar to the early reflection function, offset by a constant, allowing us to establish a simple, linear relationship between JND_{er} and JND_{lr} in terms of the mean-free path.

To establish a mathematical relationship between the two JNDs, we consider the ratio of the mean-free paths in both cases. The resulting figure is shown in Fig. 5.7. The linear fits are almost coincident after adding a constant offset of 0.02, i.e.

$$\frac{\mu_{JND_{lr}}}{\mu_1} + 0.02 = \frac{\mu_{JND_{er}}}{\mu_1}, \tag{5.5}$$

which gives a simple relationship between the two JND values:

$$\mu_{JND_{lr}} = \mu_{JND_{er}} - 0.02\mu_1, \tag{5.6}$$

where μ_1 is the mean-free path in Room 1 = 2m. Hence $\mu_{JND_{lr}} = \mu_{JND_{er}} - 0.04$ is the simple mathematical relationship or $P - Reverb$ for the JND values of the two signals. Given $\mu_{JND_{er}} = 0.06m$ as derived above, we can easily compute the value of $\mu_{JND_{lr}}$ as being $0.02m$ for a reference room (Room 1) $\mu = 2m$, *giving us the percentage change (*$\frac{\mu_{JND_{er}}}{\mu_{Room1}} = 1\%$*) in the mean-free path values that constitute the JND for late reverberation*, when using early reflections.

It turns out that Eq. 5.6 can be interpreted as a "first-order" approximation to a function that expresses the mathematical relationship between two multi-dimensional perceptual phenomena that are dependent on frequency, edge length, method of rendering, material parameters, etc. However, any function that accounted for the small frequency dependencies in the observed psychometric data and accommodated the effects of more complex environments and material parameters would have to be substantially more complicated than the linear relationship that we derive here. The value of the present formulation lies in its reasonable approximation of the observed effects with only one derived parameter.

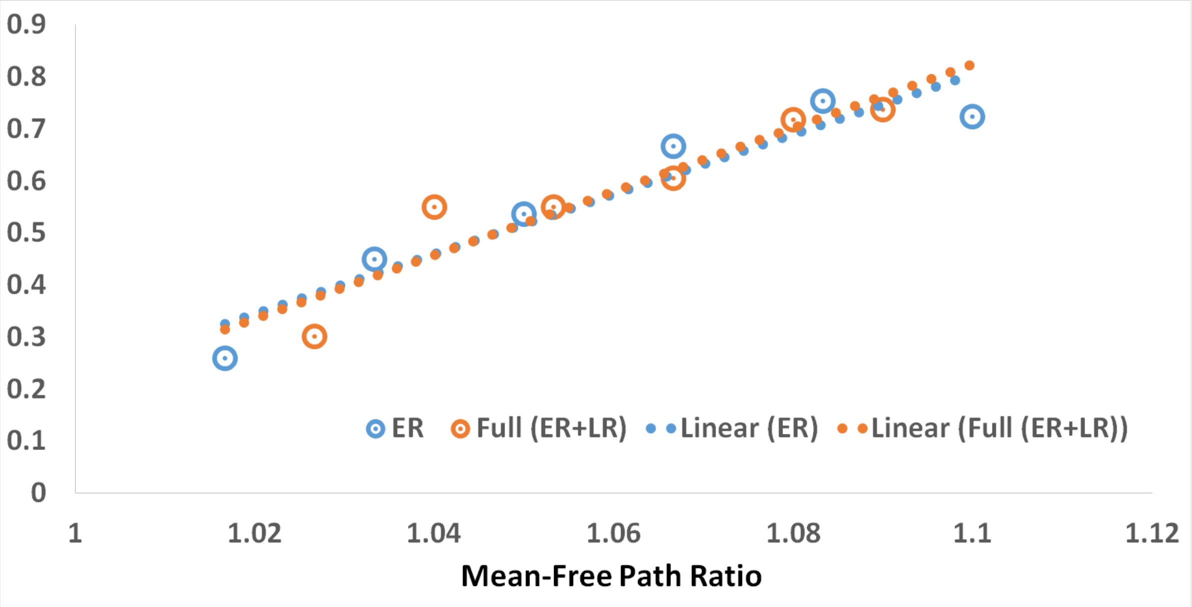

Figure 5.7: The psychometric function with a constant offset adjustment. We consider the ratio of the mean-free path for the different rooms to the mean-free path of Room 1. The resulting linear fits for the two cases (early reflections and full impulse response) coincide once a constant offset of 0.02 is added to the ratio for the full impulse response. This highlights the accuracy of our model.

We would also like to note that, although psychometric functions are usually fitted using sigmoid functions, our design did not require us to do so. A sigmoid function approach would have been suitable had we started with a value somewhere in the middle and taken a range of values above and below. This would have yielded two end-cases with the non-standard stimulus being judged smaller 100 % of the time; similarly, the larger non-standard stimulus would be judged as such 100 % of the time. In our approach, however, we never tested anything smaller than the standard, which led us to values that rose to the ceiling. Consequently, a linear fit to this function accounted for most of the variance (93%). A better fit could be achieved using a quadratic fit (accounting for 99% of the variance), but at the expense of adding a parameter. A sigmoid function, too, would add another parameter without yielding much gain. Therefore given the fact that our linear fit accounts for most of the variability, we chose to not use a sigmoid fit.

5.4 Results & Evaluation

Our approach consists of two primary numerical steps: computing the mean-free path (μ) using early reflections (ERs), and predicting RT_{60} using our perceptually established $P - Reverb$ metric. We first validate the use of early reflections (ERs) to compute the mean-free path (μ) in various environments. Next, we highlight the validation of the $P - Reverb$ metric in terms of its accuracy in predicting RT_{60}.

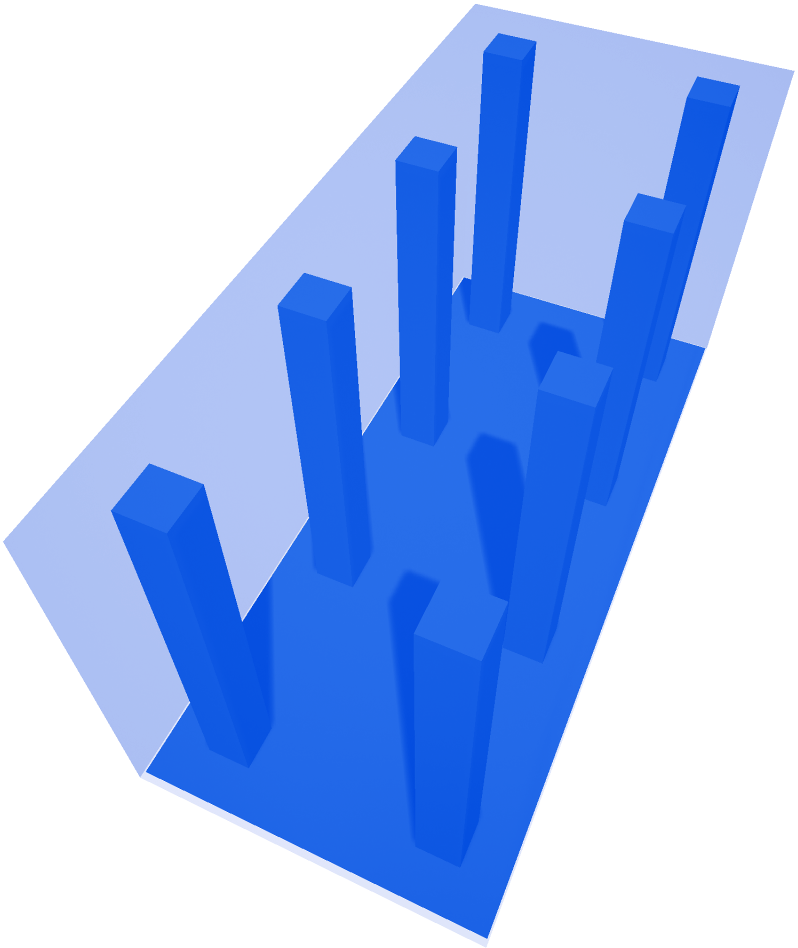

Figure 5.8: Room with Pillars: We illustrate the room with 8 pillars and use this benchmark to estimate the effectiveness of our mean-free path computation in complex environments with obstacles. We observe less than 3% error using our early reflection based method.

5.4.1 Mean-Free Path Computation

Our $P - Reverb$ metric depends on the numerically computed mean-free values that are computed using early reflections. The mean-free path is the average distance a sound ray would travel between collisions and we use ERs to estimate this distance. As mentioned, Eq. (5.3) can be used to compute mean-free path values in terms of the volume (V) and surface area (S). Table 5.1 highlights the accuracy of our computed mean-free path values (μ_{er}) as compared to the analytical values given by Eq. 5.3. We use 500 rays and 20 bounces for each ray to compute our μ_{er} value as:

$$\mu_{er} = \frac{\sum d_i}{n \times b}, \tag{5.7}$$

where d_i is the distance traveled by a sound ray on the i^{th} bounce, n is the total number of rays, and b is number of bounces per ray.

Shape	$\mathbf{Dim.(m)}$	$\mu_{\mathbf{er}}\mathbf{(m)}$	$\mu_{\mathbf{an}}\mathbf{(m)}$	$\mathbf{\%error}$
Cube	5	3.3	3.33	1
Rect. Prism	(2,3,4)	1.87	1.85	1.3
Sq. Pyramid	(2.8, 3) (b, h)	1.16	1.18	1.7
Room with Pillars	(5,6,12)	3.14	3.04	3

Table 5.1: **Mean-free path Computation:** We show the accuracy of computing μ_{er} using early reflections for differently shaped rooms. The closed-form expression in Eq 5.3 gives us the analytical value for the mean-free paths in each of the rooms μ_{an}. We observe that ERs can closely approximate the analytically obtained μ_{an}. The Room with Pillars is shown in Fig. 5.8. Even for a scene with multiple obstacles, our method computes the mean-free path while inducing a maximum error of only 3%.

5.4.2 RT_{60} using $P - Reverb$ Computation

The $P - Reverb$ metric predicts regions in a scene where the late reverberation is likely to vary imperceptibly. Conversely, it can estimate regions where the late reverberation would vary in a perceptually noticeable manner. We demonstrate the effectiveness of the $P - Reverb$ metric in finding regions of similar reverberation characteristics by considering a scene shown in Fig. 5.9. The scene is composed of different interconnected rooms of varying shapes and volumes. Since reverberation is a function of the volume and shape of the room, it is likely to vary as one moves from one room to another. We consider a path that traverses three different connected rooms and compute the mean-free path along the path using ERs. Fig.

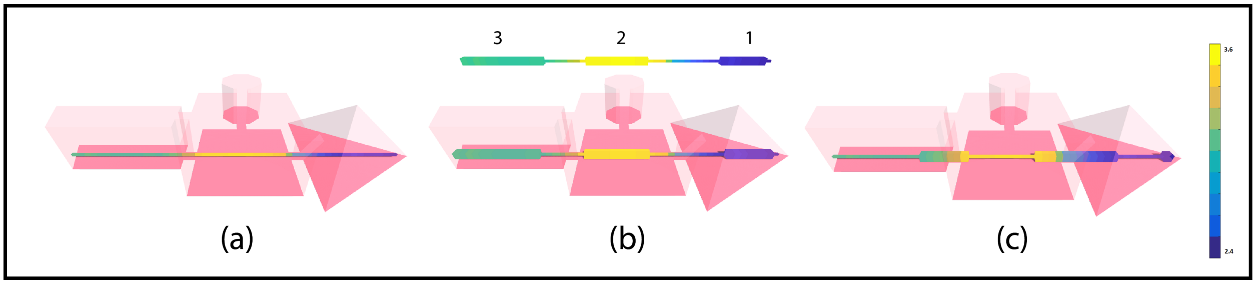

Figure 5.9: We highlight the application of $P - Reverb$ metric to predict variations in RT_{60} in a scene composed of interconnected rooms of different shapes and volumes: (a) shows the variation in μ along a path that goes through three rooms with volumes 135 m^3, 256m^3, and 125 m^3 from left to right; (b) shows three regions $(1, 2, 3)$ along the path roughly corresponding to the three rooms, where μ changes within the JND specified by the $P - Reverb$ metric. This indicates that the reverberation in these regions would vary imperceptibly, as is indicated by the uniformity of the μ values; (c) shows rapidly varying μ values as one approaches the apertures between the connected rooms, indicating that RT_{60} would also vary rapidly. This is expected because the geometry varies rapidly in these regions and validates the accuracy of our perceptual metric $P - Reverb$.

5.9(a) shows the variation in μ as we move along the path. We group the regions along the path where μ varies within the JND threshold computed using our $P - Reverb$ metric (as shown in Fig. 5.9(b)), as Regions 1, 2, and 3. Based on the $P - Reverb$ metric, each such region is likely to have an imperceptible sound in terms of RT_{60}. We illustrate this in Table 5.2. The μ_{mean} corresponds to the average mean-free path value for the entire region and $Diff.^{\mu}_{max}$ corresponds to the maximum difference from the μ_{mean} for all the points in that region (i.e., a measure of variance). The RT_{60}^{mean} represents the average value of the reverberation times for the region, while $Diff.^{RT_{60}}_{max}$ corresponds to the maximum difference from RT_{60}^{mean}. For regions where μ varies within the JND specified by the $P - Reverb$ metric, the RT_{60} values vary within 5% of the RT_{60}^{mean}. This is within established JND values for RT_{60}, as specified in ISO 3382-1 (ISO, 2009) and correspond to imperceptible changes in late reverberation.

Fig. 5.9(c) shows rapidly varying μ values, as one moves from one room to another. This indicates that the reverberation or RT_{60} would vary rapidly in these regions. Since none of these values falls within the JND specified by $P - Reverb$, they cannot be grouped to create regions where reverberation would be imperceptible. This is expected because coupling of spaces is known to affect the sound energy flow and the change of RT_{60} close to the coupling aperture (Jing and Xiang, 2008).

Region	$\mu_{\mathbf{mean}}(\mathbf{m})$	$\mathbf{Diff}^{\mu}_{\cdot\mathbf{max}}$	$\mathbf{RT}^{\mathbf{mean}}_{\mathbf{60}}(\mathbf{s})$	$\mathbf{Diff}^{\mathbf{RT_{60}}}_{\cdot\mathbf{max}}$
1	2.45	1.1 %	0.65	4.6 %
2	3.64	0.5 %	1.27	2.4 %
3	3.11	1.1 %	1.03	3.6 %

Table 5.2: **Mean-Free Path and Reverberation Time Computation:** We show the average values computed using high-order ray tracing for the three different regions shown in Fig. 5.9 and the differences from the average values. Each of these rooms corresponds to imperceptible regions based on our $P - Reverb$ metric. The numerical value shows a maximum variation of 5%, which is within JND values of RT_{60}. The exact RT_{60} was computed using high-order ray tracing with 300 bounces.

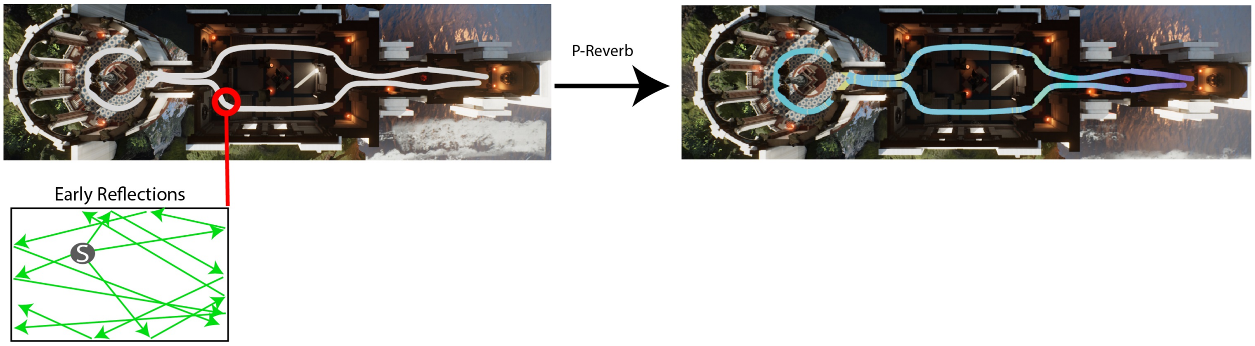

Figure 5.10: The figure shows how the $P - Reverb$ metric can be used to estimate regions where RT_{60} would vary imperceptibly in a scene. The left figure shows a typical listener path in a scene. At each point along this path, we compute the mean-free path μ using the early reflection based method described. Then using the $P - Reverb$ metric, we cluster the points based on the JND to give us clusters along the path where RT_{60} would vary imperceptibly as shown in the right figure.

5.5 Interactive Sound Propagation

In this section we describe how the $P - Reverb$ metric can be used for interactive sound propagation. As described in Sections 1. & 2., the sound reaching the listener from a source has three components: direct sound, early reflections, and late reverberation as shown in Fig. 5.1. Geometric sound propagation algorithms use methods such as ray tracing to compute the ERs and LRs in the scene. Although early reflections can be computed cheaply, late reverberation computation remains a major bottleneck as it requires very high-order ray bounces in the scene for accuracy making these methods resource heavy. This prevents the use of these methods in interactive environments such as games, which tend to use cheap filter-based approaches (digital reverberation filters) to simulate late reverberation. Reverberation filters require parameters such as RT_{60} to approximate late reverberation in an environment. One way in which reverberation filters can be parameterized accurately is to precompute the RT_{60}s along the listener's path in the scene using a high-fidelity geometric sound propagation algorithm such as (Schissler et al., 2014a), and then use these precomputed

RT_{60} values at runtime in the filter. This would avoid costly high-order ray tracing to simulate reverberation at runtime, but can incur a high precomputation cost requiring us to run high-order ray tracing for every point along the listener's path. We now describe how using our $P - Reverb$ metric can reduce the precomputation cost of computing RT_{60} values in the scene.

5.5.1 Sound Propagation using $P - Reverb$

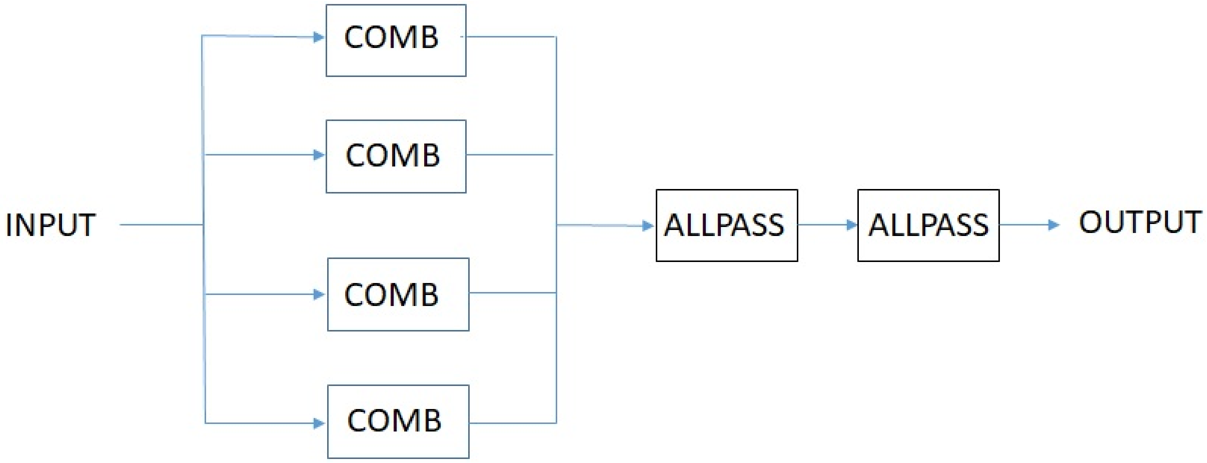

Figure 5.11: The figure shows the schematic of a typical Schroeder-type filter used in our implementation. The input is processed through a parallel bank of comb filters that create the delayed version of the input signal. The output of this parallel bank goes through a series connection of allpass filters. These filters require parameters like RT_{60} to approximate the late reverberation in a scene.

5.5.1.1 Precomputation

We use our $P - Reverb$ metric to accelerate the pre-computation of late reverberation for an interactive sound propagation system using a Schroeder-type reverb filter to simulate late reverberation (Fig 5.11). We sample a given scene at multiple points along the listener's path and use a geometric sound propagation method (Schissler et al., 2014a) to compute early reflections by placing an omni-directional sound source tracing 20 orders of specular reflections at each of these points. Next, using Eq. 5.3 we compute the mean-free paths at each of these points. Using the $P - Reverb$ metric, we clusters points on the path where μ varies within its JND, indicating that these regions will have perceptibly similar RT_{60} values (Fig. 5.10). Finally, using (Schissler et al., 2014a), we compute the RT_{60} values once for each computed region using high-order

(300 bounces) reflections to get a high quality estimate. Table 5.3 shows the speed-up obtained using the $P - Reverb$ metric in precomputation stage. The results were obtained on a multi-core desktop using single thread for the computations.

5.5.1.2 Runtime

At runtime, the direct sound computation is done through visibility testing; if a source is visible to the listener, its distance to the listener is used to attenuate the sound pressure according to the inverse distance law. The late reverberation computation is performed using the precomputed RT_{60} values in the previous stage. Given the listener position, a look-up is performed to ascertain the cluster (precomputed in the previous step) the listener position belongs to. Since, an RT_{60} value is associated with each cluster, this is now used as a parameter into the reverberation filter. As long as the listener in within this cluster, $P - Reverb$ metric tells us that RT_{60} value would vary imperceptibly.

5.5.2 Benchmarks

Sun Temple This scene consists of spatially varying reverberation effects. As the listener moves throughout the scene, the reverberation characteristics vary from being almost dry in the semi-outdoor part of the temple to being reverberant in the inner sanctum.

Shooter Game This scene showcases the ability of our method to handle very large scenes. It shows an archetypal video game with multiple levels. As the listener moves from part of the scene to another, it shows our method's ability to handle highly varying, large, virtual environments.

Tuscany This scene has two different structures, a house and a cathedral, separated by an outdoor garden. The two structures have very different reverberant characteristics owing to their different geometries, and as the listener moves from the house to the cathedral going through the outdoor garden, the reverberant varies accordingly.

5.6 Conclusion, Limitations and Future Work

We present a novel perceptual metric that highlights the relationship between the JNDs of early reflections and late reverberation. Our metric is based on two user studies and can be used for fast computation of mean-free-paths and reverberation time in complex environments without high-order ray tracing. Our metric

$\mathbf{Scene}$	$\mathbf{\#Vert.}$	$\mathbf{\#P}$	$\mathbf{T_{ER}(ms)}$	$\mathbf{T_{LR}(ms)}$	$\mathbf{\#P}$	$\mathbf{Speed-up}$
Sun Temple	215k	2301	40.2	124.2	53	3x
Tuscany	135k	1945	47.5	150.7	110	3x
Shooter Game	49k	3235	16.7	68.4	43	4x

Table 5.3: **Precomputation Performance Analysis:** We highlight the speed-up in precomputation stage using the $P - Reverb$ metric. $\#P$ is the number of points along the listener path, T_{ER} is the avg. time taken at each point using ERs,T_{LR} is the average time taken at each point using LRs, and $\#P$ is the number of clusters found using our $P - Reverb$ metric.

can be used to predict regions in an environment where the reverberation time is likely to vary within its JND value. We evaluate the accuracy of these perceptual metrics and find their accuracy within 5% of the actual values on our benchmarks.

Our approach has some limitations. Our $P - Reverb$ metric computation may not work in totally open environments since the mean-free path computation depends on the presence of collisions with the obstacles in the scene. Our $P - Reverb$ metric can be regarded as an approximation to a complex function that corresponds to a multi-dimensional perceptual phenomenon dependent on source frequency, scene dimensions, method of sound rendering, material parameters, etc. As a result, we need to perform more evaluations that take other parameters into account. While we observe high accuracy in our current benchmarks, the accuracy could vary in more complex scenes. Further, our metric tends to be conservative and overestimates the number of regions with similar RT_{60} resulting in running more full simulations than optimal. That being said, it still significantly reduces the number of full simulations as shown in Table 5.3. Our experimental work also has limitations, including the restricted range of room sizes (motivated by the psychophysical goal), the fixed listener, and the restriction to mono rendering. As part of future work, we would like to overcome these limitations and further evaluate our approach on complex scenes and use them for multi-modal rendering.

CHAPTER 6: SUMMARY AND CONCLUSIONS

6.1 Summary of Results

In this dissertation, we have presented user-studies of interactive sound propagation algorithms to compare their perceptual effectiveness in simulating complex acoustic phenomena such as diffraction and reverberation. Our results show that accurate sound propagation methods result in better perceptual differentiation. Based on the results of these studies, we further present novel methods for interactive sound propagation that incorporate modal sounds, diffraction, and reverberation. In contrast to prior methods for modal sound propagation, our method incorporates the dynamic directivities of each mode of a vibrating rigid body and propagates them into the environment rather than just free space. Further, we propose a perceptually-driven approximation of computationally expensive Hankel functions that makes our method interactive for multiple sources in complex environments. Secondly, we have proposed a diffraction method that can be easily integrated into existing geometric sound propagation methods significantly enhancing their diffraction handling capabilities. Our method can handle complex, highly-tessellated objects that were not possible using prior geometric acoustics system. Further, we also propose a novel source placement algorithm that significantly reduces the precomputation time for evaluating the diffraction kernels. Our method has been integrated with the Unreal game engine and is able to generate diffraction effects for complex objects interactively. Finally, we present a novel metric that perceptually characterizes the early and late reflections. Our novel metric ($P - Reverb$) relates the just-noticeable differnce (JND) of early reflections with the JND of late rerverberation in terms of the mean free path (μ) of the scene. We conduct two extensive, online user evaluations that establish the JNDs of early reflections and late reverberation in terms of mean-free path. We then relate the two JNDs to give us the $P - Reverb$ metric. The use of the metric shows significant speed up in the precomputation of late reverberation parameter such as (RT_{60}).

6.2 Limitations

In this section, we discuss the limitations of the proposed techniques. Our psychoacoustic evaluations of sound propagation algorithms were done in simple virtual environments. The choice of using simple virtual environments was made so that the approximate methods could work to their full potential. Since most virtual environments are significantly more complex, our results provide more of a baseline value for the approximate methods. The actual performance of the approximate methods in more complex environments is likely to be worse than what we observed. Our implementation of mode-aware sound propagation is limited to rigid objects and modal sounds. Moreover, the time complexity tends to increase with the mode frequency as higher-order basis functions are required to represent high-frequency radiation patterns. Further, the precomputation time required for the evaluating the basis functions is high and requires running high-frequency wave-based simulation. Our diffraction kernels framework, while offering many benefits over existing geometric acoustic methods, is less accurate than purely wave-based propagation methods. This approach is mainly designed for scenes with well-separated rigid objects, whose scattering behavior does not change at runtime. Further this framework does not take into account the inter-object interactions in the scene which might result in the approach not working well in certain scenarios. Also, our current formulation of diffraction kernels only considers the magnitude of the scattered field but ignores the phase. Our $P - Reverb$ metric may not work in totally open environments since the mean-free path computation is dependent on collisions in the environment. Further the metric is an approximation to a multi-dimensional complex function of source frequency, scene dimensions, material parameters, etc. and might reduce accuracy in complicated environments.

6.3 Future Work

There are many avenues of future work. The user evaluations can be extended in multiple ways. It would be interesting to vary the spectral content of the source in the diffraction experiment, since diffraction is a frequency dependent phenomenon and evaluate the subjects responses. The subjects distance from the source can also be varied and evaluated. The reverberation experiment could be verified by constructing real world rooms and using an actual sound source to verify the logarithmic relation of subjects responses to changing room size. Further, it would interesting to observe the effect of visuals on the diffraction experiment. The reverberation experiments too can be augmented with visuals to study the combined effect of sound and vision in assessing visual depth in virtual environments.

In order to accelerate the precomputation of our mode-aware sound propagation, we can integrate acceleration techniques such as mode compression and mode culling. It would also be useful to consider the radiation efficiency of each mode which might allow to further accelerate the precomputation and the runtime stages. It would be interesting to integrate non-rigid synthesized sounds, e.g., liquid and soft-body sounds into our framework. Further, it would interesting to integrate the method into a wave-based sound propagation system and evaluate the results.

The diffraction kernel framework can be improved by modeling other interactions such as first order surface scattering based on Kirchoff approximation or wave-based geometric acoustics to model inter-object wave interactions. The precomputation stage can be accelerated further by considering progressive mesh simplifications on the side not facing the sound source. We could also consider multiple concentric spheres around the object to represent different distances hence capturing the near and far-field scattering effects rather than just the far-field characteristics. It would be also useful to design approximate schemes that can also model phase, as that is needed for certain applications, such as seat-dip effects in concert halls. We would also like to perceptually evaluate our diffraction kernel approach in other applications such as social VR and telepresence, where it is important to simulate diffraction effects and generate smooth sound fields.

Finally, the $P - Reverb$ metric can benefit from more evaluations that take the multi-dimensional nature of the psychometric function (i.e., source frequency, scene dimension, material parameters, etc.) into account. We could also improve our user study design by including a broader range of room sizes, having a moving listener, and rendering the sound using spatialization. We would also like to incorporate the $P - Reverb$ metric in game engines to automatically evaluate the reverb zones in the scene.

BIBLIOGRAPHY

Abramowitz, M. and Stegun, I. A. (1972). *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*. Number 55. Courier Dover Publications.

Adrien, J.-M. (1991). The missing link: Modal synthesis. In *Representations of musical signals*, pages 269–298. MIT Press.

Aldrich, K. M., Hellier, E. J., and Edworthy, J. (2009). What determines auditory similarity? the effect of stimulus group and methodology. *The Quarterly Journal of Experimental Psychology*, 62(1):63–83.

Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C. (2001). The cipic hrtf database. In *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*, pages 99–102. IEEE.

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

Antani, L., Chandak, A., Savioja, L., and Manocha, D. (2012a). Interactive sound propagation using compact acoustic transfer operators. *ACM Trans. Graph.*, 31(1):7:1–7:12.

Antani, L., Chandak, A., Taylor, M., and Manocha, D. (2012b). Efficient finite-edge diffraction using conservative from-region visibility. *Applied Acoustics*, 73(3):218–233.

Arnoldi, W. E. (1951). The principle of minimized iterations in the solution of the matrix eigenvalue problem. *Quarterly of Applied Mathematics*, 9(1):17–29.

Barbone, P. E., Montgomery, J. M., Michael, O., and Harari, I. (1998). Scattering by a hybrid asymptotic/finite element method. *Computer methods in applied mechanics and engineering*, 164(1):141–156.

Barron, M. (1971). The subjective effects of first reflections in concert hallsthe need for lateral reflections. *Journal of sound and vibration*, 15(4):475–494.

Bate, A. and Pillow, M. (1947). Mean free path of sound in an auditorium. *Proceedings of the Physical Society*, 59(4):535.

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

Begault, D. R. et al. (1994). *3-D sound for virtual reality and multimedia*, volume 955. Citeseer.

Begault, D. R. and Trejo, L. J. (2000). 3-d sound for virtual reality and multimedia.

Berenger, J.-P. (1994). A perfectly matched layer for the absorption of electromagnetic waves. *Journal of computational physics*, 114(2):185–200.

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

Blauert, J. and Lindemann, W. (1986). Auditory spaciousness: Some further psychoacoustic analyses. *The Journal of the Acoustical Society of America*, 80(2):533–542.

Borish, J. (1984a). Extension of the image model to arbitrary polyhedra. *The Journal of the Acoustical Society of America*, 75(6):1827–1836.

Borish, J. (1984b). Extension of the image model to arbitrary polyhedra. *The Journal of the Acoustical Society of America*, 75(6):1827–1836.

Bradley, J., Sato, H., and Picard, M. (2003). On the importance of early reflections for speech in rooms. *The Journal of the Acoustical Society of America*, 113(6):3233–3244.

Bronkhorst, A. W. and Houtgast, T. (1999). Auditory distance perception in rooms. *Nature*, 397(6719):517.

Cabrera, D., Jeong, D., Kwak, H. J., Kim, J.-Y., and Duckjin-gu, J. (2005). Auditory room size perception for modeled and measured rooms. In *Internoise, the 2005 Congress and Exposition on Noise Control Engineering, Rio de Janeiro, Brazil, 7-10 August 2005*. Citeseer.

Cao, C., Ren, Z., Schissler, C., Manocha, D., and Zhou, K. (2016). Interactive sound propagation with bidirectional path tracing. *ACM Transactions on Graphics (TOG)*, 35(6):180.

Chadwick, J. N., An, S. S., and James, D. L. (2009). Harmonic shells: a practical nonlinear sound model for near-rigid thin shells. In *ACM Transactions on Graphics (TOG)*, volume 28, page 119. ACM.

Chandak, A., Lauterbach, C., Taylor, M., Ren, Z., and Manocha, D. (2008). Ad-frustum: Adaptive frustum tracing for interactive sound propagation. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1707–1722.

Cheng, A. H.-D. and Cheng, D. T. (2005). Heritage and early history of the boundary element method. *Engineering Analysis with Boundary Elements*, 29(3):268–302.

Chinnery, P. A., Humphrey, V. F., and Zhang, J. (1997). Low-frequency acoustic scattering by a cube: Experimental measurements and theoretical predictions. *The Journal of the Acoustical Society of America*, 101(5):2571–2582.

Christensen, C. and Koutsouris, G. (2013). Odeon manual, chapter 6.

Chu, D., Stanton, T. K., and Pierce, A. D. (2007). Higher-order acoustic diffraction by edges of finite thickness. *The Journal of the Acoustical Society of America*, 122(6):3177–3194.

Ciskowski, R. D. and Brebbia, C. A. (1991). *Boundary element methods in acoustics*. Computational Mechanics Publications Southampton, Boston.

Cohen-Steiner, D. and Morvan, J.-M. (2003). Restricted delaunay triangulations and normal cycle. In *Proceedings of the nineteenth annual symposium on Computational geometry*, pages 312–321. ACM.

Dindart, J.-F., Embrechts, J.-J., and Sémidor, C. (1999). Use of bidirectional reflectance distribution function in a particle tracing method. *The Journal of the Acoustical Society of America*, 105(2):1198–1198.

Djelani, T. and Blauert, J. (2001). Investigations into the build-up and breakdown of the precedence effect. *Acta Acustica united with Acustica*, 87(2):253–261.

Dubois, E., Gray, P., and Nigay, L. (2009). *The engineering of mixed reality systems*. Springer Science & Business Media.

Dubois, E., Gray, P., and Nigay, L., editors (2010). *The Engineering of Mixed Reality Systems*. Springer, London.

Durlach, N. and Mavor, A. (1995). *Virtual Reality Scientific and Technological Challenges*. National Academy Press.

Egelmeers, G. P. and Sommen, P. C. (1996). A new method for efficient convolution in frequency domain by nonuniform partitioning for adaptive filtering. *IEEE Transactions on signal processing*, 44(12):3123–3129.

Fastl, H. and Zwicker, E. (2007). *Psychoacoustics: Facts and models*, volume 22. Springer Science & Business Media.

Fletcher, H. (1953). Speech and hearing in communication.

Fletcher, H. and Galt, R. H. (1950). The perception of speech and its relation to telephony. *The Journal of the Acoustical Society of America*, 22(2):89–151.

Franinovic, K. and Serafin, S. (2013). *Sonic Interaction Design*. MIT Press.

Funkhouser, T., Carlbom, I., Elko, G., Pingali, G., Sondhi, M., and West, J. (1998a). A beam tracing approach to acoustic modeling for interactive virtual environments. In *Proc. of ACM SIGGRAPH*, pages 21–32.

Funkhouser, T., Carlbom, I., Elko, G., Pingali, G., Sondhi, M., and West, J. (1998b). A beam tracing approach to acoustic modeling for interactive virtual environments. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, pages 21–32. ACM.

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

Galster, J. A. (2007). *The Effect of Room Volume on Speech Recognition in Enclosures with Similar Mean Reverberation Time*. PhD thesis, Vanderbilt University.

Gescheider, G. A. (2013). *Psychophysics: the fundamentals*. Psychology Press.

Giguère, C. and Abel, S. M. (1993). Sound localization: Effects of reverberation time, speaker array, stimulus frequency, and stimulus rise/decay. *The Journal of the Acoustical Society of America*, 94(2):769–776.

Granier, E., Kleiner, M., Dalenbäck, B.-I., and Svensson, P. (1996). Experimental auralization of car audio installations. *Journal of the Audio Engineering Society*, 44(10):835–849.

Gumerov, N. A. and Duraiswami, R. (2009). A broadband fast multipole accelerated boundary element method for the three dimensional helmholtz equation. *The Journal of the Acoustical Society of America*, 125(1):191–205.

Haas, H. (1951). Über den einfluβ eines einfachechos auf die hörsamkeit von sprache. *Acta Acustica united with Acustica*, 1(2):49–58.

Hameed, S., Pakarinen, J., Valde, K., and Pulkki, V. (2004). Psychoacoustic cues in room size perception. In *Audio Engineering Society Convention 116*. Audio Engineering Society.

Hampel, S., Langer, S., and Cisilino, A. (2008). Coupling boundary elements to a raytracing procedure. *International journal for numerical methods in engineering*, 73(3):427–445.

Hartmann, W. M. (1983). Localization of sound in rooms. *The Journal of the Acoustical Society of America*, 74(5):1380–1391.

Hendrix, C. and Barfield, W. (1996). The sense of presence within auditory virtual environments. *Presence: Teleoperators & Virtual Environments*, 5(3):290–301.

Hughes, D. E., Thropp, J., Holmquist, J., and Moshell, J. M. (2006). Spatial perception and expectation: factors in acoustical awareness for mout training. In *Transformational Science And Technology For The Current And Future Force: (With CD-ROM)*, pages 339–343. World Scientific.

ISO, E. (2009). 3382-1, 2009, acousticsmeasurement of room acoustic parameterspart 1: Performance spaces,. *International Organization for Standardization, Brussels, Belgium*.

James, D. L., Barbič, J., and Pai, D. K. (2006a). Precomputed acoustic transfer: output-sensitive, accurate sound generation for geometrically complex vibration sources. In *ACM Transactions on Graphics (TOG)*, volume 25, pages 987–995. ACM.

James, D. L., Barbič, J., and Pai, D. K. (2006b). Precomputed acoustic transfer: output-sensitive, accurate sound generation for geometrically complex vibration sources. *ACM Transactions on Graphics (TOG)*, 25(3):987–995.

Jing, Y. and Xiang, N. (2008). Visualizations of sound energy across coupled rooms using a diffusion equation model. *The Journal of the Acoustical Society of America*, 124(6):EL360–EL365.

Jot, J.-M. and Chaigne, A. (1991). Digital delay networks for designing artificial reverberators. In *Audio Engineering Society Convention 90*. Audio Engineering Society.

Kapralos, B., Jenkin, M., and Milios, E. (2005). Acoustical modeling using a russian roulette strategy. In *Proceedings of the 118th Convention of the Audio Engineering Society*, pages 28–31.

Kawai, T. (1981). Sound diffraction by a many-sided barrier or pillar. *Journal of Sound and Vibration*, 79(2):229–242.

Kistler, D. J. and Wightman, F. L. (1992). A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *The Journal of the Acoustical Society of America*, 91(3):1637–1647.

Kleiner, M., Dalenbäck, B.-I., and Svensson, P. (1993). Auralization-an overview. *Journal of the Audio Engineering Society*, 41(11):861–875.

Knudsen, V. O. (1932). Architectural acoustics.

Kouyoumjian, R. G. and Pathak, P. H. (1974). A uniform geometrical theory of diffraction for an edge in a perfectly conducting surface. *November*, 88:1448–1461.

Krokstad, A., Strom, S., and Sørsdal, S. (1968a). Calculating the acoustical room response by the use of a ray tracing technique. *Journal of Sound and Vibration*, 8(1):118–125.

Krokstad, A., Strom, S., and Sorsdal, S. (1968b). Calculating the acoustical room response by the use of a ray tracing technique. *Journal of Sound and Vibration*, 8(1):118–125.

Kuttruff, H. (2007). *Acoustics: An Introduction*. CRC Press.

Kuttruff, H. (2016). *Room acoustics*. Crc Press.

Lam, Y. W. (2005). Issues for computer modelling of room acoustics in non-concert hall settings. *Acoustical science and technology*, 26(2):145–155.

Larsson, P., Vastfjall, D., and Kleiner, M. (2002a). Better presence and performance in virtual environments by improved binaural sound rendering. In *Audio Engineering Society Conference: 22nd International Conference: Virtual, Synthetic, and Entertainment Audio*. Audio Engineering Society.

Larsson, P., Vastfjall, D., and Kleiner, M. (2002b). Better presence and performance in virtual environments by improved binaural sound rendering. In *Virtual, Synthetic, and Entertainment Audio conference*.

Lentz, T., Schröder, D., Vorländer, M., and Assenmacher, I. (2007). Virtual reality system with integrated sound field simulation and reproduction. *EURASIP Journal on Advances in Singal Processing*, 2007:187–187. Article ID 70540, 19 pages.

Liu, Q. H. (1997). The pseudospectral time-domain (pstd) method: A new algorithm for solutions of maxwell's equations. In *Antennas and Propagation Society International Symposium, 1997. IEEE., 1997 Digest*, volume 1, pages 122–125. IEEE.

Liu, Y. and Nishimura, N. (2006). The fast multipole boundary element method for potential problems: a tutorial. *Engineering Analysis with Boundary Elements*, 30(5):371–381.

Lokki, T., Southern, A., Siltanen, S., and Savioja, L. (2011). Studies of epidaurus with a hybrid room acoustics modelling method. *Acoustics of Ancient Theaters Patras, Greece*.

Löllmann, H. W. and Vary, P. (2008). Estimation of the reverberation time in noisy environments. In *Proc. of Intl. Workshop on Acoustic Echo and Noise Control (IWAENC)*.

Mehra, R., Antani, L., Kim, S., and Manocha, D. (2014a). Source and listener directivity for interactive wave-based sound propagation. *IEEE transactions on visualization and computer graphics*, 20(4):495–503.

Mehra, R., Antani, L., Kim, S., and Manocha, D. (2014b). Source and listener directivity for interactive wave-based sound propagation. *IEEE Transactions on Visualization and Computer Graphics*, 19(4):567–575.

Mehra, R., Raghuvanshi, N., Antani, L., Chandak, A., Curtis, S., and Manocha, D. (2013). Wave-based sound propagation in large open scenes using an equivalent source formulation. *ACM Transactions on Graphics (TOG)*, 32(2):19.

Mehra, R., Raghuvanshi, N., Savioja, L., Lin, M. C., and Manocha, D. (2012). An efficient gpu-based time domain solver for the acoustic wave equation. *Applied Acoustics*, 73(2):83–94.

Mehra, R., Rungta, A., Golas, A., Lin, M., and Manocha, D. (2015). Wave: Interactive wave-based sound propagation for virtual environments. *IEEE transactions on visualization and computer graphics*, 21(4):434–442.

Mershon, D. H. and King, L. E. (1975). Intensity and reverberation as factors in the auditory perception of egocentric distance. *Perception & Psychophysics*, 18(6):409–415.

Meshram, A., Mehra, R., Yang, H., Dunn, E., Franm, J.-M., and Manocha, D. (2014). P-hrtf: Efficient personalized hrtf computation for high-fidelity spatial sound. In *Mixed and Augmented Reality (ISMAR), 2014 IEEE International Symposium on*, pages 53–61. IEEE.

Mitra, N. J., Guibas, L. J., and Pauly, M. (2006). Partial and approximate symmetry detection for 3d geometry. In *ACM Transactions on Graphics (TOG)*, volume 25, pages 560–568. ACM.

Moorer, J. A. (1979). About this reverberation business. *Computer music journal*, pages 13–28.

Moss, W., Yeh, H., Hong, J.-M., Lin, M. C., and Manocha, D. (2010). Sounding liquids: Automatic sound synthesis from fluid simulation. *ACM Transactions on Graphics (TOG)*, 29(3):21.

Nordahl, R., Serafin, S., and Turchet, L. (2010). Sound synthesis and evaluation of interactive footsteps for virtual reality applications. *Proc. of IEEE VR*, pages 147–153.

O'Brien, J. F., Shen, C., and Gatchalian, C. M. (2002a). Synthesizing sounds from rigid-body simulations. In *Proceedings of the 2002 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pages 175–181. ACM.

O'Brien, J. F., Shen, C., and Gatchalian, C. M. (2002b). Synthesizing sounds from rigid-body simulations. In *The ACM SIGGRAPH 2002 Symposium on Computer Animation*, pages 175–181. ACM Press.

Ochmann, M. (1999). The full-field equations for acoustic radiation and scattering. *The Journal of the Acoustical Society of America*, 105(5):2574–2584.

Painter, T. and Spanias, A. (2000). Perceptual coding of digital audio. *Proceedings of the IEEE*, 88(4):451–515.

Pierce, A. D. et al. (1981). *Acoustics: an introduction to its physical principles and applications*. McGraw-Hill New York.

Polk, T. A., Behensky, C., Gonzalez, R., and Smith, E. E. (2002). Rating the similarity of simple perceptual stimuli: asymmetries induced by manipulating exposure frequency. *Cognition*, 82(3):B75–B88.

Pop, C. B. and Cabrera, D. (2005). Auditory room size perception for real rooms. In *Proceedings of the Australian Acoustical Society Conference*.

Raghuvanshi, N., Allen, A., and Snyder, J. (2016). Numerical wave simulation for interactive audio-visual applications. *The Journal of the Acoustical Society of America*, 139(4):2008–2009.

Raghuvanshi, N. and Lin, M. C. (2006). Interactive sound synthesis for large scale environments. In *Proceedings of the 2006 symposium on Interactive 3D graphics and games*, pages 101–108. ACM.

Raghuvanshi, N., Narain, R., and Lin, M. C. (2009). Efficient and accurate sound propagation using adaptive rectangular decomposition. *IEEE Transactions on Visualization and Computer Graphics*, 15(5):789–801.

Raghuvanshi, N. and Snyder, J. (2014). Parametric wave field coding for precomputed sound propagation. *ACM Transactions on Graphics (TOG)*, 33(4):38.

Raghuvanshi, N., Snyder, J., Mehra, R., Lin, M., and Govindaraju, N. (2010). Precomputed wave simulation for real-time sound propagation of dynamic sources in complex scenes. In *ACM Transactions on Graphics (TOG)*, volume 29, page 68. ACM.

Rakerd, B. and Hartmann, W. (1985). Localization of sound in rooms, ii: The effects of a single reflecting surface. *The Journal of the Acoustical Society of America*, 78(2):524–533.

Ratnam, R., Jones, D. L., Wheeler, B. C., OBrien Jr, W. D., Lansing, C. R., and Feng, A. S. (2003). Blind estimation of reverberation time. *The Journal of the Acoustical Society of America*, 114(5):2877–2892.

Rayleigh, L. (1907). Xii. on our perception of sound direction. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 13(74):214–232.

Ren, Z., Mehra, R., Coposky, J., and Lin, M. C. (2012). Tabletop ensemble: touch-enabled virtual percussion instruments. In *Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, pages 7–14. ACM.

Ren, Z., Yeh, H., Klatzky, R., and Lin, M. C. (2013). Auditory perception of geometry-invariant material properties. *Visualization and Computer Graphics, IEEE Transactions on*, 19(4):557–566.

Ren, Z., Yeh, H., and Lin, M. C. (2010). Synthesizing contact sounds between textured models. In *Virtual Reality Conference (VR), 2010 IEEE*, pages 139–146. IEEE.

Richards, D. G. and Wiley, R. H. (1980). Reverberations and amplitude fluctuations in the propagation of sound in a forest: implications for animal communication. *The American Naturalist*, 115(3):381–399.

Rocchesso, D., Serafin, S., Behrendt, F., Bernardini, N., Bresin, R., Eckel, G., Franinovic, K., Hermann, T., Pauletto, S., Susini, P., and Visell, Y. (2008). Sonic interaction design: Sound, information and experience. *Proc. of ACM SIGCHI*, pages 3969–3972.

Román, S. R. M., Svensson, U. P., Šlechta, J., and Smith, J. O. (2016). A hybrid method combining the edge source integral equation and the boundary element method for scattering problems. *The Journal of the Acoustical Society of America*, 139(4):2202–2202.

Rothbaum, B. O., Hodges, L., Alarcon, R., Ready, D., Shahar, F., Graap, K., Pair, J., Hebert, P., Gotz, D., Wills, B., et al. (1999). Virtual reality exposure therapy for ptsd vietnam veterans: A case study. *Journal of Traumatic Stress: Official Publication of The International Society for Traumatic Stress Studies*, 12(2):263–271.

Rungta, A., Rewkowski, N., Klatzky, R., Lin, M., and Manocha, D. (2017). Effects of virtual acoustics on dynamic auditory distance perception. *The Journal of the Acoustical Society of America*, 141(4):EL427–EL432.

Rungta, A., Rust, S., Morales, N., Klatzky, R., Lin, M., and Manocha, D. (2016). Psychoacoustic characterization of propagation effects in virtual environments. *ACM Transactions on Applied Perception (TAP)*, 13(4):21.

Savioja, L. and Svensson, U. P. (2015). Overview of geometrical room acoustic modeling techniques. *The Journal of the Acoustical Society of America*, 138(2):708–730.

Schissler, C. and Manocha, D. (2015). Interactive sound propagation and rendering for large multi-source scenes. Technical report, Department of Computer Science, University of North Carolina at Chapel Hill.

Schissler, C. and Manocha, D. (2016). Interactive sound propagation and rendering for large multi-source scenes. *ACM Transactions on Graphics (TOG)*, 36(1):2.

Schissler, C. and Manocha, D. (2017). Interactive sound propagation and rendering for large multi-source scenes. *ACM Transactions on Graphics (TOG)*, 36(1):2.

Schissler, C., Mehra, R., and Manocha, D. (2014a). High-order diffraction and diffuse reflections for interactive sound propagation in large environments. *ACM Transactions on Graphics (TOG)*, 33(4):39.

Schissler, C., Mehra, R., and Manocha, D. (2014b). High-order diffraction and diffuse reflections for interactive sound propagation in large environments. *ACM Trans. Graph.*, 33(4):39:1–39:12.

Schröder, D. (2011). *Physically based real-time auralization of interactive virtual environments*, volume 11. Logos Verlag Berlin GmbH.

Schroeder, M. R. (1962). Natural sounding artificial reverberation. *Journal of the Audio Engineering Society*, 10(3):219–223.

Schroeder, M. R. and Logan, B. F. (1961). " colorless" artificial reverberation. *IRE Transactions on Audio*, 9(6):209–214.

Serafin, S. (2004). *The Sound OF Friction: Real-Time Models, Playability and Musical Applications*. PhD thesis, Stanford University.

Shilling, R. D. and Shinn-Cunningham, B. (2002). Virtual auditory displays. *Handbook of virtual environment technology*, pages 65–92.

Siltanen, S., Lokki, T., Kiminki, S., and Savioja, L. (2007). The room acoustic rendering equation. *The Journal of the Acoustical Society of America*, 122(3):1624–1635.

Skålevik, M. (2010). Reverberation time–the mother of all room acoustic parameters. In *CD Proceedings of 20th International Congress on Acoustic, ICA*, volume 10.

Sloan, P.-P. (2013). Efficient spherical harmonic evaluation. *Journal of Computer Graphics Techniques (JCGT)*, 2(2):84–83.

Southern, A., Siltanen, S., and Savioja, L. (2011). Spatial room impulse responses with a hybrid modeling method. In *Audio Engineering Society Convention 130*. Audio Engineering Society.

Steinicke, F., Visell, Y., Campos, J., and Lcuyer, A. (2015). *Human Walking in Virtual Environments: Perception, Technology, and Applications*.

Stephenson, U. M. (2010). An energetic approach for the simulation of diffraction within ray tracing based on the uncertainty relation. *Acta Acustica united with Acustica*, 96(3):516–535.

Svensson, U. P., Fred, R. I., and Vanderkooy, J. (1999). An analytic secondary source model of edge diffraction impulse responses. *The Journal of the Acoustical Society of America*, 106(5):2331–2344.

Taflove, A. and Hagness, S. C. (2005). *Computational electrodynamics: the finite-difference time-domain method*. Artech house.

Taylor, M., Chandak, A., Antani, L., and Manocha, D. (2009a). Resound: interactive sound rendering for dynamic virtual environments. In *MM '09: Proceedings of the seventeen ACM international conference on Multimedia*, pages 271–280, New York, NY, USA. ACM.

Taylor, M., Chandak, A., Mo, Q., Lauterbach, C., Schissler, C., and Manocha, D. (2012). Guided multiview ray tracing for fast auralization. *IEEE Transactions on Visualization and Computer Graphics*, 18:1797–1810.

Taylor, M. T., Chandak, A., Antani, L., and Manocha, D. (2009b). Resound: interactive sound rendering for dynamic virtual environments. In *Proceedings of the 17th ACM international conference on Multimedia*, pages 271–280. ACM.

Terhardt, E. (1979). Calculating virtual pitch. *Hearing research*, 1(2):155–182.

Torres, R., Kleiner, M., Svensson, U., and Dalenbäck, B. (2001a). Edge diffraction and surface scattering in auralization. *Proceedings of SIGGRAPH Campfire*.

Torres, R. R. and Kleiner, M. (1998). Audibility of edge diffraction in auralization of a stage house. *The Journal of the Acoustical Society of America*, 103(5):2789–2789.

Torres, R. R., Svensson, U. P., and Kleiner, M. (2001b). Computation of edge diffraction for more accurate room acoustics auralization. *The Journal of the Acoustical Society of America*, 109(2):600–610.

Tsingos, N., Dachsbacher, C., Lefebvre, S., and Dellepiane, M. (2007). Instant sound scattering. In *Proceedings of the Eurographics Symposium on Rendering*, pages 111–120.

Tsingos, N., Funkhouser, T., Ngan, A., and Carlbom, I. (2001a). Modeling acoustics in virtual environments using the uniform theory of diffraction. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 545–552. acm.

Tsingos, N., Funkhouser, T., Ngan, A., and Carlbom, I. (2001b). Modeling acoustics in virtual environments using the uniform theory of diffraction. In *SIGGRAPH 2001, Computer Graphics Proceedings*, pages 545–552.

Tsingos, N., Funkhouser, T., Ngan, A., and Carlbom, I. (2001c). Modeling acoustics in virtual environments using the uniform theory of diffraction. In *Proc. of ACM SIGGRAPH*, pages 545–552.

Tsingos, N., Gallo, E., and Drettakis, G. (2004). Perceptual audio rendering of complex virtual environments. *ACM Trans. Graph.*, 23(3):249–258.

Tsingos, N. and Gascuel, J.-D. (1998). Fast rendering of sound occlusion and diffraction effects for virtual acoustic environments. In *Audio Engineering Society Convention 104*. Audio Engineering Society.

Turchet, L. (2015). Designing presence for real locomotion in immersive virtual environments: an affordance-based experiential approach. *Virtual Reality*, 19(3-4):277–290.

Turchet, L., Spagnol, S., Geronazzo, M., and Avanzini, F. (2015). Localization of self-generated synthetic footstep sounds on different walked-upon materials through headphones. *Virtual Reality*, pages 1–16.

Valimaki, V., Parker, J. D., Savioja, L., Smith, J. O., and Abel, J. S. (2012). Fifty years of artificial reverberation. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(5):1421–1448.

van den Doel, K., Kry, P. G., and Pai, D. K. (2001). Foleyautomatic: physically-based sound effects for interactive simulation and animation. In *Proc. of ACM SIGGRAPH*, pages 537–544.

Vickers, G. (1996). The projected areas of ellipsoids and cylinders. *Powder technology*, 86(2):195–200.

Visell, Y., Fontana, F., Giordano, B. L., Nordahl, R., Serafin, S., and Bresin, R. (2009). Sound design and perception in walking interactions. 67(11):947–959.

Von Helmholtz, H. (1912). *On the Sensations of Tone as a Physiological Basis for the Theory of Music*. Longmans, Green.

Vorländer, M. (1989). Simulation of the transient and steady-state sound propagation in rooms using a new combined ray-tracing/image-source algorithm. *The Journal of the Acoustical Society of America*, 86(1):172–178.

Vorländer, M. (2000). Room acoustical simulation algorithm based on the free path distribution. *Journal of sound and vibration*, 232(1):129–137.

Vorländer, M. and Bietz, H. (1994). Comparison of methods for measuring reverberation time. *Acta Acustica united with acústica*, 80(3):205–215.

Wang, Y., Safavi-Naeini, S., and Chaudhuri, S. K. (2000). A hybrid technique based on combining ray tracing and fdtd methods for site-specific modeling of indoor radio wave propagation. *IEEE Transactions on antennas and propagation*, 48(5):743–754.

Webb, C. and Gray, A. (2013). Large-scale virtual acoustics simulation at audio rates using three dimensional finite difference time domain and multiple graphics processing units. In *Proceedings of Meetings on Acoustics*, volume 19, page 070092. Acoustical Society of America.

Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L. (1993). Localization using nonindividualized head-related transfer functions. *The Journal of the Acoustical Society of America*, 94(1):111–123.

Wilson, J., Walker, B. N., Lindsay, J., Cambias, C., and Dellaert, F. (2007). Swan: System for wearable audio navigation. In *Wearable Computers, 2007 11th IEEE International Symposium on*, pages 91–98. IEEE.

Wolfe, J., Levi, D., Kluender, K., Bartoshuk, L., Herz, R., Klatzky, R., Lederman, S., and Merfeld, D. (2014). *Sensation and Perception (4th ed.)*. Sinauer associates, inc.

Yee, K. (1966). Numerical solution of initial boundary value problems involving maxwell's equations in isotropic media. *IEEE Transactions on antennas and propagation*, 14(3):302–307.

Yeh, H., Mehra, R., Ren, Z., Antani, L., Manocha, D., and Lin, M. (2013). Wave-ray coupling for interactive sound propagation in large complex scenes. *ACM Transactions on Graphics (TOG)*, 32(6):165.

Young, D. and Serafin, S. (2003). Playability evaluation of a virtual bowed string instrument. *Proc. of Conference on New Interfaces for Musical Expression*, pages 104 – 108.

Zahorik, P. (2002). Assessing auditory distance perception using virtual acoustics. *The Journal of the Acoustical Society of America*, 111(4):1832–1846.

Zahorik, P., Brungart, D. S., and Bronkhorst, A. W. (2005). Auditory distance perception in humans: A summary of past and present research. *Acta Acustica united with Acustica*, 91(3):409–420.

Zahorik, P. and Wightman, F. L. (2001). Loudness constancy with varying sound source distance. *Nature neuroscience*, 4(1):78.

Zannini, C. M., Parisi, R., and Uncini, A. (2011). Binaural sound source localization in the presence of reverberation. In *Digital Signal Processing (DSP), 2011 17th International Conference on*, pages 1–6. IEEE.

Zheng, C. and James, D. L. (2009). Harmonic fluids. *ACM Trans. Graph.*, 28(3):1–12.

Zheng, C. and James, D. L. (2011). Toward high-quality modal contact sound. *ACM Transactions on Graphics (TOG)*, 30(4):38.

Zienkiewicz, O. C. (2005). *The finite element method for fluid dynamics*. PhD thesis, University of Wales.

Zienkiewicz, O. C., Morgan, K., and Morgan, K. (2006). *Finite elements and approximation*. Courier Corporation.

Zwislocki, J. and Goodman, D. (1980). Absolute scaling of sensory magnitudes: A validation. *Perception & Psychophysics*, 28(1):28–38.