Realistic real-time audio rendering in virtual environments

Matthias Labschütz*
TU Vienna

Abstract

Spatial audio rendering plays an important role in simulating realistic virtual environments. This report gives an overview. It contains an introduction to the physical background of sound and the psychophysical background of sound. Spatial audio in current applications such as video games is reviewed and an overview of acoustical simulations and applications with special interest in real-time application is given. Numerical and geometric methods are presented and compared.

Keywords: real-time spatial audio, acoustics, room acoustics, psychoacoustics, numerical methods, geometric methods, finite difference method, digital waveguide mesh, image source method, ray tracing, beam tracing, phonon tracing, frustum tracing, GPU audio

1 Introduction

Sight, hearing, taste, smell and touch, five senses - two of them can be tricked efficiently with todays computers in virtual environments. On the one hand this can be done by video, normally displayed on a screen, on the other hand by audio played over speakers or headphones. Sight is usually covered in the field of computer graphics while sound is somehow between a few other fields.

The main field of sound is acoustics, divided into physical and biological acoustics and acoustic engineering. For simulating acoustics in virtual environments the fields of room acoustics and psychoacoustics are of special interest. Many audio rendering techniques are similar to graphical rendering, therefore the field is also covered by computer graphics.

This report should give an overview of the different physical and psychophysical views on the subject. It tries to give a little insight on where and how real time acoustical simulations are used. Finally, based on previous chapters, it will summarize current methods to simulate spatial sound in virtual environments, as found in research papers and research implementations. Namely numerical and geometric methods will be described.

The ultimate goal is to bring audio rendering to a point where actual geometry is used to calculate an audible impression of a virtual scene. This impression should come very close to that of a real scene. In addition, the method should be able to simulate a complex, dynamic world at interactive rates.

For virtual environments dynamic sound sources and listener positions may be more important than dynamic geometry, since most current usual virtual environments consist of a large amount of static objects. For room acoustics, fast dynamic changes in geometry may be interesting to check the influence of a certain change in a rooms design.

*e-mail: e8971103@student.tuwien.ac.at

1.1 Sound rendering methods

The problem was addressed by many different methods.

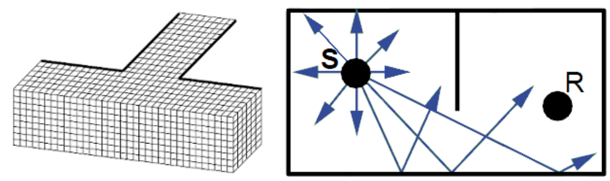

Figure 1: left: A numerical method discretizing space in cubes, right: Geometric ray casting with a source S and a listener R

Numerical methods (Figure 1 left) are physically accurate solutions of the differential wave equation, which can be achieved by numerous mathematical methods. Solutions require boundary and initial conditions. Boundary conditions specify the borders of the region of interest and initial conditions are used to model the locations of sound sources in the area. They are often slow in computation times, but allow to model many physical effects accurately. These effects may significantly increase realism of sound in virtual environments as shown in 'Precomputed Wave Simulation for Real-Time Sound Propagation of Dynamic Sources in Complex Scenes' [Raghuvanshi et al. 2010]. Numerical methods are strong for lower frequencies, because low frequencies only require lower computation resolutions, as explained in section 5.1 about Numerical Methods. Some more promising and recent solutions are described in [Raghuvanshi et al. 2010] and 'Accelerated wave-based acoustics simulation' [Tsingos 2009]. These examples of numerical methods are based on precomputation and allow moving listeners and sound sources. Their way of precomputation causes only static scenes to be considered in the first place. Dynamic scenes are hard to manage with numerical methods, because modeling the boundary conditions and solving the equation may result in high computation time.

Geometric solutions (Figure 1 right) treat acoustic waves like optical rays. They are fast but initially disregard the important effect of diffraction explained in the physical section about diffraction 2.2.3. They also introduce sampling artifacts. Still geometric approaches are widely used because of their simplicity, flexibility and speed. Early image source methods like in 'Extension of the image model to arbitrary polyhedra' [Borish 1984] tend to be only efficient for simple and rectangular scenes. Simple ray tracing approaches suffer from sampling artifacts and therefore require a lot of rays to be computed. First promising results were achieved with beam tracing, where a more recent algorithm is described in 'A beam tracing approach to acoustic modeling for interactive virtual environments'

[Funkhouser et al. 1998]. This method requires precomputation and only allows static sound sources. Another approach is found in 'AD-Frustum: Adaptive Frustum Tracing for Interactive Sound Propagation' [Chandak et al. 2008]. It allows dynamic scenes, listeners and sources, but may be slower than other methods and still suffers from sampling artifacts. Geometric solutions are explained in section 5.2 about Geometric Methods.

1.2 Categorization of scenes for sound propagation

When looking at possible scenes for calculating sound propagation, properties of these scenes can be defined as:

- dynamic vs. static environment
- enclosed vs. open environment
- organic vs. flat shapes

These attributes may be more important for certain types of scenes and therefore different methods of calculating sound propagation may be more useful to use.

In static environments of limited size precomputation is often used in virtual environments. Most scenes are still consisting of a huge amount of static geometry. Dynamic geometry is often restricted to certain objects. Precomputation is already used in lighting (e.g. lightmaps), visibility (e.g. portals for visibility) or generation of waypoints for artificial characters. For sound propagation an approach of precomputation may be sufficient as well. In addition it would be possible to add dynamic geometry effects for certain objects like doors, to increase realism. This is already used in some newer games as shown in section 4.

Two exemplary scenes used for enhanced audio rendering may be: A human made city or room scenario which are categorized as static, enclosed environments with flat shapes. A natural outdoor environment with the opposite properties. These are just two examples, a city may also include organic shapes and will include dynamic geometry. On the other hand also rough terrain like mountains, cliffs or caves are rather similar to the city example.

The city scenario is characterized by the existence of many large plane reflectors like walls. The areas are often of limited volumetric size (e.g. a room). Transmission through thinner materials is to be considered, but often of lower relevance. Noticeable reverberation occurs in these scenarios. Reflection but also diffraction have a great impact on the perceived sound. These scenarios can be computed more easily with current methods, since they evolved from the field of room acoustics and it was of major concern to simulate sound propagation in human-made structures. Finding planes or large flat faces in outdoor scenes is a lot harder than in man-made environments. Man made environments can also often make use of visibility simplifications. Many games take their computational capability from the fact that scenarios exist in enclosed space. Precomputation is of greater relevance in such an environment if it is also static, making the algorithm of [Raghuvanshi et al. 2010] more interesting.

The outdoor scenario on the other hand provide different characteristics. In a scene like that, simple reflectors are non-existent. Therefore reverberation plays a minor role. It is hardly possible to create any simplification due to visibility, except culling the scene at a certain distance. Therefore these scenarios are not very suitable for precomputation for visibility. Precomputation for sound could be possible with the use of a much coarser grid size, but this would prevent exact direction of sound signals to be perceivable.

Outdoor scenes are normally dominated by vegetation like trees or grass. These objects are on the one hand not static, on the other hand many phenomena that can be easily computed for flat shapes can not be applied directly to plants, namely reflection. Reflection requires flat faces of a large size in comparison to the wavelength. Since most common ray tracing methods mainly consider reflections they are also not suitable to handle outdoor scenes efficiently. It could be of further interest to study the influence of vegetation like trees or grass on the propagation of sound. Are there any general simplifications to be made when propagating sound through grass or leaves?

Atmospheric temperature change and the caused refraction are also of higher interest in large outdoor scenes.

In the field of spatial sound simulation for room acoustics normally human made scenarios are of main interest. Luckily many virtual environments and game engines follow the same principle (e.g. Quake Engine, Source Engine). So room acoustic methods could be applied directly, or with reduced accuracy. This would require stable physical simulations even for low grid sizes and a low amount of rays. Stable in this context mean that no additional artifacts occur when reducing calculation accuracy.

Another aspect of special interest in sound propagation, is the frequency dependency of phenomena. It is known that refraction and diffraction are frequency dependent. Reflection depends on frequencies since different materials absorb different frequencies more or less, and the amount of diffuse reflection on a material is also frequency dependent. On relatively small objects no reflection of low frequencies occurs.

A closer observation of the relevance of frequency dependent simulations to acoustical perception could be useful. This could be done by evaluating real scenes or with use of auditory experiments. Specially in the field of noise control it is known that low frequencies can travel more easily through materials or around objects. This effect will most likely also have a great influence on calculating realistic sound propagation.

1.3 Overview

Advanced spatial audio rendering in virtual environments, specially in video games, can improve the realism of these simulations. It also provides a solution to the problem of monotonous sounds, by simulating acoustical properties more precisely. Sounds will not seem that repetitive, if they sound different in all different positions within a room or a scene [Piringer 2010]. Due to improved realism, also the grade of immersion may increase significantly.

I strongly recommend to listen to some examples of the following projects, since those results can not be added to a visual report: *Accelerated wave-based acoustics simulation* [Raghuvanshi et al. 2010], *AD-Frustum* [Chandak et al. 2008], *Efficient and Accurate Sound Propagation Using Adaptive Rectangular Decomposition* [Raghuvanshi et al. 2009] and *RESound* [Taylor et al. 2009].

The following work will be categorized in four chapters. The first chapter (2) will be about the basic physics of sound and wave propagation starting from the very basics. The second chapter (3) will summarize selected psychophysical phenomena and tries to give an overview of the psychophysics of sound. The third chapter (4) picks out a few examples of virtual environments using spatial audio and the last chapter (5) will explain numerical and geometric methods of calculating sound propagation.

2 Physics of Sound

This section gives a brief overview of the physical models of sound waves (2.1), sound propagation (2.2) and some other acouustic effects (2.3). This lies mainly within the research fields of acoustics and room acoustics. Wave propagation is an important physical problem as found in the propagation of light, sound or radio waves in air or other gases. Wave propagation can also be observed traveling through liquids and solid materials. These different cases are for example used in seismology or ultrasonic techniques. For the purpose of simulating sound in virtual environments the propagation of waves in air is our main concern.

2.1 Wave models

Figure 2: Waves can be found in water and wheat

Waves are a disturbance traveling through a medium. Imagine throwing a stone in a pond, waving wheat or sound. On throwing a stone in a pond circular waves will be emitted around the point of impact. The movement along the direction of propagation of a wave does not make the medium flow or move constantly in a direction. Imagine the wheat moving, even if it moves and pushes its neighbors, it will stay at the spot it grows at. (Figure 2) So not the medium is transported along the direction of propagation, but the medium pushes and falls back to its original position.

Waves can be separated into transverse waves and longitudinal waves. Transverse waves are moving in perpendicular direction to the direction of propagation. For example water waves on the surface are transverse waves and therefore moving up and down on the surface. Another example would be the movement of the string of a musical instrument. On the other hand there are longitudinal waves. They describe a movement along the direction of the propagation and can be seen as pressure waves of low and high pressure along the direction of the propagation. One example for longitudinal waves would be sound traveling through air.

The following chapter will first concentrate on wave models and oscillations, then some important effects of wave propagation are explained. It is meant to explain all that is needed to understand the preceding sections, in specific section (5). It will give an introduction to notations, as well as an explanation of physical phenomena that have to be modeled to simulate realistic sound propagation.

2.1.1 Harmonic oscillation

In a simplified model we only consider continuous harmonic waves, which are waves with constant amplitude and frequency.

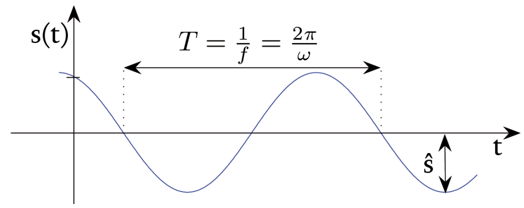

Figure 3: Harmonic oscillation with amplitude $\hat{s}$, period or wavelength T [wikipedia]

Waves can be described with cosine (or sine) functions. Such a function can be negative or positive. A longitudinal wave would describe an overpressure followed by an under-pressure and so forth, this corresponds to a positive pressure followed by a negative pressure. This way waves travel through a medium. The most simple way to describe a continuous harmonic wave mathematically is [Kuttruff 2007]:

$$s(t) = \hat{s}cos(\omega t + \varphi) \tag{1}$$

Where $\hat{s}$ is the amplitude, $\omega = 2\pi f$ the angular frequency (where f is the frequency), t the time and φ the phase angle (describing an offset of the original cosine function from the coordinate origin). A wave function is shown in Figure 3. In this representation the shown function is in a way ambiguous. It can be either seen as the amplitude of a certain point over time. Otherwise if we would define t as a location in space, the same function would describe the shape of a wave at a fixed time step as it is occurring in space.

When considering a sound wave the relation between the wave parameters and the hearing result is: The frequency of a sound wave refers to the pitch of a sound. A low frequency will be audible as low pitch, a high frequency as high pitch. The amplitude of sound is in relation to the loudness perceived. The impact of sound on human perception will be discussed in the chapter about perception.

An alternative, widely used expression for harmonic oscillations can be written with use of a complex notation also called phase vector (shown in Figure 4):

$$s(t) = Re\{\hat{s}e^{j(\omega t+\varphi)}\} \tag{2}$$

where Re is the real part of the term in brackets. The complex 'phasor' moves in circles with time t. The real part $Re\{s\}$ of the 'phasor' is the amplitude of the wave at the time t. In simplified notation, by omitting the Re symbol, as:

$$s(t) = \hat{s}e^{j(\omega t+\varphi)}. \tag{3}$$

It can be easily shown that equation (1) equals equation (2) by inserting into Euler's formula:

$$e^{jz} = cosz + jsinz \tag{4}$$

with only considering the left real part. (j is the complex number sometimes written as i.)

One advantage of that notation is that differentiating with respect to time only requires a multiplication by $j\omega$. Figure 4 shows the complex notation.

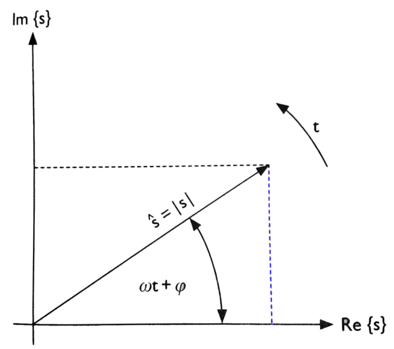

Figure 4: Complex 'phasor' of a harmonic oscillation [Kuttruff 2007]

2.1.2 Fourier analysis

Natural waves are normally not pure cosine waves. Often multiple waves with different wavelength (or frequency), amplitude and phase offset are overlaying. These underlaying waves are added up, like mentioned earlier to form a more complex shape. The process of breaking these waves up into their underlaying frequencies can be achieved by Fourier transformation. The original function is called the time function. The resulting function after Fourier transform is called the frequency spectrum. The frequency spectrum of a saw tooth approximating function is shown in 5.

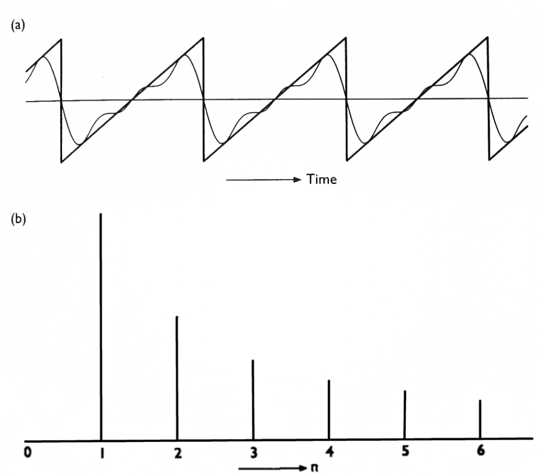

Figure 5: a) periodic saw tooth time function with an approximating function of periodic Fourier components b) discrete frequency spectrum of the approximating function

For a periodic oscillation, the frequency spectrum will be discrete. The Fourier transform that has to be used in this case is called discrete Fourier transform. For non-periodic signals the frequency spectrum is continuous. Non-periodic signals can only be transformed with the use of the continuous Fourier transformation if the integral over the function is not infinite, which means that the Cauchy principal value has to exist [Drmota et al. 2008]. Signals that never vanish and that are non periodic, also called signals without existing Cauchy principal value, can not be transformed with the continuous Fourier transformation. The Cauchy principal value has solutions for certain improper integrals that are ill defined when using the usual Riemann integral. To transform a function with infinite Cauchy principal value, only a finite time interval of the time function can be considered to create the frequency spectrum.

Material	Temperature (°C)	Sound velocity (m/s)	wavelength (at 440 Hz in cm)
Helium	0	965	219.3
Oxygen	0	316	71.8
Air	0	331	75.2
Air	20	343	78.0
Water	20	1483	337.0

Table 1: Propagation speeds in different mediums [Kuttruff 2007]

The Fourier series can be written as:

$$S_f(t) = \sum_{k=-\infty}^{\infty} c_k e^{jk\omega t} = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n cos(n\omega t) + b_n sin(n\omega t)) \quad (5)$$

where a_n, b_n and c_n are the Fourier coefficients, which can be calculated and converted into each other by certain formulas.

When comparing these series with equation (1) it can be seen that the Fourier series is a sum of harmonic oscillations. The phase offset in equation (1) is not considered. So for modeling a harmonic oscillation with phase offset a combination a sine and a cosine function has to be used. Also a connection to the complex phase vector notation (2) can be seen. The sum of the Fourier series also includes negative values for k, these negative values would correspond to a wave with negative angular frequency or frequency. There is no real physical explanation for this, instead it is required to make the complex part of the equation vanish, that occurs when using Euler's formula to express the exponential function as a sum of sine and cosine functions. For a non-complex wave the negative an positive parts are mirrored with respect to 0 and therefore vanish. The functions geometric representation would be the sum of spinning phasors, rotating with k multiples of ω. These phasors are mirrored along the imaginary origin, in a way that the imaginary parts cancel.

For sound propagation the input sounds will be of finite length and normally based on discrete measurements. Sound waves will be of finite size, since in reality damping occurs. This has to be modeled explicitly in physical models and is sometimes disregarded. So in general we will be fine with the above mentioned Fourier transformations and their inverse transformations.

2.1.3 Wavelength and speed of sound

The wavelength of a certain frequency is dependent on the medium the wave travels through. For example a sound wave with a frequency of 440 Hz will have a wavelength of about 78 cm in air and about 337 cm in water. A faster transmission in water occurs because the water molecules are more connected to each other than air molecules and therefore propagate the wave faster. The wavelength has to stretch because the frequency stays constant while the propagation speed increases.

On the other hand the speed of sound is connected to the temperature of the medium. The higher the temperature the higher the speed of sound. (see Table 1 for propagation speeds of waves in different materials)

The (maybe unintuitive) phenomenon of higher speeds of sound at higher temperatures result from the fact that with increasing temperature not necessarily the density changes. If thinking about atoms or molecules it may become more clear that already moving particles can easier bump into other particles to propagate a wave, as well as particles close to each other can also bump easier into other particles to propgatate a disturbance.

These different propagation speeds influence the direction of propagation of waves due to gradually changing temperatures in gases and liquids; it is a special case of refraction as explained in 2.2.2. One example is the wave propagation in the atmosphere. A graduate temperature change will bend the sound waves.

2.2 Wave propagation phenomena

When a wave reaches an obstacle or a boundary, it can be *reflected*, *diffracted* by the obstacle or transmitted and *refracted* into the obstacle. Figure 6 schematizes those effects. It is important to realize that the latter two observations vary by wavelength (and therefore frequency, dependent on the medium of propagation). The specular reflection angle is independent of the wavelength.

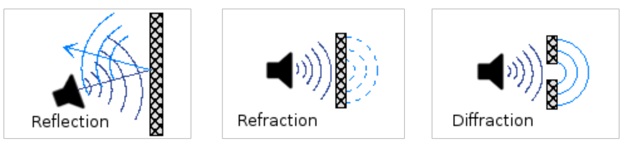

Figure 6: Reflection, refraction and diffraction

2.2.1 Reflection

Due to the law of *reflection* [Kuttruff 2007] the direction of incoming waves (incident waves) and the direction of outgoing waves (reflected waves) make the same angle with respect to the surface normal n:

$$\theta_{in} = \theta_{out} \tag{6}$$

as shown in figure 7. This is valid for light as well as for sound waves. Reflections mentioned above are called specular reflections. In a simplified model, that approximates rough surfaces by material properties, also diffuse reflections can be modeled. Diffuse reflections distribute the incoming over a larger area and are in general much more important for light than sound.

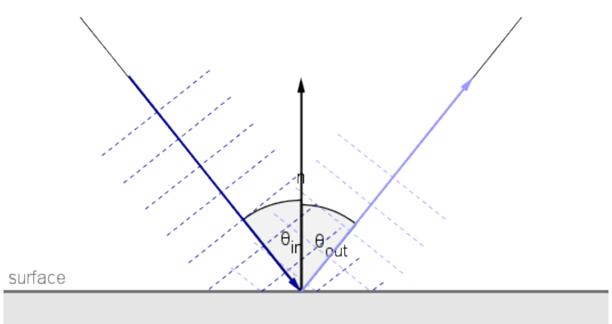

Figure 7: Reflection of a planar wave

A wave reflected at a surface will loose some of its pressure amplitude and may undergo a jump in phase. A jump in phase means that the phase angle changes abruptly. See section 2.2.4. for more information on phase jumps. Therefore a reflection factor R can be written as:

$$R = |R|e^{j\chi}. \tag{7}$$

This factor is multiplied to the incident wave, to model the change in amplitude (with $|R|$) and phase (with χ) of the reflected wave. The result of this multiplication with a regular harmonic oscillation is:

$$r(t) = |R|\hat{s}e^{j(\omega t+\varphi+\chi)}. \tag{8}$$

If $|R| = 1$ perfect reflection occurs, the amplitude of the reflected wave will be equal to the amplitude of the incident wave. This value is approached by real walls, but never reached. Also notice that the reflection factor is frequency dependent [Kuttruff 2007, page 177] and can be angle dependent [Kuttruff 2007, page 103].

In practice, the absorption coefficient α:

$$\alpha = 1 - |R|^2 = \frac{I_i - I_r}{I_i} \quad \text{or} \quad |R| = \sqrt{1-\alpha} \tag{9}$$

of a material is used. It is the amount of sound energy lost during reflection divided by the total incoming sound energy. I_i is the intensity of the incoming wave, I_r the intensity of the reflected wave. Table 2 shows some typical absorption coefficients for different materials with a random angle of incidence.

Material	Octave band center frequency (Hz)					
	125	250	500	1000	2000	4000
Hard surface (concrete, bricks)	0.02	0.02	0.03	0.04	0.05	0.05
Parquet	0.04	0.04	0.07	0.06	0.06	0.07
Carpet 5 mm	0.02	0.03	0.05	0.10	0.30	0.50
Glass window	0.35	0.25	0.18	0.12	0.07	0.04
Water	0.01	0.01	0.01	0.02	0.02	0.03
Adult person	0.25	0.35	0.42	0.46	0.50	0.50

Table 2: Absorption coefficients [Kuttruff 2007], these values are only typical examples and can not be seen as totally accurate. More values can be found on the internet.

Another quantity used in this context is the wall impedance. From the reflection coefficient the wall impedance can be calculated and vice versa. Impedance will be explained later in more detail.

2.2.2 Refraction

Refraction of waves can be described by:

$$\frac{c}{sin\theta} = \frac{c'}{sin\theta'}. \tag{10}$$

This is called 'Snell's law' in optics and is shown in figure 8. The refraction angle is dependent on the sound velocities c and c' in both media.

Alternatively Snell's law can be written as [Kuttruff 2007]:

$$\frac{sin\theta}{sin\theta'} = \frac{n'}{n}. \tag{11}$$

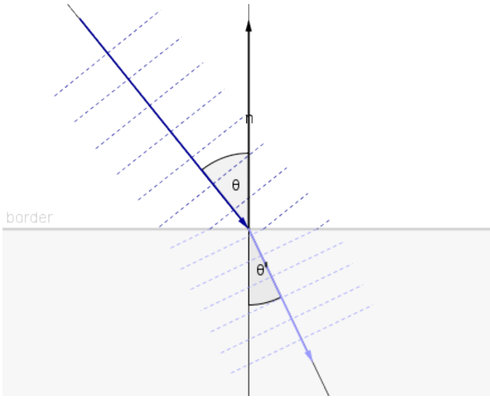

Figure 8: Refraction of a planar wave

where n are the refractive indices. Notice that the refractive index can be dependent on the wavelength of the wave. So the diffraction angle may vary with wavelength. (e.g. prisms)

Refraction does not necessarily occur at a border. Refraction can also occur on a graduate change of a material's density or a material's temperature.

In reality reflection and refraction often occur at the same time, dependent on the angle towards the boundary. At a certain angle, *total reflection* occurs. That means that the incident wave is totally reflected and not refracted.

To understand the mathematical explanation of this phenomenon it is easier to think about a beam of light at first. Light is propagating faster in air and slower in water, contrary to sound. A faster movement results in larger distances between the wavefronts. When looking at figure 8, imagine a beam of light coming from the lower gray area moving up into the white upper area. If the angle towards the surface boundary increases the intersection points of the wavefronts with the surface will move closer towards each other. At a certain angle, there is no way for an outgoing wave to be created in the upper white region, since the wavefronts are too close to each other. For sound total reflection only occurs when moving from air to water because of the different propagation speeds of sound.

Formally the following has to hold for total reflection to occur:

If $c < c'$ and $sin\theta > \frac{c}{c'}$.

Then $sin\theta'$ would have to be greater than 1. This is impossible and therefore total reflection occurs.

This principle is made use of in fiber optics.

2.2.3 Diffraction

Diffraction is a bending of waves along corners or into openings. The amount of diffraction increases with wavelength, meaning that high frequencies will be easier blocked and less diffracted by obstacles than low frequencies. Noticeable diffraction only occurs if the wavelength is bigger than the obstacle or opening. Usually visible light waves will be less affected by diffraction than sound waves, since their wavelength is a lot smaller (0.4 to 0.8 μm). Sound propagation on the other hand is heavily influenced by diffraction. Even human hearing depends on diffraction as explained in the section about psychophysics.

Diffraction is sometimes called scattering for very small objects compared to the wavelength.

Diffraction is difficult to account for. The basic approach is the Huygens-Fresnel principle. It sees each point of an advancing wavefront as the seed point for a new spherical wave (or wavelet). This is shown in figure 9. The lines of the figure are corresponding to wave fronts. At the opening the yellow dots are the centers of the wavelets.

The larger the space between wave fronts, the higher the wavelength. As already mentioned, for a large wavelength diffraction is more visible. In the center of the opening, most of the points of adjacent wavelets cancel each other out. Only points radiating directly in direction of the original wave remain. The wavelets at the borders are missing some adjacent wavelets. At these points they can not be canceled out and radiate also sidewards. These wavelets at the borders of an opening also affect the wavefront traveling directly straight through the opening. The do not only 'brighten the shadows' behind the wall, but also have an impact on the total sound field that propagates through the opening [Kuttruff 2007].

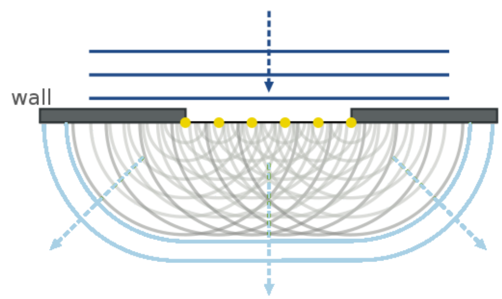

Figure 9: Diffraction due to Huygens principle: Top down view, the yellow points are sources of wavelets. The incoming wave is dark blue, the exiting wave is light blue. [http://en.wikipedia.org/]

Different cases are to be considered for diffraction through large and small openings (also called apertures) compared with the wavelength. Also solutions to diffraction around corners and diffraction around small objects are to be formulated.

The Helmholtz-Kirchoff integral theorem considers a function based on the Huygens-Fresnel principle. It can be used for diffraction through large openings and around corners. Diffraction on small objects depend on the object's shape. For a sphere, the diffracted wave is in shape of a spherical wave combined with a dipole wave. A dipole wave is a directional sound source that can be modeled by a combination of two point sources. The dipole wave emits waves in two opposite direction whereas the waves radiating orthogonal to the direction of propagation cancel each other out. All those cases are explained in higher detail and with further references in [Kuttruff 2007].

There are also geometric solutions to diffraction. The geometric theorem of diffraction for optics has also been applied to acoustics. It can be found in [Keller 1962].

2.2.4 Reflections and refractions at boundaries

On a closer look reflections and refractions of waves occur at boundaries between mediums. There are two kinds of reflections at boundaries.

One kind of reflection happens if the wave encounters a *soft boundary*. If you imagine a wave as a transverse oscillation of a string and the end of the string is not fixed in space it can oscillate as well. Reflection of waves on soft boundaries are simply reversed back to where they came from. No phase jump occurs. Reflections against soft boundaries may occur for example when a sound wave under water hits the water surface.

The second way of reflection is when the wave encounters a *hard boundary*. Waves will be inverted when they are reflected. You can imagine a string tied to a fixed point. If you wave it you will encounter an inverse reflection of the incoming wave next to the fixed point. If the wave is a sound wave (and therefore waving in positive and negative direction) a phase jump along half a wavelength will occur. That means an incoming wave's pressure changes sign while hitting the boundary. Reflections at hard boundaries occur when sound hits a wall.

Since most reflections are non total reflections, also transmission into the material will occur. Transmission is never inverted. The amount of a wave that is reflected and refracted depends on the materials, the sound frequencies and the angle. Certain materials allow more transmission for certain frequencies than others. Also reflection depends on the surface of an obstacle. A plain surface reflects more than a rough surface.

2.2.5 Interference

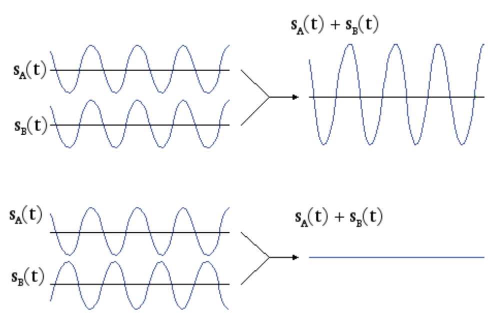

Figure 10: Superposition: constructive interference top, destructive interference bottom

When two or more different waves occur at the same location, their amplitude is simply added at this location. This principle of addition of wave amplitudes is called *superposition* of waves. If two or more waves appear and therefore interfere at a shared position, two phenomena can be monitored. *Constructive interference* happens if the amplitude of two waves added result in a larger amplitude at a point. There is also *destructive interference*, when waves may cancel each other out. This occurs when the positive amplitude of another wave interferes with the negative amplitude of a wave. The absolute value of the resulting amplitude can be smaller than the absolute values of the amplitudes of the individual waves. Total destructive interference occurs if the amplitude at the location equals zero. Therefore the amplitudes at a position and time must be in relation: $s_A(t_I) = -s_B(t_I)$. The resulting amplitude will be $s_R = s_A(t_I) + s_B(t_I) = 0$. Both effects are demonstrated in Figure 10.

An interesting question that may arise about interference, is what happens to the energy during destructive interference. An easy answer would be that the destructive interference occurring at a certain point in space and time will create constructive interference somewhere else. But a more accurate solution is out of the scope of this report.

2.2.6 Amplitude, intensity and impedance

The last aspect of waves we did not include yet is the *amplitude* or *intensity*. For sound there is on the one hand air absorption that reduces a wave over time. Air absorption is frequency dependent. In general higher frequencies are absorbed earlier. Lower frequencies tend to travel further. On the other hand a 3 dimensional wave spreads out in all directions. This means that intensity is decreased in a quadratic way. The inverse square law states that the amplitude of a wave is inverse proportional to the squared distance to the source.

Due to the inverse square law, the intensity of a point source which radiates in all directions is:

$$I_r = \frac{I_P}{4\pi r^2} \tag{12}$$

where I_r is the intensity at distance r to the point source. The point source intensity is I_p [Begault 1994].

Another term is the impedance. The characteristic impedance is defined as

$$Z_0 = \rho_0 c = \frac{p}{v_x}. \tag{13}$$

The pressure p is in relation to the particle velocity v_x. c is the speed of sound. ρ_0 is the density of the medium at rest. Like the reflection coefficient, the wall impedance has a real and an imaginary part. The characteristic impedance is a material property. If a high pressure is required to excite a relatively low velocity the characteristic impedance of the material will be high. Otherwise if a relatively low pressure is required to excite a high velocity, the impedance is low [Kuttruff 2007].

Another impedance concept is the wall impedance. It can be written in the same form as the characteristic impedance but v_x is normal component (to the wall) of the particle velocity. The wall impedance can be used to model boundaries. More on boundary conditions is found in section 5.1.5.

2.3 Other acoustic effects

2.3.1 Doppler effect

For moving sound sources the *Doppler effect* should be considered. This phenomenon is noticeable as a change in frequency when listening to a moving sound source. For example the pitch of a passing ambulance car would drop as the vehicle passes. While a sound source moves the effective wavelength in direction of the movement is shortened, since the propagation speed of sound in air stays constant. The effective wavelength in the opposite direction of the movement is increased. Figure 11 schematises the Doppler effect.

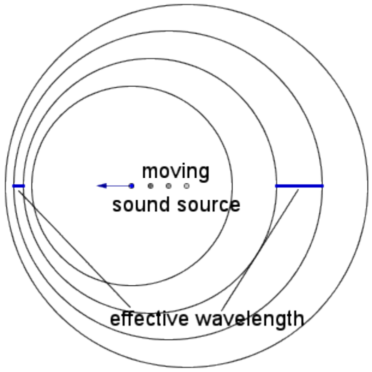

Figure 11: Doppler effect with effective wavelengths.

2.3.2 Normal modes

When waves propagate in an enclosed space with finite dimensions certain frequencies will 'fit' into the space. Imagine the one dimensional example of a tube with two rigid (perfectly reflective) boundaries. The wave propagation in this tube will be limited to certain frequencies, since the maximum frequency is limited by the length of the tube. Because of the boundary conditions for the start and end of the tube, which are perfectly reflective (and therefore the pressure at the endings is 0), only an integral number N of half-wavelengths can fit into the tube. These frequencies are called 'Eigenfrequencies' and can be written as:

$$f_n = n\frac{c}{2L}. \tag{14}$$

where L is the length of the tube and $n = 0, 1, 2, ...$ and c the speed of sound. The waves that oscillate with the 'Eigenfrequencies' are called normal modes of the tube. Figure 12 shows some exemplary normal modes for tubes with closed and open ends. Closed ends were already described with perfectly reflective ends, that means that their impedance is 1. The open end's impedance is 0.

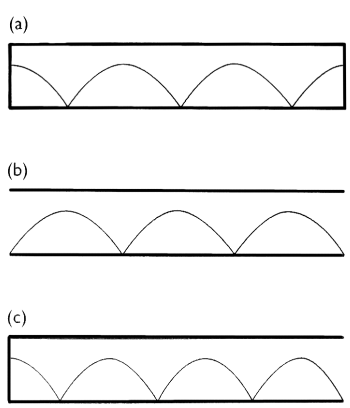

Figure 12: Normal modes for a tube with a) two closed ends b) two open ends and c) one open and one closed end [Kuttruff 2007]

A modal analysis can be performed on an arbitrary surface to find its normal modes. When an oscillating force is applied to a structure with varying frequency, certain frequencies will propagate much stronger then others through the structure. These are the normal modes.

Normal modes are important in the design of many different objects to prevent oscillations that may damage the structure (e.g. bridges) or to prevent noise. In music the normal modes of a string are the harmonics or overtones of the string.

3 Human Perception

The modeling of sound waves using exact physical models can be very resource expensive. A way to reduce its complexity is to consider only effects that affect the human perception the most. One example of the successful application of knowledge about human perception is the mp3 compression. The display of sound happens normally through a limited amount of speakers while a real sound usually occurs in 3d space. Therefore it is important to draw a connection from physical signals to human perception. These considerations may on the one hand help to improve sound simulations and help in further experiments of human sound perception since reproduction of experiments can be managed more easily. On the other hand, it can improve the field of sound pattern recognition and sound recording. (For example source position finders or cocktail party processors [Blauert 1997, page 393].)

This part starts off with looking at human perception in general, especially in relation to virtual environments, to give a little motivation to increase the realism of sound rendering in virtual environments. After that a sound specific section will give an introduction to the field of psychophysics for sound.

3.1 Perception in the real world vs. virtual environments

Imagine a walk through a city scenario, for example walking home from university. What will your eyes do and what will you hear. First of all it is important to realize that both senses are limited. A human being creates an overall impression with help of all available information and additionally through the previously learned.

The eyes will be mainly used to check the path for obstacles and to watch uncertain possibly dangerous or interesting spots. They will act as a local scanner picking points of interest for closer observation. But they only cover a limited part of the world around us. Without the aid of our auditory system, anytime, danger could approach us from behind or from the sides. (For example a car when crossing a crossroad.) Some opinions go even further and state: The ears are steering the eyes [Röber and Masuch 2004a].

This may be to some point correct, in a way that auditory perception sometimes covers a different part of space, dependent on the available information and the physiology of eyes and ears. Firstly our visual system has a limited field of view, only covering roughly 180 degrees and still the radius of clear foveal vision is a lot smaller. Caused by the physical properties of light we can not see around corners or through thin light-absorbing materials. Also in the absence of light or during sleeping sound may play a more important role than light.

As seen in the physical section, sound waves, unlike usual light waves, can be diffracted around obstacles and reflection of sound waves usually hold a higher specular portion than diffuse portion, in comparison to light. This is due to the larger wavelength of sound. Larger wavelength reflections are not that much influenced by a little surface roughness, leading to cases when our auditory system can hear further than our eyes can see. (e.g. around corners) Additionally, our auditory system is able to perceive a 360 degree radius.

But unfortunately not all visible objects emit sound [Röber and Masuch 2004b]. Nevertheless also non-emitting objects and environments do produce an audible impression due to reverberation. A human person may emit sound and the reverberation information of this sound can contribute to the listeners imagination of his surrounding. e.g. A person in a dark unknown room may make some noise, to find out how large the room approximately is.

So our visual system could be seen as the more reliable sense in a way that most existing objects, connected in a line to a perceiver, can be seen. Also it is more accurate for localization. The visual system is able to distinguish a resolution of less than an arch minute. While the auditory system, at its best, can sense a difference in location of about one degree. That is about two orders of magnitude less than the visual system [Blauert 1997, page 38].

Reaction times on the other hand are slightly faster for sound events.

In real scenarios all senses and knowledge are combined to create an overall impression. So sound may enhance visual and visual can enhance sound.

Effects caused by the combination of visual and auditory cues are: The *ventriloquism effect*, which states, that a sound signal that may correspond to a visual impression, is imaginary moved to match the visual, even if occurs at a physically different position. Another aspect of the connection of visual and auditory input is multi-modal task facilitation. It can improve the performance of actions using the combined input of multiple modalities.

3.1.1 In virtual environments

Typical virtual environments for most of us are computer games. In common first person games the field of view only covers 75 degrees [Valve 2010]. As mentioned earlier, our visual system covers about 180 degrees. The field of view can be broken down to foveal view (center of view) and peripheral view. The foveal view only covers about 2 degrees of the visual field.

So what does this mean for typical virtual environments on typical computer screens? The field of view is even more limited than in reality. So on perceiving roughly 285 degrees of a virtual scene an actor will rely on audio signals. It should be mentioned that there are tries to increase the field of view by adding multiple monitors (by ATI), head mounted displays and multiple projection areas (e.g. CAVE). Still the use of a single monitor is the most common way interacting with todays virtual environments. Therefore audio is a very important part to create a 360-degree perception of a virtual environment.

It should be mentioned that perception in virtual environments is still an active research field. Some studies use the term of immersion to describe a degree of being pulled into a virtual environment. It has been proved that a combination of auditory and visual cues enhance the sense of immersion [Tsingos et al. 2004a]. One of the major questions of perceptional experiments is: Are observations in a virtual environment also valid for reality and vice versa?

Other improvements to current audio rendering techniques may include additional user input, such as head movements or speech, since audio is a very interactive way of perceiving the world. There are still many gaps to close that lie outside the task of 'simply' simulating audio.

3.2 Psychophysics of sound localization

Psychophysics measure the human sensory system.

Psychophysics of human sound localization regard the fundamental question: 'What conditions must be fulfilled inside and outside an organism for a particular object to appear in the sensory world of that organism? [Blauert 1997]'

Therefore two terms are introduced. Sound source or sound event, to describe a physical phenomenon and auditory event, to describe what is perceived auditorily. An auditory event may be caused by a sound event, but this is not necessarily required. (e.g. tinnitus) One of the first concerns is how those two events are connected to each other.

In psychophysics, experiments are particularly important. One simple example of an experiment would be to point in the direction of an auditory event caused by a sound event. Experiments where the experimenter is the subject and experiments in which other persons are the subjects are possible. In the latter case an auditory event can only be accessed indirectly by the subjects' description. One important aspect of experiments in general is their *objectivity* or general validity of measured results. Objectivity is achieved when measured results are always the same. This can be the case if multiple measurements with the same subject, or a single measurement for multiple subjects are taken. Measurements of objectivity should also be included in psychophysical experiments.

In numerous auditory experiments the localization and localization blur of auditory events in certain conditions were examined. Localization is the rule by which a location of an auditory event is related to attributes of a sound event. Localization blur on the other hand is the smallest change in an attribute of a sound event to be changed in order to perceive a change in position of an auditory event. Localization blur exists, because auditory space is perceived at a lower resolution than physical measurements can achieve. Localization blur depends on the direction of a sound source and on the frequencies within a sound signal.

Figure 13 shows the horizontal plane and the median plane. The center of this coordinate system lies halfway in between the entrance of both ear channels. Dependent on what plane a sound source lies, two different cases of localization can be observed.

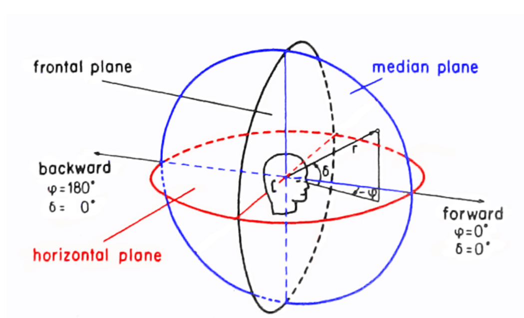

Figure 13: Head related coordinate system showing the horizontal and the median plane [Blauert 1997].

3.2.1 Localization in the horizontal plane

For localization in the horizontal plane two effects are to be considered. *Interaural time difference* (ITD) and *Interaural level difference* (ILD).

Interaural time difference occurs because the path lengths of a sound signal approaching both ears from a position not laying on the median plane differ. The time difference at the ear can be a phase difference for low frequencies, or an amplitude difference for higher frequencies with amplitude fluctuations. In general this phenomenon is more significant for frequencies below about 1000 Hz. For a harmonic wave with constant amplitude, it would be impossible to find the correct time difference between higher frequencies, since it would be impossible to distinguish between the leading and the following wavefront. This is because multiple wavefronts lie in between the leading and the following wavefront, as the waves fluctuate to fast. The auditory system deals with this problem with the ITD envelope cue [Begault 1994].

Interaural level difference on the other hand is important for wavelengths smaller than the diameter of the head. Frequencies with greater than about 1500 Hz are shadowed by the head. This shadow effect is increased for higher level frequencies. ILD is the important cue for stereo recordings. Stereo recordings only consider ILD since ILD is relevant for a range down to at least 200 Hz. Most recordings have frequencies laying above that frequency.

The physical cause for ITD and ILD is diffraction. Low frequencies are diffracted around the head because of the diameter of the head which is small compared to the wavelength of low frequencies. Therefore no audible ILD occurs, because the diffracted sound reaches both ears with about the same sound intensity. High frequencies are diffracted with a lesser amount around the head. In this case the head creates a 'sound shadow' and causes ILD.

Figure 14 shows the result of tests on localization blur (corresponding to a resolution of auditory events) along the horizontal plane. That data is based on experiments by Preibisch-Effenberger and by Haustein and Schirmer, who let subjects position speakers in alignment with a fixed speaker.

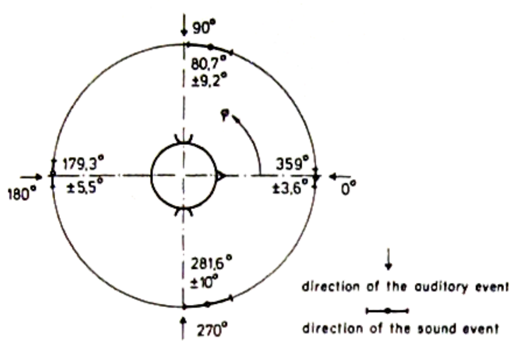

Figure 14: Localization blur in the horizontal plane [Blauert 1997].

Another problem arises from the normally static positioning of the head in experiments. Sometimes mirroring of an auditory event occurs. This means that a sound event in forwards direction causes an auditory event that is mirrored along the frontal plane and therefore occurs behind the listener. Head movements are very important to place such an auditory event in the right direction. It is pointed out that these can be tracked with a head tracker. It is worth mentioning that in today's first person computer games head movements are possible with moving the mouse cursor.

3.2.2 Localization in the median plane

ITD and ILD do not occur along the median plane, since distances to the sound source are constant in that case. Nevertheless localization is possible along the median plane.

An experiment regards the localization blur in the median plane for the speech of a familiar person. The results can be seen in Figure 15.

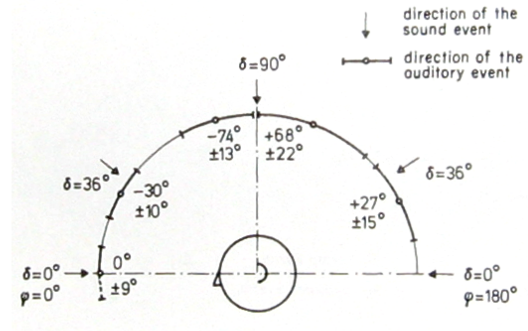

Figure 15: Localization blur in the median plane [Blauert 1997].

A closer look on localization in the median plane revealed that the location of an event is based on the frequency of a signal. This is caused mostly by the shape of the ears. It has been shown for a narrow band signal with a certain frequency that the location of the auditory event only depends on the frequency of the signal. Figure 16 shows the frequencies and their locations.

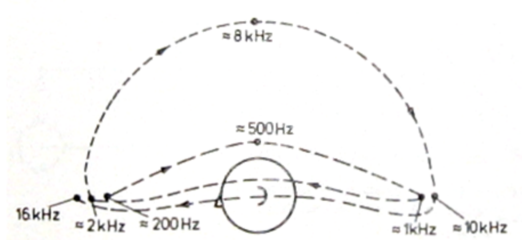

Figure 16: Location of an auditory event dependent on the center frequency of a narrow band signal in the median plane [Blauert 1997].

3.2.3 Distance

The property is the distance of an auditory event. For distance hearing the familiarity to a certain sound is the most important cue. Best localization was found for familiar sounds. For unfamiliar sound events the auditory event depends mostly on the perceived loudness. Loudness is the interpreted intensity of a sound event. Therefore a familiar (or know to be close) relatively silent sound can appear closer than a familiar louder sound heard at the same time. E.g. In a room the silent clicking sound of a clock may appear closer than the loud sound of a car passing outside. If sounds are unknown (e.g. unfamiliar animal sounds), the perceived distance depends mostly on loudness in a way that louder sounds appear closer.

Other possible distance cues can be frequency, interaural time difference and reverberation. Frequency dependent cues are weak compared to loudness, but they may be added for a more precise simulation. Interaural time difference is insignificant for distance localization [Begault 1994, page 96]. Reverberation can be also an important distance cue.

3.2.4 Law of the first wavefront

The law of the first wavefront, also called the precedence effect, states that the reflection is not taken into account for localization of an auditory event. In a physical expression, only the direct sound, or first arriving wavefront is important for sound localization. In a setup of two loudspeakers radiating the same sound, many different thresholds were found with experiments. If the time difference between the signals coming from both loudspeakers is about 0.6 to 1 ms, the law of the first wavefront starts to apply. That means, the sound seems to come directly from the loudspeaker activated first. Many other thresholds considering the law of the first wavefront have been formulated and are found in [Blauert 1997, page 222ff].

3.2.5 Reverberation

Reverberation is an effect caused by multiple reflections, refractions and diffractions of a sound wave that normally follow the direct wave from the source. Reverberation can be displayed with use of the diagram of an impulse response. An impulse response shows the pressure of the medium over time caused by the propagation of an impulse in a certain room at a specific listener position. An impulse is a short signal covering all frequencies. By the shape of the impulse response also frequency dependent information is decoded due to the fluctuation of the pressure values over time. With Fourier transformation a frequency response can derived from the impulse response.

Normally reverberation is considered to be a room specific parameter. A certain room has certain reverberation parameters and therefore a certain impulse response. This can be measured by sending a short signal into the room and measuring the pressure amplitude arriving at a listener in the same room over time. It should be mentioned that this is only a simplification and the impulse response depends on source and listener positions within a room.

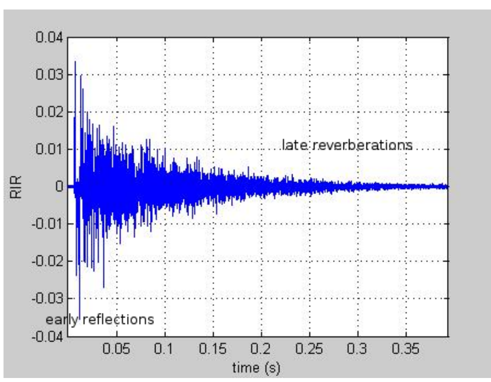

Figure 17: Examplary impulse response showing early reflections to the left and late reverberations to the right. [www.mathworks.com]

Figure 17 shows early and late reflections dependent on the time of arrival to the listener. Early reflections occur within the first 80 milliseconds of the measurement. They are followed by late reverberation. Late reverberation is spatially more diffuse. In reverberation models normally only the early reflections are modeled directly, while late reverberations are often approximated by a Gaussian distribution with decaying envelope. The reverberation decay in a larger room is much slower than in a small room.

3.2.6 Head related transfer function

In psychophysics it has been revealed that the pinna (outer ear) codes spatial attributes to a sound. The physiology of the ears but also the rest of the head add a linear time-invariant, frequency dependent amplitude and time-delay to incoming sound. This effect can be measured or simulated and expressed as a head related transfer function (HRTF).

HRTFs can be measured for each individual, but since this is not very practicable normally non-individualized HRTFs are used. These general HRTFs can be derived from averaging over measured ones. The use of HRTF for localization has been verified by experiments. Spectral HRTF created cues are of no or low relevance for locating horizontal sounds. For location in the horizontal plane ITD and ILD are more relevant [Begault 1994, page 62]. The main role of HRTF cues lies in distinguishing between forward and backwards laying sound and the elevation cue (up and down).

HRTFs are the currently known most accurate technology for using headphones (also called binaural). If speakers are used (called transaural) no HRTF is needed since the signal will reach the ears naturally. The advantage of using headphones for sound localization, is that surrounding room parameters play a very minor or no role at all. When playing sound over speakers other phenomena have to be considered, like cross-talk cancellation. There are also surround systems, which model sound by a greater number of speakers. As already mentioned the actual room acoustics of the room in which the speakers are used can heavily influence the perceived auditory impression. But the technology behind those systems is beyond the scope of this report.

3.2.7 Auralization

For simulating room acoustics using HRTF technology in virtual environments, the following simplified model is used [Blauert 1997, page 379]. First an impulse response of a room has to be generated. How this can be achieved will be explained in further detail in section 5 "'Suggested Methods of Advancement"'. It can be used to apply room specific properties to an audio recording. This impulse response is in general calculated for each sound source and listener pair in the scene. For two ears it may be required to calculate two impulse responses. The last part of making an arbitrary sound signal audible within this environment is called auralization.

To manage auralization digital signal processing (DSP) filters have to be applied to a preferable anechoic recorded sound signal. Anechoic means a dry signal with no reverberation. Three steps of signal processing have to be taken. A temporal delay dependent on the distance of the sound source. A spectral filter that modifies the sound signal dependent on the absorption of frequencies during the waves travel. In the end a spatial filter convolves the signal with an external ear impulse, when using HRTF.

As shown above, spatial audio is an important part of human perception in the real world, as well as in virtual environments. The succeeding part of this report will give a summary on current and influential past ways of rendering audio in virtual 3d worlds. After giving this overview I will represent some suggested methods of advancement to the currently common ways of rendering audio.

4 Audio in Virtual Environments

There are many possible applications for spatial audio. Reaching from entertainment, to education and exploration of data. Since this report is about spatial audio I will not discuss all different kinds of virtual reality applications, but rather take a few examples and explain how they deal with spatial audio.

4.1 Audio in 3d games

Todays probably most frequently used applications for real-time spatial audio are computer games. Usually bringing sounds into a game is a task of taking audio files and use one of the many available audio libraries to place these sound sources in a 3 dimensional coordinate system.

Most audio libraries use a 3 dimensional coordinate system to place listeners and sound sources. The listener is obviously the location from where the audio scene will be perceived. The sound sources can be turned on or off, change volume, increase or decrease play back speed, have a certain spatial velocity and more. Depending on the library you want to use, both, listener and sound sources can change their positions over time. All you have to do is to deliver coordinates, orientation (specially for the listener) and sometimes velocities to the library. So this seems really straight forward.

One major disadvantage that can already be derived from this kind of input/output system is that there is no geometric input to the sound library. So sound propagation only depends on positions, distances and orientations. A lot is already possible with this simple model. But obviously this system also produces many errors. Imagine a level with rooms very close to each other but no connecting doorways. The sound would be audible through the wall as if there was none. This is often unaccounted in older or lower budget games (e.g. Counterstrike). Another problem that is not solved in many games are dynamical changes to audio environments. The most basic would be closing a door, other possible applications could be dynamical changing audio depending on, for example, the furniture of a room.

More advanced sound libraries like Fmod, support the input of geometry for simple occlusions. Due to hardware restrictions they are described as very simple ray-tracing methods with additional reverberation parameters. Eax supports parameters for occlusion and obstruction, the actual calculations must be done separatly from the eax library.

4.2 Enhancing spatial sound

Let us take a closer look on how certain games enhance audio with geometric information. One very basic thing is to add *material types to textures*. For example to make the sound of footsteps dependent on a material type tied to a certain (floor) texture. Also interactions (shooting at, throwing things at) walls can be enhanced be additional texture information. This is a standard for most shooter games out there. But some games use that more than others.

Another way to add additional depth, is the use of (many) *ambient sounds*. Those can be either placed in actual 3d space or be a non spatial soundtrack. The advantage of placing each individual sound source is a more realistic game world, the disadvantage is that it can be a lot more work and requires more hardware resources. Notice that even a not explicitly spatial soundtrack can still be pseudo spatialized with simply using a 3 dimensional recording of a real environment. Ambient sounds which include far animal or traffic sounds with a typical far field frequency filter applied can greatly increase the perceived realism of a virtual environment. Also multiple layers of ambient sounds can be considered.

A newer approach is to calculate *room dependent reverberation parameters*. Those can be applied to the sound sources during runtime. On entering or leaving rooms sounds are normally gradually changed.

Objects of relevance in an interactive environment can be supported with sounds to focus the player's interest.

In games often the interactivity is of special importance. Sounds can greatly improve the immersion of a user in an interactive process in virtual environments. So dynamic sounds that react on player actions are very important. One example in Crysis would be the sound of birds flapping away and silence after a gunshot of the player.

Figure 18: Unreal 3, Half Life 2 and Call of Duty

In the following the parts of the audio system of three games shown in Figure 18 will be explained shortly.

4.3 Spatial audio in the Source Engine

One more advanced spatial audio engine is included in Valve's Source Engine used for a variation of games. On compilation of a level the compiler calculates room dependent reverberations of a level. That will be applied during runtime. One special addition to the engine's audio is the use of Soundscapes. Soundscapes in this context are looped and randomly played ambient sounds and can be stored in specific text files. When moving from one Soundscape to another, the previous is fading out while the new is fading in.

4.4 Spatial audio in the Unreal 3 Engine

Unlike in the Source Engine the default unreal engine does not calculate reverberation parameters for standard levels. Special reverberation brushes have to be set, to modify sound reverberation settings for an area. Those reverberation settings will be applied to all game sounds. To allow additional flexibility these brush based reverberation areas can also be applied to ambient sounds. (named ambient zones)

4.5 Spatial audio in Call of Duty: World at War

According to an article on the Call of Duty homepage [Treyarch 2009] they calculate occlusions for sound sources also depending on the degree of occlusion. So a sound behind a wall would sound different than a sound behind a low wall cover. This is done by adding a low pass filter to ”‘muffle”’ the sound. The second enhancement explained is a system called flux, that generates moving sound sources for shots and shock waves.

The problem of verifying currently used audio technology is heavily dependent on what companies allow to become public. Therefore only a few examples could have been picked out. Along spatial audio technology, also background music or interface sound feedback are important aspects of game audio. Music can be used like in movies, but it is important to notice that games tend to have a longer play time than movies, so music can get repetitive and also dynamic changes are possible because of the non linearity of games. This opens up another wide area of normally non spatial audio aspects of games. The problem of repetitive sound effects should also always be considered. For example if there is only one sound for each footstep it may get annoying to hear it all over again, or at least not very realistic [Piringer 2010].

5 State-of-the-art in research

This last chapter will concentrate on research papers of the field. With the background knowledge of the previous sections it should be possible for developers, as well as other interested people in the field, to use some of that knowledge to improve new applications.

The ultimate goal of this section would be to model sound propagation in a virtual environment with as little as possible resource usage and the most realistic output.

The problem of sound propagation is connected to the problem of global illumination. Like in global illumination wave (or light) reflections and refractions are to be considered. When thinking back to the physical section, some differences to solving the problem of global illumination are noticeable [Funkhouser et al. 2002].

- *Wavelength* of light is far smaller than the wavelength of sound and sound waves cover a much larger spectrum of different wavelengths. This means, that for the modeling of sound, less detail of geometry is required. On the other hand as mentioned in the physics section, the wavelength of a sound wave has to be known to precisely model its propagation. For example diffraction (a not so noticeable effect for normal lighting) depends on the size of a possible occluder and must be calculated.
- *Speed* needs to be considered since it is perceptable. (e.g. reverberation)
- *Coherence* of waves needs to be taken into account. As mentioned in the physics section coherence of waves leads to effects called destructive or constructive interference.
- The *number of reflections* has to be higher than for global illumination, since the time difference will cause even low amplitudes (volumes) to be audible. Also specular reflections are stronger for sound, because of the wavelength which makes sound reflection less sensitive to surface roughness.

Even though these differences exist, the possible solutions are somewhat similar to solving the problem of global illumination. Methods of solving the problem will be divided into two sections: numerical methods and geometrical methods.

Detailed summaries about the topic can be found in [Manocha et al. 2009] and [Funkhouser et al. 2002].

5.1 Numerical methods

Numerical methods are very computation-expensive and are therefore not suggested for real-time audio rendering. Recent applications have shown that they can be used for real-time audio rendering for static geometry. Nevertheless this section will give a short overview.

The basic approach of numerical methods is to formulate a wave equation with certain boundary and initial conditions. The area inside the boundaries (specified by the boundary conditions) describes a pressure field. Sound sources (which are pressure disturbances) can be placed inside this pressure field either by the initial conditions, which normally describe two time steps of the pressure field, or by functions for sound sources which are included in the wave equation. So the wave equation is either forced by zero if there are no radiating sounds in the area or forced by a function of a sound source. Without damping any initial force will never stop to be propagated through the pressure field.

With the solution to such a wave equation any point in time and space can be directly calculated. Nevertheless, some numerical methods do not express such an explicit solution and only provide a mechanism to calculate succeeding time steps.

The following will give an overview to understand the analytical ground truth of the practical methods described later. Moreover, this should raise the readers awareness of the assumptions and simplifications being made in the computational optimizations, which are particularly mandatory for real time applications.

5.1.1 Wave equations

A one dimensional wave is a wave propagating in one direction, for example a planar wave traveling through a tube.

The one dimensional wave equation can be written as:

$$\frac{\partial^2 p}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2}. \tag{15}$$

p(x,t) is the pressure function, x is the propagation axis, t the time axis and c a constant value that turns out to be to the sound velocity or speed of sound in acoustics. c can be a different constant value for other types of waves.

The wave equation is a linear homogeneous partial differential equation. Partial differential equation means it contains a function $p(x,t)$ in multiple variables and some of their partial derivatives with respect to different variables. In specific, it contains the second order partial derivative with respect to x on the left hand side and with respect to t on the right hand side. It is also linear, it does not contain orders higher then one, of the unknown function p or its derivatives. Homogeneous will be explained when introducing the inhomogeneous wave equation at the end of this section.

This basic one dimensional wave equation can be derived from two fundamental equations. The first equation:

$$\frac{\partial p}{\partial x} = -\rho_0 \frac{\partial v_x}{\partial t} \tag{16}$$

is a simplified equation of the relation of the difference force, acting on both sides of an infinitesimally small volume in a tube, to the total velocity of that volume. ρ_0 is the density of the medium at rest and v_x the particle velocity in x direction. The considered volume element moves along the x axis in this first equation. The second equation:

$$-\rho_0 \frac{\partial v_x}{\partial x} = \frac{1}{c^2}\frac{\partial p}{\partial t} \tag{17}$$

is specified for a fixed volume element. Once again it is a simplified version of an original equation that relates the net influx of mass into the volume element with its containing density.

These two equations can be partially differentiated with respect to x for equation (16) and with respect to t for equation (17) to form the originally mentioned wave equation.

The one dimensional wave equation is only valid for gases and liquids. For solids it is more complicated because the shape of a solid does not deform itself and therefore so called shear stresses occur. These are the forces acting on all sides of a solid body. The detailed derivation of the wave equation and the wave equation for solids can be found in [Kuttruff 2007]. If the pressure is replaced by a displacement value, also a wave traveling through a string can be computed with that formula.

5.1.2 Solution of the planar wave equation

A function $p = f(x,t)$ that combines the variables x and t in a way that $x - ct$, with existing second derivatives, is a solution to the one dimensional wave equation. This can be shown by calculating the second derivatives for a function p with use of $u = x - ct$. After applying the chain rule and substituting $x - ct$ for u, the following holds true:

$$\frac{\partial^2 p}{\partial x^2} = \frac{\partial^2 p}{\partial u^2} \cdot \left(\frac{\partial u}{\partial x}\right)^2 + \frac{\partial p}{\partial u} \cdot \frac{\partial^2 u}{\partial x^2} = \frac{\partial^2 p}{\partial u^2} \cdot 1^2 + \frac{\partial p}{\partial u} \cdot 0 = \frac{\partial^2 p}{\partial u^2}$$

and

$$\frac{\partial^2 p}{\partial t^2} = \frac{\partial^2 p}{\partial u^2} \cdot \left(\frac{\partial u}{\partial t}\right)^2 + \frac{\partial p}{\partial u} \cdot \frac{\partial^2 u}{\partial x^2} = \frac{\partial^2 p}{\partial u^2} \cdot c^2 + \frac{\partial p}{\partial u} \cdot 0 = c^2 \frac{\partial^2 p}{\partial u^2}.$$

Plugging these back in the one dimensional wave equation proves that $p = f(x - ct)$ is a solution. Another solution is $p = g(x + ct)$. Therefore the solution of the one dimensional wave equation is:

$$p(x,t) = f(x - ct) + g(x + ct). \tag{18}$$

This equation moves the shape of a function at $t = 0$ with varying t either to the left or to the right along the x axis with speed c, in our case the sound velocity or speed of sound. This process is shown in figure 19.

5.1.3 2 and 3 dimensional wave equation

Now the one dimensional wave equation can be extended to support the two and three dimensional case that can be used for calculating 2 dimensional wave or 3 dimensional sound propagation. This is achieved by also taking the particle velocities v_y and v_z into account. After extending our formulas in (16) and (17) and setting them equal once again, the 3 dimensional wave equation reads:

$$\nabla p = \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2} \tag{19}$$

where

$$\nabla p \equiv \frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} \quad \text{or} \quad \nabla p \equiv \frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2} \tag{20}$$

is called the Laplace operator which is sometimes also written as Δ or ∇^2.

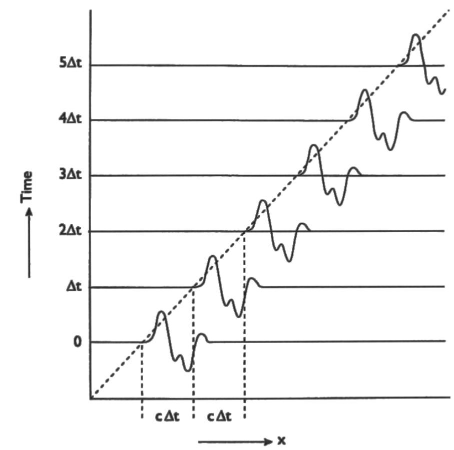

Figure 19: Pressure disturbance traveling in time.

5.1.4 Sound sources

The wave equation (19) can also be an inhomogeneous equation in the form of [Raghuvanshi et al. 2008]:

$$\frac{\partial^2 p}{\partial t^2} - c^2 \nabla p = F(x,y,z,t). \tag{21}$$

The difference to the previous version of the wave equation is that the right hand side term, also called forcing term is not zero. If the right hand side term is zero, the wave equation can be called homogeneous. The function $F(x,y,z,t)$ is a known function. It is non zero if there is a sound source in the considered area. So it describes a sound source or sound sources radiating sound. In contrary to putting pressure inside the sound field with initial conditions, the forcing term can be used to model sound sources radiating sound over time.

The pressure function of point sources can be derived similar to the plane wave. But in this case a spherical polar coordinate system is used. In this coordinate system r is the distance of a point to the origin. At each r the pressure distribution is constant. The actual formula and its derivation will not be represented here and can be also found in [Kuttruff 2007].

The equation of the pressure function for a point source is:

$$p(r,t) = \frac{j\omega\rho_0\hat{Q}}{4\pi r} e^{j\omega(t - \frac{r}{c})}. \tag{22}$$

Where $Q(t)$ is the change of the volume velocity, the volume of the expansion or contraction occurring per second. ρ_0 is the density of the medium at rest. In acoustics different directional sound sources can be modeled with a combination of point sources.

Point sources are particularly important in use of the Huygens principle. Point sources can also be used to model different cases of directional sound sources. For example a dipole sound source, which is modeled with use of two point sources.

5.1.5 Initial and boundary conditions

The 3 dimensional wave equation in (19) for the pressure p with given boundary and initial conditions is difficult to solve for arbitrary boundaries. Boundaries in this context may be walls. For perfectly reflective walls a boundary condition that can be used to solve the differential wave equation is:

$$\frac{\partial p}{\partial n} = 0 \tag{23}$$

It is called Neumann boundary condition. n is the direction of the boundary normal, pointing into the boundary, so the opposite way of surface normals normally used in computer graphics for visibility or lighting. Another boundary condition is the Dirichlet boundary condition that directly specifies a value of the function on the boundary (e.g. $p = 0$ for each point laying on the boundary).

Due to the time dependency of the problem, also initial conditions are required. These initial conditions specify a value for every point in space for a certain time (e.g. at time t_0).

The number of conditions per variable required to solve the equation is equal to the order of the highest derivative with respect to that variable. So in case of the 1D wave equation (15), two boundary conditions are required for the spatial variable x and two initial conditions for the time variable t.

5.1.6 Example: Simulating 2D wave propagation in a rectangular area

Before explaining more advanced numerical methods a basic example of the simulation of wave propagation will be given for a 2 dimensional case.

The starting point is the 2 dimensional wave equation (19). Boundaries of the wave propagation should lay between 0 and a in x-direction and 0 and b in y-direction and are walls where the pressure can never exceed 0. Therefore the boundary conditions are:

$$p(x,0,t) = p(x,b,t) = 0, \quad 0 \le x \le a, \; t \ge 0 \tag{24}$$

$$p(0,y,t) = p(a,y,t) = 0, \quad 0 \le y \le b, \; t \ge 0 \tag{25}$$

Still initial conditions are required. A very general approach is to define the initial conditions as:

$$p(x,y,0) = f(x,y), \quad \frac{\partial p}{\partial t} = 0 \tag{26}$$

So with the solution of this wave equation which is explained in greater detail in [Drmota et al. 2008, page 333] it will be possible to plug any suitable equation into the solution as initial condition.

To solve this partial differential equation first a separation of variables can be applied. To separate the t term from the x and y terms. For the x and y terms a second separation of variables is applied. This reduces the problem to a problem of solving two ordinary differential equations in x and y. For separation of variables it is required that the boundary conditions also only depend on the x and y variables respectively.

After the problem was reduced to solving two ordinary differential equations, a constant value (often called λ) introduced by separation of variables needs to be determined. Only non trivial solutions (solutions $\neq 0$) and solutions that satisfy the boundary conditions are to be considered. To find them, the initial conditions are plugged in the solution of the ordinary differential equation to form a system of equations. A non trivial solution is a solution where the determinant of the system matrix of these equations is zero. So the determinant of this system matrix is written down and solved for λ. This process is repeated for the x and y term. (actually not a single λ is sufficient, since the separation was executed twice) The solutions are called Eigenvalues. With these Eigenvalues also Eigenfunctions can be specified (just a suitable assignment for the constants resulting from solving differential equations need to be found; once again the Eigenfunctions should be non trivial).

After solving for x and y, the results are plugged into the first separation of variables which still possesses two arbitrary constants. With superposition and Fourier series expansion finally values for the coefficients and the solution to the equation can be written as:

$$a_{n,m} = \frac{4}{ab} \int_0^a \int_0^b f(x,y) sin\left(\frac{n\pi}{a}x\right) sin\left(\frac{m\pi}{b}y\right) \, \mathrm{d}y \, \mathrm{d}x \tag{27}$$

$$u(x,y,t) = \sum_{n,m \ge 1} a_{n,m} cos\left(\pi c \sqrt{\frac{n^2}{a^2} + \frac{m^2}{b^2}} t\right) sin\left(\frac{n\pi}{a}x\right) sin\left(\frac{m\pi}{b}y\right) \tag{28}$$

This last relatively complicated task of superposition and use of Fourier series is mainly required, because we want to consider an arbitrary function for wave propagation, while the boundary conditions on the other hand are only constant values. For simulation in Matlab boundaries for $a = 40$, $b = 40$, $n = 60$ and $m = 60$ where chosen. For the initial function a 2 dimensional Gauss function was used. These values for n and m seemed to be necessary. With less resolution the underlaying frequencies that make up the function shape were too visible. Result images are shown in figure 20.

The advantage of this solution, which can be solved efficiently due to the relatively simply boundary conditions, is that with the given solution, any point in time can simply be calculated by plugging into the solution. For a simulation where things may change on the fly or every step should be visualized, this approach may not be very suitable.

5.1.7 Finite Difference Methods *FDM*

FDMs partition space into a discreet grid of points. The lines along the grid lay parallel to the coordinate axes. Typically an equidistant grid is used.

First the method will be explained in a little more detail for the one dimensional wave equation. The inhomogeneous wave equation, tightly connected to homogeneous wave equation (15), looks as following:

$$\frac{\partial^2 p}{\partial t^2} - c^2 \frac{\partial^2 p}{\partial x^2} = F(x,t). \tag{29}$$

To solve the equation with use of FDM, discretization of the space and also of the time has to be applied.

In the case of a function that does not depend on time, a differential equation only requires a certain number of boundary conditions and

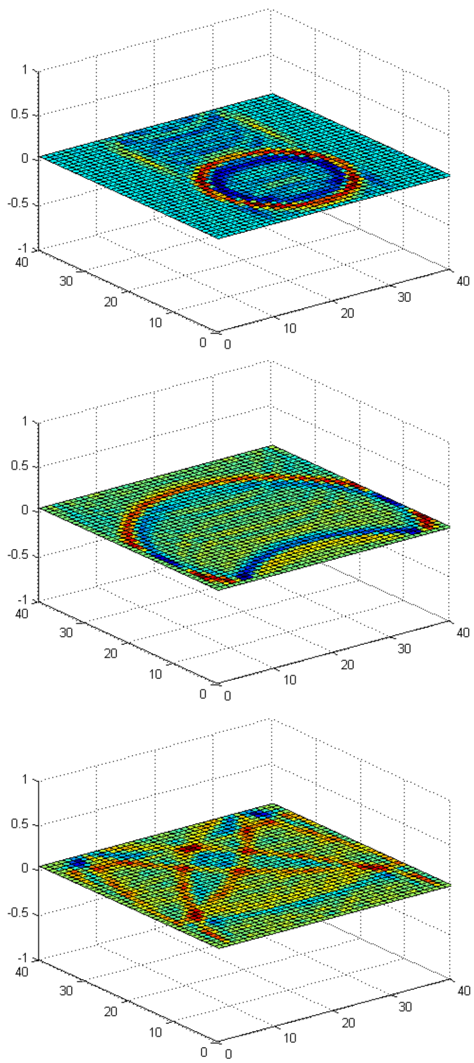

Figure 20: 2D rectangular wave propagation simulation in Matlab.

the problem can be called boundary value problem. Since the function also contains the time axis, they also contain the initial value problem. Along the time axis only information of the two directly preceding and/or following time steps are required.

The FDM makes use of the difference quotient, that can be used instead of exact derivatives. The approximation for a function $f(x)$ in a one dimensional case can be written as:

$$\frac{\partial f}{\partial x} = \frac{f(x+\Delta x) - f(x)}{\Delta x} + error. \qquad (30)$$

Δx is supposed to be smaller then 1 and is the distance between grid points. The equation makes use of the forward finite difference. Backward and central finite difference can be found in [Karow 2003] or [Li 2001]. The equation for the second order derivative, which is particularly important in wave propagation, is:

$$\frac{\partial^2 f}{\partial x^2} = \frac{f(x+\Delta x) - 2f(x) + f(x-\Delta x)}{\Delta x^2} + error. \qquad (31)$$

The discretization partitions space along the x axis in points of:

$$x_i = x_0 + i\Delta x \qquad for \quad i = 1, ..., N. \qquad (32)$$

For each point a pressure value is approximated as:

$$P_i \approx P(x_i) \qquad for \quad i = 1, ..., N-1. \qquad (33)$$

With use of those two equations the discretization of space and time can be applied for the wave equation as:

$$\frac{P_i^{m-1} - 2P_i^m + P_i^{m+1}}{\Delta t^2} + c^2 \frac{P_{i-1}^m - 2P_i^m + P_{i+1}^m}{\Delta x^2} = F_i^m. \qquad (34)$$

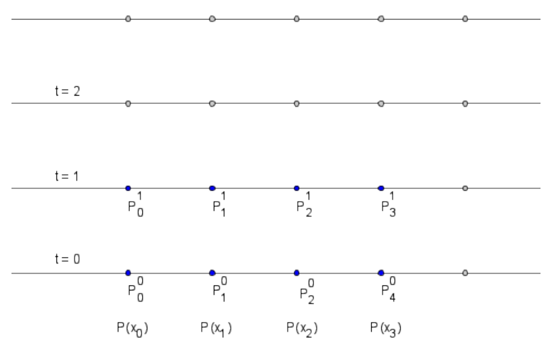

Figure 21: FDM for a one dimensional case. The horizontal lines can be imagined as a real physical object. The vertical axis symbolizes the time steps.

The indices of m stand for time steps (that have to stay within certain boundaries, also see [Karow 2003]). The indices of i denote the spatial steps in x direction. Figure 21 shows the pressure values P. The horizontal axis refers to the distribution along the x axis. A horizontal tube in which sound can propagate could be imagined. The vertical axis corresponds to the time steps. i and m are only the indices for the discrete elements P. But those elements lie within the real scene with a distance of Δx along the propagation axis and Δt along the time axis.

After this discretization two ways of solving the problem can be used. Either all equations in the form of (34) and the boundary conditions form a linear system of equations that can be solved. Or the equation can be written in a way that P_i^{m+1} can be determined directly by the two preceding time steps at m and $m-1$ which are already given by the initial conditions. When the latter is used, the step size needs to be small. This way would be a nice approach for a fast simulation of the whole wave field.

The resulting system of linear equations is often written in the form of matrices and vectors. For example the simple function of stationary heat conduction in a one dimensional case:

$$\lambda \frac{\partial^2 T(x)}{\partial x^2} + s(x) = 0 \qquad (35)$$

where λ is a constant and s a function we are not really interested in what they refer to. (This example has only been taken for its simplicity. Still it is better to show it on an applicable formula like this one instead of a constructed one to only show the idea.)

After discretization the formula looks, similar to the previous example, like:

$$-T_{i-1}+2T_i-T_{i+1}=\frac{\Delta x^2 s(x_i)}{\lambda}. \tag{36}$$

Now the system of linear equations can be written in matrix and vector form as:

$$\begin{pmatrix} -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \end{pmatrix}\begin{pmatrix} T_1 \\ \vdots \\ T_{N-1} \end{pmatrix}=\begin{pmatrix} \Delta x^2 s(x_1)/\lambda \\ \vdots \\ \Delta x^2 s(x_N)/\lambda \end{pmatrix} \tag{37}$$

FDMs are sometimes also called finite difference time domain (FDTD) methods.

5.1.8 Finite Element Methods *FEM*

FEMs try to solve the wave equation through numerically stable approximations. They divide space into a finite number of elements. Once again these methods try to reduce the problem to solving a system of linear equations. The number of required grid cells is proportional to the highest frequency [Bertram et al. 2005] . They are normally performed in the frequency domain.

A detailed introduction to FEM can be found in [Javier Sayas 2008]. The idea behind FEM is a little more involved and will not be discussed here. In practice often special software is used to calculate solutions with the FEM.

5.1.9 Digital Waveguide Mesh

The Digital Waveguide Mesh is a structure of digital waveguides, which are connected with scattering junctions [Duyne et al. 1993].

A digital waveguide is an element that implements the solution of the one dimensional wave propagation. It can be modeled with use of two so-called directional digital delay lines. Delay lines simply delay a signal for a specified time (also called sample units in digital signal processing). This model is tightly connected to the solution of the planar wave equation in equation (18). The amplitude of the pressure at the waveguide is the sum of the left and the right traveling waves. A digital waveguide is shown in figure 22.

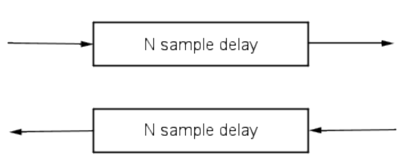

Figure 22: Digital waveguide

Waveguides can be imagined as strings or as pressure traveling through a tube. The delay of a waveguide would depend on its length and will be related to the propagation speed of sound in the considered medium.

The next concept used in digital waveguide meshes are junctions. The basic case is a lossless junction of N waveguides. Junctions need to distribute an incoming wave along all directions. The equation for such a junction can be written as:

$$v_J=\frac{(2\sum_i R_i v_i^+)}{\sum_i R_i} \tag{38}$$

for the junction velocity and

$$v_i^i = v_J - v_i^+ \tag{39}$$

for each of the outgoing waves. The variable v_i^+ is the incoming velocity. The variable R is the impedance of a string. If we want to use the characteristic impedance Z_0 from equation (13), they are related by $R=1/Z_0$, which can be verified with help of [Murphy et al. 2001].

With waveguides and junctions a 2 dimensional waveguide mesh can be created like shown in figure 23. Also a 3 dimensional version is possible.

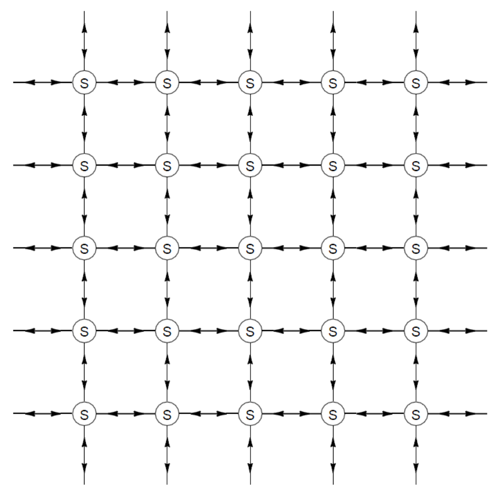

Figure 23: 2 dimensional waveguide mesh. Junctions S are connected with waveguides

In figure 24 the 2 dimensional transversal wave propagation is shown that is described in [Duyne et al. 1993].

5.1.10 Boundary Element Methods *BEM*

These methods, in contrary to the finite element method, only discretize the boundaries of a region. Like FEM they are also regularly performed in the frequency domain.

5.1.11 An exemplary numerical method using a FDM

The following explanation of a numerical method based on the finite difference method and modal analysis was presented in 'Accelerated Wave-based Acoustics Simulation' [Raghuvanshi et al. 2008].

The initial problem is the three dimensional wave (19) equation already mentioned a few times.

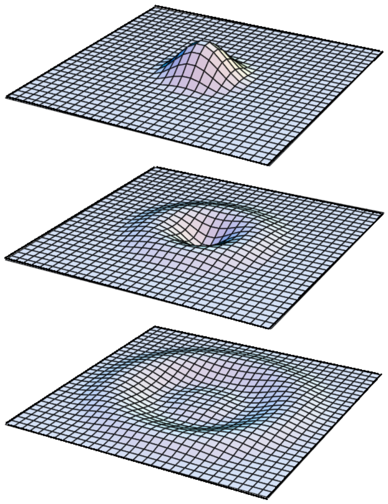

Figure 24: Application of the 2 dimensional waveguide mesh for transversal wave propagation in a membrane.

- First only space (not time) is discretized using an FDM. A grid size has to be chosen. The required grid must suffice the highest frequency (or shortest wavelength). A grid size of at least 1/8 to 1/10 of the shortest wavelength needs to be taken, since no assumption of where the actual waves may lie can be made. The resulting formula after discretization includes a matrix K similar to the matrix in equation (37) and a vector of pressures at discrete points stored in P:

$$\frac{\partial^2 P}{\partial t^2} + KP = F(t). \tag{40}$$

  Notice that the function F or forcing term, only depends on t. This result corresponds to the matrix notation in FDMs. The used boundary condition is the Neumann boundary condition. (It takes the first derivative of the function as boundary condition values.) It is also integrated into the matrix K. This system of ordinary differential equations has still to be solved.

- Next step is to diagonalize the matrix K. This is achieved by modal analysis which uses the eigenvectors and eigenvalues, the formula can be found in [Raghuvanshi et al. 2008].

- Damping due to air absorption is inserted in the formula by adding:

$$\alpha \frac{\partial p}{\partial t} \tag{41}$$

  to the original wave equation. The equation therefore is:

$$\frac{\partial^2 p}{\partial t^2} + \alpha \frac{\partial p}{\partial t} - c^2 \nabla p = F(x,t). \tag{42}$$

- After diagonalization the remaining system of ordinary differential equations can be formulated with use of a vector called the mode vector M. The equations are now to be solved for the elements of the mode vector m_i. This mode vector is required instead of the pressure vector P to express the diagonal matrix in multiplication with only a single vector in the equation. This results in the system of equations to be explicitly solved, since only a term of the form of

$$\frac{\partial^2 M}{\partial t^2} + \Lambda M = \tilde{F}(t) \tag{43}$$

  remains. Where Λ is the diagonal matrix of the Eigenvalues.

- From the mode vector M the pressure value P can be directly calculated.

The diagonalization of K only needs to be performed once. The mode vector M has to be found for every time step by applying the new pressure values.

The complete paper [Raghuvanshi et al. 2008] includes more formulas. Still the amount of detail is not that high in the paper, so a direct implementation may be a bit more involved. Additional literature about FDM and matrix diagonalization with use of modal analysis are recommended.

In addition to this basic approach also an additional partitioning of the space was applied during precomputation. Otherwise the precomputation time of about an hour would have been much too high. The domain partitioning lead to a reduction of performance during runtime, which had to be solved as well.

The method was used to simulate sound propagation in real time. For a room of 12m x 13m x 7m the computation speed reached about 3 FPS. The advantage of this too slow method is that also interference and diffraction are accounted for. Still the computation requirements are too high for use on todays personal computers. Wall impedances are also not taken into account, as all walls were set to be perfectly reflective.

5.1.12 A Precomputed method using a numerical solution

The project 'Precomputed Wave Simulation for Real-Time Sound Propagation of Dynamic Sources in Complex Scenes' [Raghuvanshi et al. 2010] which will be explained now, is the succeeding paper to [Raghuvanshi et al. 2008]. It uses a numerical method to precompute exact acoustical properties for static scenes and moving sources and listeners. It was tested in Valve's Source engine.

The overview of the approach is:

- A numerical solution with use of FDTD (FDM) is precomputated. Space is divided in a grid. Listeners and sound sources are placed at each grid point. An impulse response (see Section 3.2.5) for a pair of listener and sound source position is saved. To reduce storage and computation times only a two dimensional positioning of the listener on a ground plane is used.

- During runtime interpolation between the impulse responses at two grid points is performed. The resulting impulse response is then combined with use of a HRTF for each ear. Finally the sound file is convolved with the complete impulse response for each ear to create the spatial impression.

The algorithm uses the psychophysical knowledge of early reflections and late reverberations (see Section 3.2.5) to simplify calculations. Once again, early reflections are the main spatial cue to locate a sound source, while late reverberations are a diffuse impression creating the feel of a room.

The precomputation starts with a late reverberation simulation in the center of each room with a listener at the same position for 1-2 seconds. This approach is to determine on the one hand when the early reflection transition to the late reverberation takes place. This is done by a threshold and will be explained later. On the other hand the simulation is taken to form the late reverberation impulse response (LRIR) filter. The length of the late reverberation and the transition from early reflection to late reverberation depend mostly on the size of the room.

After calculation of the LRIR for every room, the scene is divided into a grid and for each possible listener and sound source position an early reflection impulse response (ERIR) is calculated.

For the numerical simulation the approach of [Raghuvanshi et al. 2008] or [Raghuvanshi et al. 2009] is used, which were explained in the previous section. The latter approach is improved with specific wall impedances, which determine how much absorption at a wall occurs. It uses additional partitioning of space to increase computation time. Due to the high computational demands of the method, only frequencies below 1000 Hz are considered with a corresponding grid of about 10 cm. Higher frequencies are extrapolated with help of these low frequencies. The spatial resolution for listener positions during run time is decreased from the mentioned 10 cm to about 1 m to improve performance during runtime.

Since band limiting to 1000 Hz may cause the signal to be of regular steps of $0.5ms$, a low random factor is added or subtracted. This is because higher frequencies are not taken into account. The maximum resolution therefore is a periodic 1000 Hz signal. The threshold between early reverberations (ER) and late reverberation (LR) is taken when the number of peaks in the impulse response (IR) is more than 500.

Again due to the band limited calculations, not enough peaks in the impulse response for late reverberations are occurring. Gaps are therefore filled stochastically. After that step the final LRIR is ready to be used. The late reverberations are as already mentioned only calculated for each room, in their approach the user has to define rooms manually.

The ERIR holds a few problems: due to band limitation it muffles the sound (frequencies over 1000 Hz were not calculated) and it needs a lot of storage capacity. To solve this problem the ERIR is split into a few significant amplitude peaks in time domain and a frequency trend. The frequency trend is found with dividing the original peak signal in frequency domain through the simplified significant peak signal in frequency domain. The only thing left to manage is extrapolating the frequency trend for frequencies higher than 1000 Hz. The peaks and the frequency trend are stored. Figure 25 shows this process.

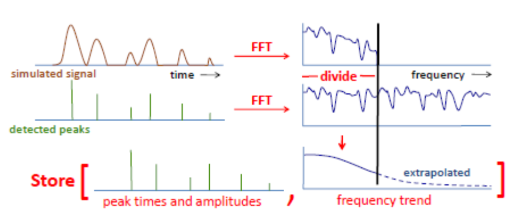

Figure 25: Solving ERIR problems. First row shows the original peaks created through numerical simulation. Second row only looks at the selected peaks. In the last row the way an ERIR is stored is shown [Raghuvanshi et al. 2009].

During runtime the algorithm needs to interpolate between grid points. Interpolation between impulse responses is a difficult problem since peaks of two successive impulse responses correspond to each other in a way that they belong to the same wavefront. The correct way to interpolate peaks would be to try to find corresponding peaks and interpolate their position in time and amplitude. Their algorithm only considers peaks close enough to each other along the time axis (wavefronts can not move faster than the speed of sound in air) and then creates peak pairs dependent on their amplitude difference. These peak pairs can then be interpolated over time and amplitude.

The final step includes the actual interpolation between 8 source positions and 4 listener positions. Due to binaural perception, the peaks within the calculated signal from the source to the listener would require a spatialization. Which is an application of a HRTF, that perturbs the signal due to its direction. This would make the storage of all directions of incoming peaks/wavefronts in the IR necessary. Since this is hardly manageable the algorithm makes use of the psychoacoustic *precedence effect*. This is also called the law of the first wavefront and states that the direction of an incoming sound is in the first place dependent on the direction of the first incoming wavefront. The algorithm therefore only spatializes the first peak in the IR correctly. Subsequent peaks will be placed in random spatial locations. The importance of correct positioning of subsequent wavefronts regarding the imagined position of a listener within the scene are still to be examined.

Disadvantages of this method are that only 10 sources can be used and the very high memory requirements during runtime. In addition it is not simple to implement and has many special cases to be considered, but it allows spatial sound to be simulated in real-time, including diffraction and reflections for moving sounds and a listener.

5.2 Geometric methods

The basic idea of geometric methods is to model acoustic waves as rays, similar to methods used in computer graphics or optics.

The initial problem is to find propagation paths of sound waves influenced by the physical phenomena of reflections. Also diffraction can be modeled with advanced geometrical methods. Refraction is normally not considered in basic methods.

The following part will explain the most currently suggested and used methods of geometric sound propagation.

5.2.1 Image source methods

Image source methods model reflections by adding additional sound sources to the scene. These sound sources are called virtual sound sources. Each virtual source models one reflection and is therefore placed behind a reflecting plane. For a stationary sound source and room geometry, but a moving listener, these source positions can be precomputed. To simulate mirroring along a plane, only a mirror-inverted duplicate needs to be created behind that plane. Figure 26 shows a top down view at a scene with virtual sound sources to simulate reflections.

This is particularly simple in a rectangular room. For irregular shapes it is more difficult to find the virtual sound positions. Additionally, it is required to compute visibility in more complicated or non convex rooms. There are some criteria to determine if these virtual sound sources should be generated or activated [Borish 1984].

- **Validity:** Only reflection along one side of a boundary is valid. Reflection along the backside of a boundary is invalid.
- **Proximity:** Sound sources exceeding a certain distance are discarded. This is required for termination.

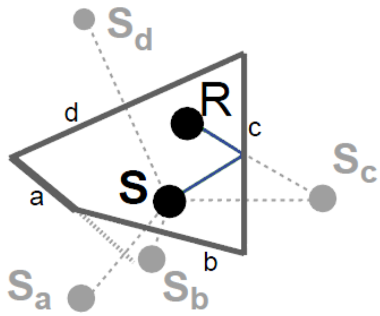

Figure 26: Image source method: with sound source S and virtual sound sources S_a, S_b, S_c, S_d [Funkhouser et al. 2002].

- **Visibility:** When obtuse angles in geometry occur, sound sources become invisible because they are outside the reflection angle. To calculate visibility a line is generated from the listener to the virtual source and intersected with the reflecting plane. If the point lies within the reflecting polygon boundary it is visible. One way to solve this if only convex polygons are used, is to generate vectors from the intersection point to each vertex of the face. The cross product of two successive vectors always point in the same direction if the point is inside the boundary. Otherwise if the point is outside the polygon, not all crossproducts will point in the same direction. The visibility problem is shown in figure 27.

  For higher order of reflections additional visibility needs to be calculated. To do this, the listener needs to be mirrored along reflective planes before higher order visibility calculations. Paths of higher order need additional listener mirroring to figure out visibility.

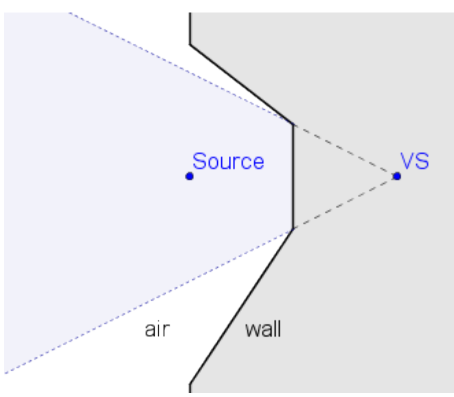

Figure 27: Visibility for the image source method. The listener inside the blue shaded region can hear the virtual sound source VS.

Advantages and disadvantages

The advantage of image source methods is that they are geometrically accurate and free from aliasing issues. They guarantee that all specular reflections up to a certain order of reflections or distance are found.

Their disadvantage lies in the exponentially increasing number of virtual sound sources. For p surface planes and a number of r reflections, $O(p^r)$ virtual sound sources have to be generated. Additionally for all of those virtual sound sources validity, proximity and visibility checks have to be performed. They also only consider specular sound reflections and no refractions or diffractions. Computation cost for dynamic and interactive applications may be too high.

As a result image source methods are normally used in hybrid combinations, or for simple or rectangular scenes with a small number of sound sources.

5.2.2 Ray Tracing

Ray tracing methods generate rays from sound sources in all directions. These rays are reflected at surfaces. All rays that hit a typically volumetric listener (for example a sphere) are audible. This problem is very similar to the problem of solving global illumination, so suitable solutions may also be found in global illumination as started in [Whitted, 1980]. Ray tracing in computer graphics normally refers to the process of casting rays from a point of view into a scene. A schematized image to represent ray tracing is given in Figure 28.

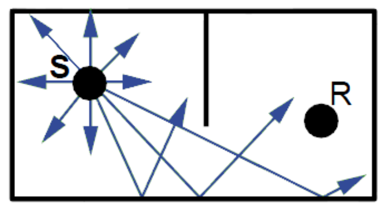

Figure 28: Ray tracing [Funkhouser et al. 2002].

There are many different versions of ray tracing. Variations can be the way of emitting rays from sources. One way is to send them out in fixed directions. Alternatively ray directions can be random (Monte Carlo). Another aspect is that surfaces absorb parts of an incoming wave. Therefore, surfaces need to have absorption coefficients. The best results will be generated if those coefficients differ in wavelength. So the absorption will differ dependent on the wavelength of the incoming sound. Some ray tracing methods can also handle diffuse reflections. Additionally, a diffuse or scattering coefficient can be applied to surfaces. Rays may be reflected in arbitrary directions, dependent on this reflection coefficient. Splitting up rays is normally avoided since it significantly increases computation time [Svensson 2008].

The information gathered by ray tracing will be used to create an impulse response for a sound source. Dependent on this impulse response, a filter can be generated and applied to a sound file, recorded in an anechoic room.

Diffraction for ray tracing has been suggested. This can be achieved with use of the Uniform Theorem of Diffraction (UTD). It has been shown that this theorem originally developed for high frequencies can be used for frequencies of 150 Hz upwards [Tsingos et al. 2001a]. These are based on the Geometric Theory of Diffraction (GTD) that was originally formulated for light hitting edges or corners [Keller 1962].

Advantages and disadvantages

Advantages are its speed and efficiency on GPU hardware. Also higher order reflections can be examined without significant computation cost increase. Even dynamic scenes can be handled. Diffuse reflections can be modeled.

The disadvantage are sampling artifacts and that this is an aggressive method of sound simulation. Aggressive in a way that paths may be lost due to the limited sampling steps. There are no guarantees that all significant paths will be considered. Diffraction is theoretically possible but can not be solved efficiently.

The bigger and more complex the scene, the more rays need to be computed. When propagating rays from a point source these rays will diverge and the farther away a sound source is, the more rays will need to be computed for an evenly distributed sound field. Reflections can further split up ray paths (e.g. reflections on a convex surface or edges). So in general, the more distant a ray has been traveled, the more sampling artifacts will occur. In addition many rays are computed that will never reach the listener.

A way of using ray tracing more efficiently may be to only use direct (non-reflected) rays reaching the listener for exact spatial sound. The remaining rays can be used for estimating area reverberation parameters for a volumetric area the listener is located in. Therefore more computed rays can be used and are not simply discarded.

5.2.3 Beam Tracing

In beam tracing instead of a ray a pyramidal beam area is cast into the scene. A set of volumetric beams can be cast into the scene and therefore cover the whole space. Beams hitting surfaces can be split up and reflected. Also diffraction and transmission are possible. The process of splitting up beams dependent on the surface they hit can be quite complicated and is related to the problem of calculating exact visibility culling in computer graphics. After tracing beams one approach is to model virtual sound sources like in the image source method or use a ray tracing approach. The advantage is that through beam tracing many virtual sound sources can be discarded.

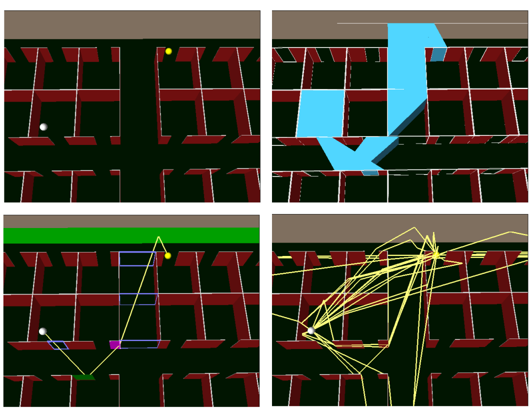

Figure 29: Beam tracing: a) scene with source (white) and receiver (yellow), b) traced beam in cyan, c) path generated for one beam, diffraction is visible at the cyan part of the beam d) paths for all possible beams (there are more than one) [Tsingos et al. 2001b].

One beam tracing algorithm by [Funkhouser et al. 1998] divides the problem in 4 stages. The first two are precomputed, the rest are used in real-time:

- Spatial subdivision: To define spatial regions a Binary Space Partition (BSP) is created. Resulting cells are convex regions separated by planes. Then an adjacency graph can be constructed for easy traversal into connected regions. Each BSP cell is represented by a node within the graph.
- Beam tracing: The beam tracing part requires the positions of sound sources. At each sound source an infinite *current beam* is generated covering all space. This *current beam* starts in the *current cell* the sound source is located in. For all cell boundaries the algorithm recursively calculates beam traversal. If the beam hits an occluding wall the *current beam* is cut off behind that wall. If the beam hits a reflecting surface, specular reflections are created by reflecting the beam on this boundary surface. The remaining *current beam* is the approximated audible region of the sound source. For easy traversal of the beams a *beam tree* data structure is used. The result of beam tracing is shown in Figure 29 b).

  In real-time for every beam:

- Path generation: Path generation is done in real-time. The cell within the BSP tree has to be found. In that cell the *beam tree* nodes have to be checked with the listener position for validity. If a *beam tree* node has been found its ancestors model the traversal of sound from the source to the listener. The changes of the sound due to reflection and transmission are stored within the *beam tree*. Additionally an exact length to the sound source has to be found. This can be easily solved since mirrored virtual sound sources caused by reflections are also stored in the *beam tree*. So the distance is either a direct path from source to listener, or if reflections occur, a path from a virtual sound source to the listener. Generated paths for an example are shown in Figure 29 c) and d).
- Auralization: Finally if paths from the source to the listener are found an overall impulse response can be generated. This is a function of time and intensity modeling the effect of reverberation. For each path from the source to the listener a component is added to the impulse resonse dependent on its time and intensity loss through the scene. In the end the impulse response is applied to an anechoic sound to create the audible result.

This basic beam tracing approach has been modified and improved in succeeding papers [Tsingos et al. 2001b]. One of the major changes to the original version is the introduction of diffraction. Diffraction can be considered using the Huygens-Fresnel principle, the Helmholtz-Kirchoff integral theorem, both mentioned in the physics section, or the Uniform Theorem of Diffraction (UTD). The UTD fits the solution using rays the most. It models higher frequency waves hitting an edge with use of the ray theory. The angle of the incident ray generates a cone of rays when hitting an edge. The aperture angle of the cone equals the incident angle with the edge. The beam tracing and path generation parts are modified to allow diffraction.

Advantages and disadvantages

The advantages of this approach are that there are no aliasing issues and it models the problem geometrically accurate. Also it is interactive for a moving listener. Diffraction can be considered.

The disadvantages are that it can not handle dynamic sound sources or geometry. This is because of the preprocessing step that significantly reduces computation during run time. But the preprocessing does not allow dynamic changes to the geometry or sound sources. Similar to creating lightmaps in computer graphics, where geometry and lights can not move during runtime. Diffuse reflection is also not possible.

5.2.4 Phonon Tracing

Phonon tracing [Bertram et al. 2005] is a particle tracing approach specially for higher level frequencies. It was inspired by photon mapping [Jensen 1996].

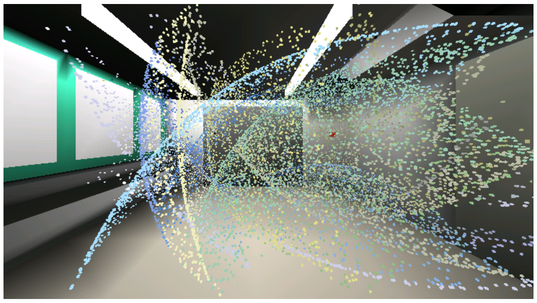

Figure 30: Visualization of phonons with color-coded spectral energy [Bertram et al. 2005].

Phonon paths are frequency dependent 'sound' particles which are traced through a scene starting from each sound source. These phonons carry an energy spectrum, position, direction and distance travelled. The audible frequency spectrum from about 20 to 20000 Hz is simulated. To realize this, the frequency spectrum is divided into a certain amount of intervals and each phonon belongs to one of those intervals. The phonons are emitted randomly from the source with an equal distribution in the shape of a sphere. If phonons hit a surface they are reflected and their energy is reduced according to the material properties at the surface.

Additional the phonons at the reflection points contribute to the phonon map and are stored with the scene, this can be later used for visualization. Maybe this phonon map is created because the method is very similar to photon tracing. Nevertheless for the actual hearing result a 3d Gaussian function is used to find the incident phonons at a listener position. Unlike in lighting this is a time dependent process, so we are interested in the phonons at a certain time step at the listener.

This method also allows to create a visualization of sound propagation as shown in Figure 30. It also provides images of the overall frequency distribution, which can be used to improve sound quality in a room.

This paper also talks about the use of an acoustic bidirectional reflective distribution function (BRDF). As already stated in previous chapters some physical effects are angular dependent, but total specular reflection is not dependent on the angle. The effect that would be possible to model are diffuse reflections, but as already mentioned these are not that important for sound, since most reflections are specular due to the larger wavelength of sound.

Advantages and disadvantages

Advantages are that it is very accurate, considers frequency dependent absorption and diffuse reflection. Disadvantages are that no diffraction is considered. The computation cost is high and no dynamic sound or geometry is possible. Therefore it is not really suitable for real-time application.

5.2.5 Frustum Tracing

Another promising technique is frustum tracing as explained in 'Interactive sound rendering in complex and dynamic scenes using frustum tracing' [Lauterbach et al. 2007].

Instead of tracing rays, in frustum tracing convex polyhedrons are cast into the scene, in a similar way to beam tracing. Unlike beam tracing in which a beam approaching a polygon is clipped in an exact way, in frustum tracing the frustum is uniformly split up into smaller sub-frusta. Each frustum is represented by one sample ray in its center. This ray is clipped against geometry and causes a kind of discrete clipping along geometry. Splitting up of a frustum to sub-frusta is only executed when required. The shape of such frusta is shown in Figure 31.

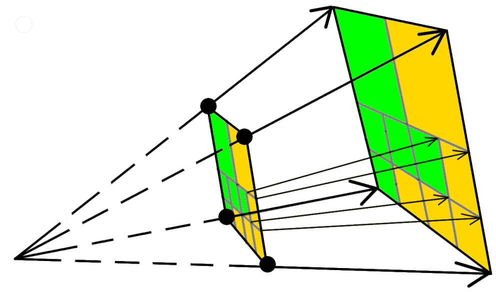

Figure 31: Frustum tracing casts volumetric frusta into the scene. These can only be split and merged in frusta with quadratic front and back planes.

Each time a frustum hits an object, additional frusta are created for transmission and reflection. If the entire frustum hits an object this process is simple. Corner edges are used to generate reflection paths. But if different sub-frusta hit different surfaces, each sub-frustum would have to be treated individually. This may lead to a very high number of new frusta and the algorithm could therefore grow exponentially. To avoid this neighbor frusta are treated together if they hit an object with the same normal and material property. These frusta are then combined.

Advantages and disadvantages

Advantages of frustum tracing are that dynamic sources, listeners and geometry are possible. Compared to other methods it is fast. Also diffraction can be accounted for. Disadvantages are that no diffuse reflection is possible. Additionally, some important contributions may not be considered because of sampling artifacts.

5.3 Solutions to spatial audio related problems

The sections above explained ways to render sound in 3d environments. This last part will now concentrate on other problems arising from the structure of common audio rendering hard- and software.

5.3.1 Limited number of available sound sources in hardware

After simulating a sound scenario, heavy DSP needs to be performed to finish auralization. This demand on computation power leads to limitations of audio hardware. Todays audio hardware supports about 16 to 128 processing channels. Each of these channels can be use to render a sound source. The usage of some of the methods explained above requires a large number of virtual sound sources. These can exceed the number of available hardware processing channels. To solve this problem a solution has been presented in 'Perceptual Audio Rendering of Complex Virtual Environments' [Tsingos et al. 2004b].

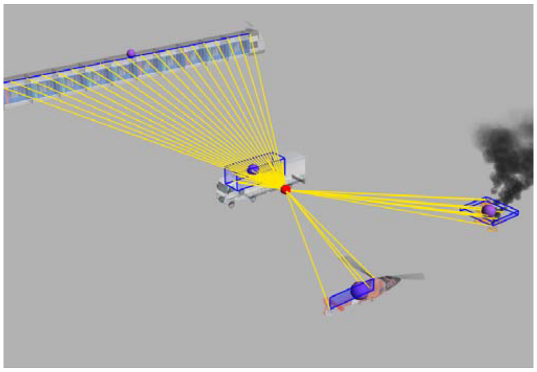

Figure 32: Clustering of sound sources. The listener is red, the sound clusters centers purple and yellow lines point towards the individual sound sources.

Their algorithm that exploits binaural masking and uses clustering which is shown in Figure 32. It can be divided in four steps:

- Sorting cues due to binaural loudness and removing those overlaid by others. First a loudness value for each sound source is calculated from distance, frequency dependent attenuation and the HRTF for each ear. This happens every time step and is averaged over a few frames for smoother values. Then these loudness values are sorted and added up starting from the maximum going to the minimum. A binaural masking threshold is calculated from the sum of preceding values until binaural masking can be observed and the sources with lower loudness can be clipped. The idea behind that threshold is further explained in [Painter et al. 1997].
- Grouping of sounds into clusters: Clusters will be new point sources representing all sources within the cluster. For clustering they use the farthest first traversal algorithm by Hochbaum-Shmoys. A number of sources with high loudness influence and a maximum distance to each other will be defined as cluster centers. Following sound sources will be added to their closest clusters.
- A sound signal is generated for each cluster. All sound sources within a cluster are then modified dependent on their path towards the listener. So each source will be modified dependent on its spatial properties. Sources are then combined to form the overall sound source for the cluster.
- In the last step the clusters can be sent to a standard (spatial) audio hardware to generate the final result.

5.3.2 Speeding up audio with the GPU

The motivation of using the GPU for sound propagation and processing lies in the fact that most conventional sound cards do not allow to run custom programs on the sound card like shaders on the GPU. Typical sound cards only use 'fixed function' pipelines, which evolve slowly [Gallo and Tsingos 2004]. Also many previously explained methods require computations like ray casting which have been already implemented efficiently on the GPU.

An overview of GPU based approaches is given in [Cowan and Kapralos 2008]. Three examples are referenced.

The first example is a GPU based ray casting algorithm for sound simulation. It is presented in [Jedrzejewski 2004]. The GPU implementation was compared with a CPU implementation. The GPU implementation showed to be on the order of 10 more computationally efficient.

The second example was a numerical approach on the GPU found in [Rober 2006]. Which uses a waveguide approach on the GPU, implemented with use of 3d textures and framebuffer objects. It has been calculated mainly for low frequencies and suggests a combination of numerical approaches for low frequencies and geometric approaches for high frequencies. With the use of GPU support this may be an interesting idea.

The third example is found as short summary in [Gallo and Tsingos 2004]. It compares two digital signal processing effects processed on the GPU with the CPU. Some difficulties were pointed out because of the lack of support of floating point textures on the GPU. Also large 1D textures which are easy to index would have been required. In a conclusion they suggest to improve audio hardware with programmable shaders, similar to the GPU.

In [Cowan and Kapralos 2008] usage of the GPU for HRTF filtering is described. Their approach shows that GPU based HRTF filtering is much faster then CPU filtering but can introduce errors. Their approach filters a signal of 60,000 samples with a HRTF filter at 249 FPS on the GPU compared to 11 FPS on the CPU.

6 Conclusion

Today there is still a noticeable gap between room acoustics and implementations in virtual environments. This is due to the fact that room acoustics focus on sound, while in virtual environments graphics are often of a much higher importance. Room acoustics try to simulate with the use of physically accurate models. Interactivity is less important than accuracy. Virtual environments often only try to create a nice impression, while keeping an application interactive, or even better, running as fast as possible. Often the amount of resources for sound in games is very limited. It is often impossible to figure out how sound rendering was achieved in current game engines, because of missing publications. It seems valid to assume that current games often do not consider real geometry for sound propagation.

Dependant on the goals of an audio application some factors may be of higher and some of less relevance. For virtual environments especially interactivity and dynamics are important. All important perceptible physical phenomena should be accounted for. Namely diffraction, refraction, diffuse and specular reflection. Also frequency dependency is very important and has to be somehow considered.

Methods shown in this report often lack one or more of these requirements. Most of them are physically correct but slow. For a useful simulation of spatial audio in virtual environments, at least movable listeners and sound sources are required. Solutions not allowing movable listeners can not be called interactive. Solutions not allowing movable sound sources are of some, but very low relevance. Static propagation of sound sources allows to simulate static, ambient sound realistically. This is useful when a movable listener moves towards or away from such sound sources and will hear various nice sound effects, such as noticeable diffraction along edges. A problem that arises from considering only static sound sources is, that in realistic virtual environments always movable sound sources will exist. A listener moving through a scenario will emit sound through movement. (e.g. footsteps, driving noise) In addition lively virtual environments will always include some other living moving beings. Excluding moving sound sources will not produce very realistic or immersive results. A system that only considers static sound sources will produce a noticeable gap between moving and static sound sources. Therefore, a way to handle movable sound sources has to be found.

The works of [Chandak et al. 2008] and [Raghuvanshi et al. 2010] allow dynamic sound sources and listeners. The first sounds promising in the way that dynamic sources, listeners and also geometry are supported. Still the approach seems to be too slow for real-time usage. Their tested game scene taken from Quake 3 Arena only runs at about 4 FPS for about 14000 polygons. The second approach [Raghuvanshi et al. 2010] seems to produce really good results, but it uses only precomputation and no dynamic geometry is supported. Anyway their results sound and look a lot more promising.

After establishing a way to simulate more realistic sound it is important to measure the influence on people using such systems. Some questions may be of interest: Is there any increase in performance or subjective realism or immersion? To what extent does realistic sound influence our perceptual positioning within a scene? Is it possible to percept positioning within a room by reverberation parameters? Does it make a difference to stand close to a wall or in the center of a hall? (This is a phenomenon not regularly accounted for in today's virtual reality applications.) Or the very subjective question of: Can sound enhance graphics?

Establishing a basis for measuring the results in a human perceived manner may be of interest. Sound, unlike graphics seems to create a more subjective impression. Often it is not easy to figure out what simulation is better than another. Implementations are often only tested by comparing the impulse response of a real room with the simulated impulse response. But especially in virtual environments a subjective impression may be of higher interest than physical measurements. At least the measured results should be compared considering their importance for human perception.

As shown above, physically correct dynamic simulations can not be handled efficiently yet. For virtual environments, more accurate spatial audio may be of interest for bigger game companies. I believe that the grade of additional realism due to realistic audio rendering could really make the difference between two graphically already stunning games.

References

BEGAULT, D. R. 1994. *3-D sound for virtual reality and multimedia*. Academic Press Professional, Inc., San Diego, CA, USA.

BERTRAM, M., DEINES, E., MOHRING, J., JEGOROVS, J., AND HAGEN, H. 2005. Phonon tracing for auralization and visualization of sound. *Visualization Conference, IEEE 0*, 20.

BLAUERT, J. 1997. *Spatial Hearing : The Psychophysics of Human Sound Localization*. M.I.T. Press.

BORISH, J. 1984. Extension of the image model to arbitrary polyhedra. *The Journal of the Acoustical Society of America 75*, 6, 1827–1836.

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*, 1707–1722.

COWAN, B., AND KAPRALOS, B. 2008. Spatial sound for video games and virtual environments utilizing real-time gpu-based convolution. In *Future Play '08: Proceedings of the 2008 Conference on Future Play*, ACM, New York, NY, USA, 166–172.

DRMOTA, M., GITTENBERGER, B., KARIGL, G., AND PANHOLZER, A., 2008. Mathematik für informatiker.

DUYNE, S. A. V., III, J. O. S., VAN, S. A., JULIUS, D., AND III, O. S., 1993. Physical modeling with the 2-d digital waveguide mesh.

EPIC, 2010. Unreal 3 engine. http://udn.epicgames.com.

FUNKHOUSER, T., CARLBOM, I., ELKO, G., PINGALI, G., SONDHI, M., AND WEST, J. 1998. A beam tracing approach to acoustic modeling for interactive virtual environments. In *SIGGRAPH '98: Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, ACM, New York, NY, USA, 21–32.

FUNKHOUSER, T., JOT, J.-M., AND TSINGOS, N., 2002. "sounds good to me!" computational sound for graphics, virtual reality, and interactive systems.

GALLO, E., AND TSINGOS, N., 2004. Efficient 3d audio processing with the gpu.

JAVIER SAYAS, F., 2008. A gentle introduction to the finite element method.

JEDRZEJEWSKI, M., 2004. Computation of room acoustics on programmable video hardware.

JENSEN, H. W. 1996. Global illumination using photon maps. In *Proceedings of the eurographics workshop on Rendering techniques '96*, Springer-Verlag, London, UK, 21–30.

KAROW, M., 2003. Finite differenzen methode.

KELLER, J. B. 1962. Geometrical theory of diffraction. *J. Opt. Soc. Am. 52*, 2, 116–130.

KUTTRUFF, H. 2007. *Acoustics: An Introduction*. Taylor & Francis Group.

LAUTERBACH, C., CHANDAK, A., AND MANOCHA, D. 2007. Interactive sound rendering in complex and dynamic scenes using frustum tracing. *IEEE Transactions on Visualization and Computer Graphics 13*, 6, 1672–1679.

LI, Z., 2001. Finite difference methods basics.

MANOCHA, D., CALAMIA, P., LIN, M. C., MANOCHA, D., SAVIOJA, L., AND TSINGOS, N. 2009. Interactive sound rendering. In *SIGGRAPH '09: ACM SIGGRAPH 2009 Courses*, ACM, New York, NY, USA, 1–338.

MURPHY, D. T., NEWTON, C. J. C., AND HOWARD, D. M. 2001. Digital waveguide mesh modelling of room acoustics: Surround-sound boundaries and plugin implementation. In *Proceedings of the Cost G-6 Conference on Digital Audio DAFX-168 Effects (DAFX-01*, 198–202.

PAINTER, T., IEEE, S. M., SPANIAS, A., AND IEEE, S. M. 1997. A review of algorithms for perceptual coding of digital audio signals, dsp-97. 179–208.

PIRINGER, J., 2010. Audio vu lecture notes tu vienna.

RAGHUVANSHI, N., GALOPPO, N., AND LIN, M. C. 2008. Accelerated wave-based acoustics simulation. In *SPM '08: Proceedings of the 2008 ACM symposium on Solid and physical modeling*, ACM, New York, NY, USA, 91–102.

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., SNYDER, J., MEHRA, R., LIN, M., AND GOVINDARAJU, N. 2010. Precomputed wave simulation for real-time sound propagation of dynamic sources in complex scenes. *To appear in ACM Transactions on Graphics (SIGGRAPH 2010)*.

RÖBER, N., AND MASUCH, M., 2004. Auditory game authoring from virtual worlds to auditory environments.

RÖBER, N., AND MASUCH, M., 2004. Interacting with sound an interaction paradigm for virtual auditory worlds.

ROBER, N., S. M. M. M., 2006. Waveguide-based room acoustics through graphics hardware.

SVENSSON, P. 2008. The early history of ray tracing in room acoustics.

TAYLOR, M. T., CHANDAK, A., ANTANI, L., AND MANOCHA, D. 2009. Resound: interactive sound rendering for dynamic virtual environments. In *MM '09: Proceedings of the seventeen ACM international conference on Multimedia*, ACM, New York, NY, USA, 271–280.

TREYARCH, 2009. World at war audio technology. http://www.callofduty.com/intel/185.

TSINGOS, N., FUNKHOUSER, T., NGAN, A., AND CARLBOM, I. 2001. Modeling acoustics in virtual environments using the uniform theory of diffraction. In *SIGGRAPH '01: Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, ACM, New York, NY, USA, 545–552.

TSINGOS, N., FUNKHOUSER, T., NGAN, A., AND CARLBOM, I., 2001. Modeling acoustics in virtual environments using the uniform theory of diffraction.

TSINGOS, N., GALLO, E., AND DRETTAKIS, G. 2004. Perceptual audio rendering of complex virtual environments. *ACM Trans. Graph. 23*, 3, 249–258.

TSINGOS, N., GALLO, E., AND DRETTAKIS, G. 2004. Perceptual audio rendering of complex virtual environments. In *Graphics, Proceedings of SIGGRAPH 2004*.

TSINGOS, N., 2009. Pre-computing geometry-based reverberation effects for games.

VALVE, 2010. Source engine. http://source.valvesoftware.com.

WHITTED, T. 1980. An improved illumination model for shaded display. *Commun. ACM 23*, 6, 343–349.