

Part I

ROOM ACOUSTIC MEETS SIGNAL PROCESSING

2 ELEMENT OF ROOM ACOUSTICS

2.1	Sound Wave Propagation	9
2.1.1	The Acoustic wave equation	10
2.1.2	... and its solution as Green's function	11
2.2	Acoustic Reflections	12
2.2.1	Large smooth surfaces, absorption and echoes	13
2.2.2	Diffusion, Scattering and Diffraction of Sound	14
2.3	Room Acoustics and Room Impulse Response	15
2.3.1	The Room Impulse Response	15
2.3.2	Simulating Room Acoustics	16
2.3.3	The Method of Images and the Image Source Model	19
2.4	Perception and Some Acoustic Parameters	20
2.4.1	The Perception of the RIR's elements	21
2.4.2	Mixing time	22
2.4.3	Reverberation Time	22
2.4.4	Direct-to-Reverberant ratio and Critical Distance	22



3 SIGNAL PROCESSING AND AUDIO INVERSE PROBLEM

3.1	Signal Model	23
3.1.1	Multichannel Mixing Process	23
3.1.2	Time-Frequency Analysis and Synthesis	23
3.1.3	Artificial Mixtures	23
3.1.4	Impulse Response Models	23
3.2	Audio Inverse Problems	23
3.2.1	General Processing Scheme	23
3.2.2	Some Audio Inverse Problems	23
3.3	Taxonomy through dichotomies	24

4 EVALUATION AND DATASETS

4.1	Metrics	25
4.1.1	Signal-based metrics	25
4.1.2	Perceptual metrics	25
4.2	Data and Dataset	25
4.2.1	Picnic of the Muses dB	25
4.2.2	d'ECHORATE	25

2

Element of Room Acoustics

- **SYNOPSIS** This chapter will build a first important bridge: from ~~the~~ physics to analog signal processing. It first defines sound and how it propagates § 2.1 **defining** the concept of impulse response. Then the interaction with the environment is **show § 2.2**, teasing out the fundamental concept of this thesis: ~~the~~ echoes. By assuming some approximation, the Room Impulse **Response** (RIR) will be defined § 2.3 describing methods to compute or approximate it. Finally in § 2.4 a description of the way the human auditory system perceives reverberation will be reported.

2.1 SOUND WAVE PROPAGATION

According to common dictionaries and encyclopedias,

sound is the sensation perceived by the ear caused by the vibration of air.

Sound has then two aspects: a physical one, characterized by the vibrating air particles; and a perceptive one, involving ~~the~~ an auditory system. Focusing on the former phenomenon, when vibrating objects excites ~~air~~, air molecules starts oscillating, producing zones with different air densities (compressions-rarefactions)¹. Such vibration of molecules takes place in the direction of the excitement, with the next layer of molecules excited by the first layer. Pushing layer by layer forward, a *longitudinal wave* is created. Under ~~a~~-this perspective,

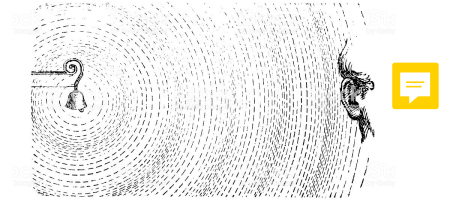
sound is a longitudinal, mechanical wave.

A *wave* is a disturbance that propagates though a medium, which could be solid, liquid or gaseous. The propagation happens at a certain speed which depends on the physical properties of the medium, such as its density and composition. The medium assumed **through out** the entire thesis **is air**. Under the fair assumption of air being homogeneous and steady, the speed of sound can be computed with the following approximated formula:

$$c_{\text{air}} = 331.4 + 0.6T + 0.0124H \quad [\text{m/s}], \quad (2.1)$$

where T is the air temperature [°C] and H is the relative air humidity [%]. The changes in **air pressure** can be represented by a *waveform*, which is a **graphic** representation of a sound.

In general, the sound field is complex **which** can be decomposed as a superimposition of several waves [Kut16].



“Sound, a certain movement of air.”
—Aristotele, De Anima II.8 420b12

Noun: from Middle English sownde, alteration of sowne, borrowed from Anglo-Norman sun, soun, Old French son, from accusative of Latin sonus.

It is legit to interrogate about where is the sounds.

¹ Sound needs a medium to travel: it cannot travel through a vacuum. Unfortunately, there is no sound in outer space.

As opposed to mechanical vibrations in a string or (drum) membrane, acoustic vibrations are *longitudinal* rather than *transversal*, i. e. the air particles are displaced in the same direction of the wave propagation.



FIGURE 2.1: Imagine a calm pool. The surface is flat and smooth. Drop a rock into it. Kerploop. The surface is now disturbed. The disturbance spreads propagates, as well know waves. The medium here is the water surface.

[Kut16] Kuttruff, *Room acoustics*

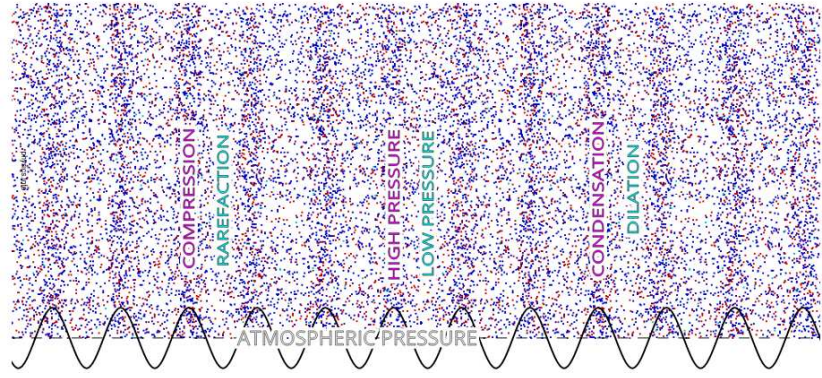


FIGURE 2.2: snapshot of a longitudinal wave in air

We think **at** this process in the light of the classic *source-medium-receiver* model of communication theory.

²example of sources are vibrating solids (e. g. **speakers membrane**), rapid compression or expansion (e. g. explosions or implosions) or air vortices with characteristics frequencies (e. g. flute and whistles).

source is anything that emits or expends energy (waves)²,

medium is the vehicle for carrying waves from one point to another, and

receiver absorbs ~~the~~ such waves.

The behavior of acoustic waves is defined by the acoustic-wave equation. The rest of **the** section is **reproduce** the **derivation** of this equation, re-arranging the derivation presented in [Kut16; Pie19; MM06; Ava19].

[Kut16] Kuttruff, *Room acoustics*

[Pie19] Pierce, *Acoustics: an introduction to its physical principles and applications*

2.1.1 The Acoustic wave equation

[MM06] Marczuk and Majkut, “Modelling of Green function in a rectangular room based upon the geometrical-filtration model”

[Ava19] Avanzini, “Sound in Space”

³ In 1746, d’Alembert discovered the one-dimensional wave equation for music strings, and within ten years Euler discovered the three-dimensional wave equation for fluids.

The symbol $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$ stands for the 3-dimensional *Laplacian* operator.

⁴In fluidodynamics, it comes with the name of the Euler’s equationⁿ

⁵meaning that the medium is an ideal gas undergoing a reversible adiabatic process^s

It is a second-order partial differential equation³ which describes the evolution of acoustic pressure p as **function** of the position \underline{x} [m] and time t [s]

$$\nabla^2 p(\underline{x}, t) - \frac{1}{c^2} \frac{\partial^2 p(\underline{x}, t)}{\partial t^2} = 0. \quad (2.2)$$

The constant c is the sound velocity in the medium with dimension $[\frac{m}{s}]$.

Assuming the propagation of the wave in a homogeneous medium, one can obtain the equation above by combining three fundamental physical laws:

- the *conservation of momentum*⁴,
- the *conservation of mass*, and
- the *polytropic process relation*⁵.

In general **medium** are not uniform and features inhomogeneities of two types: scalar inhomogeneities, e. g. due to temperature variation, and vector inhomogeneities, e. g. due to presence of fans or air conditioning. Although these affect the underlying assumption of the model, the effect are small in typical application of speech and audio signal processing. Therefore they are commonly ignored.

► THE HELMHOLTZ’S EQUATION

The wave equation 2.2 is expressed in the space-time domain (\underline{x}, t) . By applying the temporal Fourier transform, we obtain the *Helmholtz equation*,

i. e.

$$\nabla^2 P(\underline{\mathbf{x}}, f) + k^2 P(\underline{\mathbf{x}}, f) = 0, \quad (2.3)$$

where $k = \frac{2\pi f}{c}$ is known as *wave number* [m^{-1}], that relates the frequency f [Hz] and the propagation velocity c .

Both the wave equation 2.2 and the Helmholtz's equation 2.3 are *source independent*, namely no source is present in the medium. Therefore they are called *homogeneous* as the right-hand term is zero.

Normally the sound field is a complex field generated by acoustics sources. As consequence, the two equation becomes inhomogeneous as some non-zero terms needs to be added to the right-hand sides.

In presence of a sound source producing waves with distribution function $s(t, \underline{\mathbf{x}})$, the wave equation can be written

$$\nabla^2 p(\underline{\mathbf{x}}, t) - \frac{1}{c^2} \frac{\partial^2 p(\underline{\mathbf{x}}, t)}{\partial t^2} = s(t, \underline{\mathbf{x}}). \quad (2.4)$$

Then, the correspondent Helmholtz's equation writes

$$\nabla^2 P(\underline{\mathbf{x}}, f) + k^2 P(\underline{\mathbf{x}}, f) = -S(\underline{\mathbf{x}}, f). \quad (2.5)$$

For instance one can assume an infinitesimally small pulsating sphere locate at $\underline{\mathbf{s}}$ radiating constant acoustic energy at frequency f , i. e. $S(\underline{\mathbf{x}}) = \delta(\underline{\mathbf{x}} - \underline{\mathbf{s}})$. At the receiver position $\underline{\mathbf{x}} \neq \underline{\mathbf{s}}$, the Helmholtz's equation writes

$$\nabla^2 H(f, \underline{\mathbf{x}} | \underline{\mathbf{s}}) + k^2 H(f, \underline{\mathbf{x}} | \underline{\mathbf{s}}) = -\delta(\underline{\mathbf{x}} - \underline{\mathbf{s}}), \quad (2.6)$$

The function $H(f, \underline{\mathbf{x}} | \underline{\mathbf{s}})$ that satisfy Eq. (2.6) is the *Green's function* associated to Eq. (2.3), for which is also a solution.

In the next subsection, we will see that the function H can be interpreted as the free-field *Transfer Function* between the source at $\underline{\mathbf{s}}$ and the receiver at $\underline{\mathbf{x}}$.

2.1.2 ... and its solution as Green's function

THE GREEN'S FUNCTIONS are mathematical tools for solving linear differential equations with specified initial- and boundary- conditions [Duf15]. They have been used to solve many fundamental equations, among which Eqs. (2.2) and (2.3) for both free and bounded propagation.

They can be seen as *the equivalent concept of the impulse responses*⁶ used in signal processing.

Under this light the physic so-far can be rewritten in the vocabulary of the communication theory, namely *input*, *filter* and *output*.

According to Green's method, the equations above can be solved in the frequency domain for arbitrary source as follows:

$$P(f, \underline{\mathbf{x}}) = \iiint_{\mathcal{V}_s} H(f, \underline{\mathbf{x}} | \underline{\mathbf{s}}) S(f, \underline{\mathbf{s}}) d\underline{\mathbf{s}}, \quad (2.7)$$

where \mathcal{V}_s denotes the source volume, and $d\underline{\mathbf{s}} = dx_s dy_s dz_s$ the differential volume element at position $\underline{\mathbf{s}}$.

The requested sound pressure $p(\underline{\mathbf{x}}, t)$ can now be computed by taking the frequency-directional inverse Fourier transform of Eq. (2.7).

By 1950 Green's functions for Helmholtz's equation were used to find the wave motions due to flow over a mountain and in acoustics. Green's functions for the wave equation lies with Gustav Robert Kirchhoff (1824–1887), who used it during his study of the three-dimensional wave equation. He used this solution to derive his famous *Kirchhoff's theorem* [Duf15].

[Duf15] Duffy, *Green's functions with applications*

⁶impulse response in time domain, *trasfer* fuction in the frequency domain.

If one ignores the space integral, one can see the close relation with a transfer function.

Eqs. (2.8) and (2.9) are respectively the free-field transfer function and the impulse response.

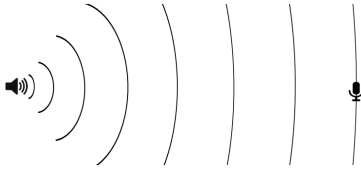


FIGURE 2.3: Visualization of the sound propagation. Since the sensor (i.e. a microphone) is drawn in the far field, the incoming waves can be approximated as plane waves.

It can be shown [Kut16] that the Green's function for Eqs. (2.3) and (2.6) writes

$$H(f, \underline{x} | \underline{s}) = \frac{1}{4\pi \|\underline{x} - \underline{s}\|} e^{-\frac{j2\pi f \|\underline{x} - \underline{s}\|}{c}} \quad (2.8)$$

where $\|\cdot\|$ denotes the Euclidean norm. By applying the inverse Fourier transform to the result above, we can write the time-domain Green's function as

$$h(t, \underline{x} | \underline{s}) = \frac{1}{4\pi \|\underline{x} - \underline{s}\|} \delta\left(t - \frac{\|\underline{x} - \underline{s}\|}{c}\right) \quad (2.9)$$

where $\delta(\cdot)$ is the time-directional Dirac delta function.

As a consequence, the free field, that is in open air without any obstacle, the sound propagation incurs a delay r/c and an attention $1/(4\pi r)$ as function of the distance $r = \|\underline{x} - \underline{s}\|$ from source to the microphone.

According to Eq. (2.9), the sound propagate around a point source with a spherical pattern. When the receiver is far enough from the source, the curvature of the wavefront may be ignored. The waves can be approximated as plane waves orthogonal to the propagation direction. This scenario depicted in Figure 2.3 is known as far-field. As opposed to, when the distance between the source and the receiver is small, the scenario is called near field.

2.2 ACOUSTIC REFLECTIONS

The equations derived so far assumed unbounded medium, i. e. free space: a rare scenario in everyday applications. Real mediums are typically bounded, at least partially. For instance in a room, the air (propagation medium) is bounded by walls, ceiling, and floor. When sound travel outdoor, the ground acts as a boundary for one of the propagation direction. Therefore, the sound wave does not just stop when it reaches the end of the medium or when it encounters an obstacle in its path. Rather, a sound wave will undergo certain behaviors depending on the obstacle acoustics and geometrical properties, including

- reflection off the obstacle,
- diffraction around the obstacle,
- and transmission into the obstacle, causing
 - refraction though it, and
 - dissipation of the energy.

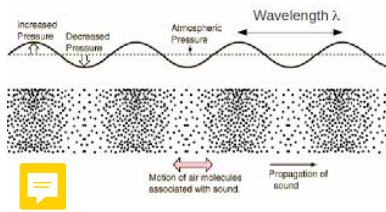
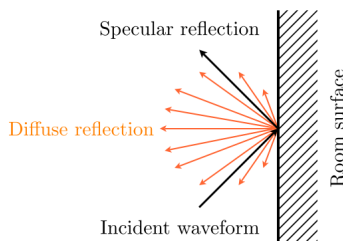


FIGURE 2.4: wavelength

In order, reflections arise typically when a sound wave hit a large surface, like a room wall. When the sound meets a wall edge or a slit, the wave diffracts, namely it bends around the corners of an obstacle. The point of diffraction effectively becomes a secondary source which may interact with the first one. The part of energy transmitted to the object may be absorbed and refracted. Object are characterized by a proper acoustic resistance, called acoustic impedance, which describes its acoustic inertia as well as the energy dissipation. The remaining contribution may continue to propagate causing resulting in the refraction phenomenon⁷.

WHEN SOUND REFLECTS ON AN SOLID SURFACE, two type of acoustic reflection can occurs: part of the sound energy



⁷This is more commonly observed when light pass through different medium, like a prism.

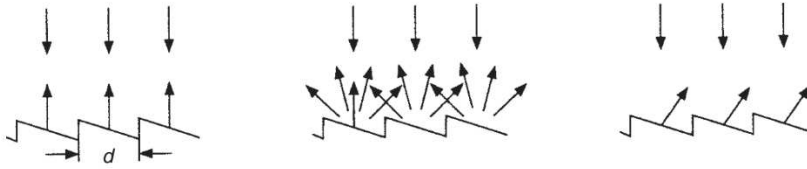


FIGURE 2.7: A reflector having irregularities on its surface with width d much smaller than the sound wavelength λ . Image courtesy of [Kut16].

- is reflected *specularly*, i. e., the angle of incidence equals the angle of reflection; and
- is reflected *diffusely* - or *scattered*, i. e., scatter in every direction).

All the phenomena occurs with different proportions depending on the acoustics and geometrical properties of surface and the frequency content of the wave. In acoustics, it is common to define the *operating points* and different *regimes*⁸ according to the sound frequencies or the correspondent wavelength [m],

$$\lambda = \frac{2\pi}{k} = \frac{c}{f} \quad [\text{m}], \quad (2.10)$$

where f is the frequency of the sound wave.

As depicted in Figure 2.4, λ measures the spatial distance between two molecules in the medium having the same value of pressure.

Using this quantity we can identify the following three response of the objects (irregularities) of size d to a plane-wave depicted in Figure 2.7

- $\lambda \gg d$, the irregularities are negligible and the sound wave reflection is of specular type;
- $\lambda \approx d$, the irregularities break the sound waves which is reflected towards every direction;
- $\lambda \ll d$, each irregularities is a surface reflecting specularly the sound waves.

ALL THIS PRESENTED BEHAVIOR can be described with the wave equation imposing opportune boundary conditions. A simplified yet effective approach - just as in optics - is to model incoming sound waves as *acoustic rays* [DF26; Kro+68]. A ray has well-defined direction and velocity of propagation, and conveys a total wave energy which remains constant. This simplified description undergoes with the name of Geometrical (room) acoustics (GA) [SS15], and share many fundamentals with geometrical optics. This model will be convenient to describe and visualize the reflection behavior hereafter.

2.2.1 Large smooth surfaces, absorption and echoes

Specular reflection are generated by surfaces which can be modelled as infinite flat, smooth and rigid (i. e. *stationary*). As mentioned above, this assumption is valid as long as the surface has dimension much bigger than the sound wavelength. Here the acoustic ray is reflected according to the *law of reflection*, stating that (i) the reflected ray remains in the plane identified by the incident

⁸for instance near- vs.. far-field

Sabine had previously used ray-based acoustics in the early 1900s to investigate sound propagation paths using Schlieren photography. Their impressive visualizations show wavefronts that are augmented with rays that are perpendicular to the wavefronts.

—Savioja and Svensson

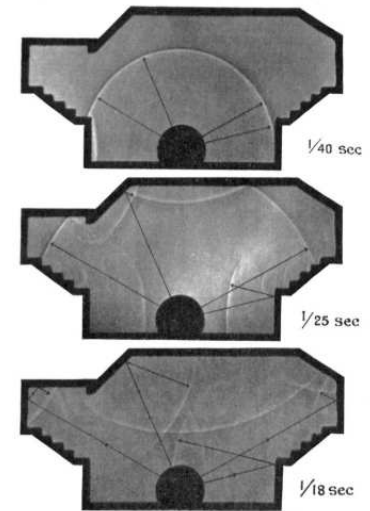


FIGURE 2.6: Photographs showing successive stages in the progress of a Sound Pulse in a Model Section of a Debating Chamber. Image courtesy of [DF26]

[SS15] Savioja and Svensson, “Overview of geometrical room acoustic modeling techniques”

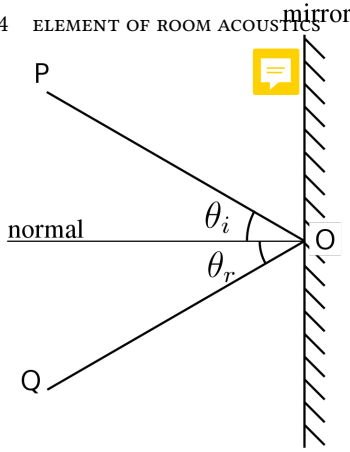


FIGURE 2.8: Specular reflection

ray and the normal to the surface, and (ii) the angles of the incident and reflected rays with the normal are equal.

If the surface S is not perfectly rigid or impenetrable, its behavior is described by the *acoustic impedance*, $Z_S(f) \in \mathbb{C}$. Analytically, it is defined as relation between sound pressure and particle velocity at the boundary. It consists of a real and imaginary part, called respectively acoustic *resistance* and *reactance*. The former can be seen as the part where energy is lost, while the latter as the part where energy is stored.

- THE REFLECTION COEFFICIENT β can be derived by the acoustic impedance for plane waves, i. e. under assuming a far-field regime between source, receiver and surface,.

It measures the portion of energy absorbed by the surface and the incident acoustic wave.

Analytically, it is defined as [Kut16; Pie19]

$$\beta(f, \theta) = \frac{Z_S(f) \cos \theta - Z_{\text{air}}(f)}{Z_S(f) \cos \theta + Z_{\text{air}}(f)}, \quad (2.11)$$

where $Z_S(f)$ and $Z_{\text{air}}(f)$ are the frequency-dependent impedance of the surface and the air respectively, and θ is the angle of incidence.

THE ABSORPTION COEFFICIENT is typically used instead in the context of GA and the audio signal processing. It comes from following approximations [SS15]. (i) The energy or intensity of the plane wave⁹, is considered instead of the acoustic pressure; (ii) dependency on the angle of incidence is relaxed in favor of the averaged quantities; (iii) local dependency on frequencies is relaxed in favor of a frequency-independent scalar or at most a description per octave-band. This assumption are motivated by the difficulty of measuring the acoustic impedance and the possibility to compute an equivalent coefficient a posteriori

Therefore, it is customary to use the absorption coefficient, defined as

$$\alpha(f) = 1 - |\bar{\beta}(f)|^2, \quad (2.12)$$

where $\bar{\beta}$ is the reflection coefficient averaged over the angles θ .

- ECHOES ARE SPECULAR REFLECTIONS which stand out in terms of energy strength or timing [Kut16]. Originally this term used to indicate sound reflections which are subjectively noticeable as a separated repetition of the original sound signal. These can be heard consciously in outdoor scenario, such as in mountain. However, they are less noticeable to the listener in close rooms. In § 2.3.1 a proper definition of echoes will be given with respect the temporal distribution of the acoustic reflection.

2.2.2 Diffusion, Scattering and Diffraction of Sound

Real-world surfaces are not ideally flat and smooth; they are rough and uneven. Examples of such surfaces are coffered ceilings, faceted walls, raw brick walls as well as the entire audience area of a concert hall. When such irregularities are in same order of the sound wavelength, *diffuse reflections* is observed.

[Kut16] Kuttruff, *Room acoustics*

[Pie19] Pierce, *Acoustics: an introduction to its physical principles and applications*

⁹since it is the square magnitude of the acoustic pressure, the phase information is lost.

The word echo derives from the Greek 'echos', litterally "sound". In the folk story of Greek, Echo is a mountain nymph whose ability to speak was cursed: she only able to repeat the last words anyone spoke to her.

In the context of GA, the acoustic ray associated to a plane-wave can be **though** as a bundle of rays traveling **parallel**. When it strikes such a surface, each individual rays are bounced off irregularly, creating *scattering*: a number of new rays are created, uniformly distributed in the original half-space. The energy carried by each of the outgoing ray is angle dependent and it is well modeled though the *Lambert's cosine law*, originally used to describe optical diffuse reflection.

The total amount of energy of this reflection may be computed a-priori knowing the *scattering coefficient* **proper** of the surface material. Alternatively, it can be derived a-posteriori with the *diffuse coefficient*, namely the ratio between the specularly reflected energy over the total reflected energy.

Diffraction waves **occurs** when the sound confronts the edge of a finite surfaces, for instance around corners or through door openings. This effect is **show** in Figure 2.9 At first the sound wave propagates spherically from the source. Once it reaches the **reflector with apertures**, the wave is diffracted, i. e. bended, behind it. It is interesting to note that the diffraction waves produced by the semi-infinite reflector edge allow the area that is “behind” the reflector to be reached by the propagating sound. This physical effect is exploited naturally by the human auditory system to localize sound sources.

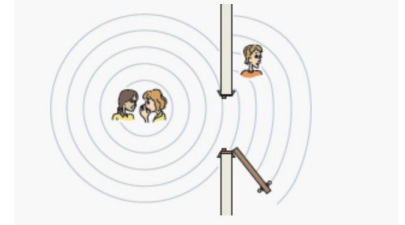


FIGURE 2.9: Sound Diffraction effect.

2.3 ROOM ACOUSTICS AND ROOM IMPULSE RESPONSE

Room acoustics **concerns** with acoustic waves propagating in air enclosed in a volumes with a set of surfaces (walls, floors, etc.), **from** which an incident wave may be **interact** as described in § 2.2. In this context, a

room is a physical enclosure containing the medium and has boundaries limit the sound propagation.

MATHEMATICALLY the sound propagation is described by the wave equation (2.2). By solving it, the Acoustic Impulse Response (AIR)¹⁰ from a source to a microphone can be obtained. In the context of room acoustics, it is commonly referred to as **Room Impulse Response (RIR)**, **usually to put attention of on the geometric relation** between reflections and the geometry of the scene. In this thesis the two terms will be used indistinctly.

¹⁰Acoustic Transfer Function (ATF) **in** the Fourier transform of the AIR

2.3.1 The Room Impulse Response

It is a fundamental concept of this dissertation and it is where physical room acoustic (Green's function/Solution of wave equation) and indoor audio signal processing meets. From now on, we will adopt **an** signal processing perspective and

The RIR is a causal time-domain filter that accounts for the whole indoor sound propagation from a source to a receiver

Figure 2.10 provides a schematic illustration of the shape of a RIR **in comparison** with **measured** one.

RIRs usually exhibits **common structure**. Based on the consideration in § 2.2, they are commonly divided into three components [Kut16]:

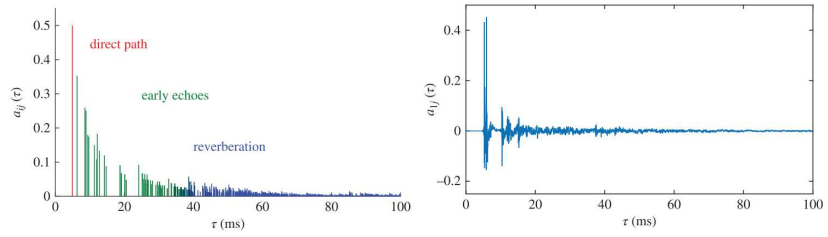


FIGURE 2.10: Schematic illustration of the shape of an RIR and the first 100 ms of a measured one.

$$h^{\text{rir}}(t) = h^d(t) + h^e(t) + h^l(t), \quad (2.13)$$

where

Direct path $h^d(t)$ is the line-of-sight contribution of the sound wave. This term coincides with the spike modeled by the free-field propagation¹¹.

Echoes or Early Refelction are included in $h^e(t)$ comprising few disjoint reflections coming typically from room surfaces. They are usually characterized by sparsity in the time domain and greater prominence in amplitude. **This** first reflections are typically specular and are well modeled in general by the **ISM**¹².

Later Reverberation or simply *reverberation* $h^l(t)$ collects many reflections occurring simultaneously. This part is characterized by a diffuse sound filed with exponentially decreasing energy.

This three components are not only “visible” when plotting the RIR against time, but they are characterized by different perceptual features, as explained § 2.4.

To conclude **with**, let $s(t)$ be the source signal, **sound received is**

$$x(t) = (h^{\text{rir}} * s)(t), \quad (2.14)$$

where the symbol $*$ is the convolution operator.

A part for certain simple scenarios, computing RIRs in closed forms is a cumbersome task. Therefore numerical solver or **approximation model** are used instead.

2.3.2 Simulating Room Acoustics

¹³ The documentation of the **Wayverb** acoustic simulator offers a complete overview of the State of the Art (SOTA) in acoustic simulator methods[Tho17].

[Hab06] Habets, “Room impulse response generator”

[SS15] Savioja and Svensson, “Overview of geometrical room acoustic modeling techniques”

[Tho17] Thomas, “Wayverb: A Graphical Tool for Hybrid Room Acoustics Simulation”

¹³ There are two main categories: geometric and wave-based methods [Hab06; SS15; Tho17].

wave-based aims at solving the wave equation numerically, while

geometric methods make some simplifying assumption about the wave propagation: they typically ignore the *wave* behavior of the sound, choosing much lighter models such as *rays* or *particles*.

► WAVE-BASED METHODS

These are iterative methods that divide the 3D bounded enclosure into a grid of interconnected nodes¹⁴. For instance, the Finite Element Method (FEM)

¹⁴e.g. mechanical unit with simple degrees of freedoms, like mass-spring system or one-sample-delay unit

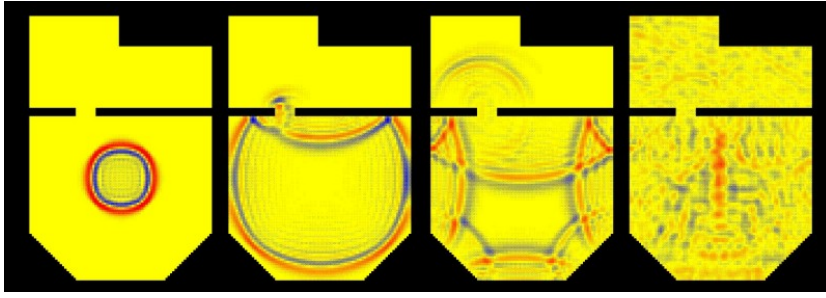


FIGURE 2.12: Simulation of Sound propagation at four consecutive timestamps using the DWM technique. A short, sharp, impulsive sound fired into the larger of two rooms causes a circular wavefront to spread out from the sound source. The wave is reflected from the walls and part of it passes through a gap into the smaller room. In the larger room, interference effects are clearly visible; in the smaller room, the sound wave has spread out into an arc, demonstrating the effects of diffraction. A short while after the initial event, the sound energy has spread out in a much more random and complex fashion.

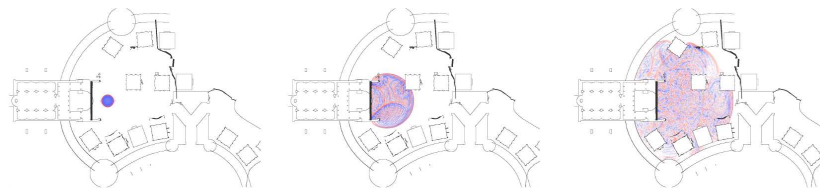


FIGURE 2.13: Sound propagation at three consecutive timestamps using the FDTD-based *Triton* simulator from Microsoft

divide the space into small volume elements smaller of the sound wavelengths, while Boundary Element Method (BEM) divide only the boundaries of the space ~~are divided~~ into surface elements. These nodes interact with each other according to the math of the wave equation. Unfortunately, simulating higher frequencies requires denser interconnection, so the computational complexity increases.

The Finite-Difference-Time-Domain (FDTD) method replace the derivatives with their discrete approximation, i. e. finite differences. The space is divided into a regular grid, where the changes of a quantity (air pressure or velocity) is computed over time at each grid point. Digital Waveguide Mesh (DWM) methods are a subclass of FDTD often used in acoustics problem.

THE MAIN DRAWBACK OF THESE METHODS is discretisation problem: less dense grid may simplify too much the simulation, while denser grid increase the computational load. Moreover, they require delicate definitions of the boundaries condition at the physical lever, like knowing complex impedances, parameters not always available in the literature.

ON THE OTHER HAND these methods inherently account for many effects such as occlusion, reflections, diffusion, diffractions and interferences. In particular by simulating accurately low-frequencies components of the RIR, they are able to well characterize the room modes¹⁵, namely collection of resonances that exist in a room and characterize it.

As stated in [Väl+16], among the wave-based methods, Digital Waveguide Mesh (DWM) are usually preferred: they run directly in the time domain, requiring typically an easier implementation, and they exhibits a natural huge level of parallelism.

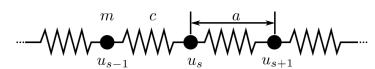


FIGURE 2.11: Example of mass-spring linear mesh used to simulate a 1D transversal wave.

¹⁵ Room modes have the effect of amplifying and attenuating specific frequencies in the RIR, and produce much of the subjective sonic “colour” of a room. Their analysis and synthesis is of vital importance for evaluating acoustic of rooms, such as concert hall, and recording studios or when producing musically pleasing reverbs.

[Väl+16] Välimäki et al., “More than 50 years of artificial reverberation”

For a detailed discussion about geometric acoustic methods, please refer to [SS15].

[Bad19] Badeau, “Common mathematical framework for stochastic reverberation models”

¹⁶such as the amount of reverberation

[Kul85] Kulowski, “Algorithmic representation of the ray tracing technique”

[Sch+07] Schröder et al., “A fast reverberation estimator for virtual environments”

[Hei93] Heinz, “Binaural room simulation based on an image source model with addition of statistical methods to include the diffuse sound scattering of walls and to predict the reverberant tail”

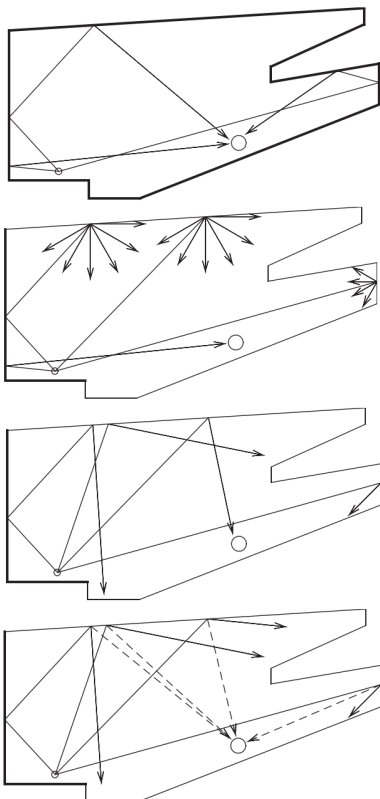


FIGURE 2.14: Visualization of ray-tracing method. From top to bottom: first the method will eventually find specular reflection; then diffuse reflections can be modeled either by

► GEOMETRIC METHODS

They can be grouped into *stochastic* and *deterministic* approaches. They typically compute the reflection path(s) between the source and the receivers, assuming that the wave behaves like a particle or a ray carrying the acoustic energy around the scene.

- **STOCHASTICS** are approximate by nature. They are based on statistical modeling of the RIRs or Monte Carlo simulation methods. The formers writes statistical signal processing models based on prior knowledge, such as probability distribution of the RIR in regions of time-frequency domain [Bad19]. Rather than the detailed room geometry, these methods generally use high-level descriptors¹⁶ to synthesize RIRs and in some application are preferable. The latters randomly and repeatedly subsample the problem space, e. g. tracing the path of random reflections, recording samples which fulfil some correctness criteria, and discarding the rest. By combining the results from multiple samples, the probability of an incorrect result is reduced, and the accuracy is increased. Typically the trade-off between quality and speed of these approaches is based on the number of samples and the qualities of the prior knowledge modeled.

RAY-TRACING [Kul85] is one the most common method that fall in this category and very popular in the field of computer graphic for light simulation. The basic idea is to collect “valid” paths of discrete rays traced around the room. Many technique have been proposed to reduce the computational load, among all the *diffuse rain algorithm* [Sch+07; Hei93] is commonly used in many acoustic simulator. Each ray trajectory is reflected in a random direction every time it hits a wall and its energy is scaled according to the wall absorption. The process of tracing a ray is continued until the ray’s energy falls below a predefined threshold. At each reflection time and for each frequency (bin or band), the ray’s energy and angle of arrival are recorded in histogram, namely a *directional-time-frequency energy map* of the room’s diffuse sound field for a given receiver location (See Figure 2.15) This map is then used as prior distribution for drawing random set of impulses which are used to form the RIR.

While neglecting some detailed description of early reflection and room modes, these methods are good to capture and simulate the statistical behavior of the diffuse sound field for low computational cost.

- **DETERMINISTIC** methods are good to simulate early reflection instead: they accurate traces the exact direction and the timing of the main reflections’ paths.

The most popular is the Allen and Barkley’s Image Source Method (ISM) [AB79]. Even if the basic idea is rather inutile and simple, the model is able to produce the exact solution to the wave equation for a 3D shoebox with rigid walls. Since it models only specular (perfect) reflections, ignoring diffuse and diffracted components. it only approximate arbitrary enclosures and the late diffuse reflections.

The naïve implementation reflects the sound source against all surfaces in the scene, resulting in a set of image sources. Then, each of these image sources is itself reflected against all surfaces. Two are the main limitation of

this method. First, in a shoebox the complexity of the algorithm is cubic in the order of reflection and for order higher 30 the algorithm become impractical. Second it models only the specular reflection, neglecting the diffuse sound field.

For these reasons, the image-source method is generally combined with a stochastic method in hybrid method to model the full impulse response.

- **HYBRID METHODS** As discussed above, the image-source method is accurate for early reflections, but slow and not accurate for longer responses. The ray tracing method is by nature an approximation, but produces acceptable responses for diffuse field. And in general geometric methods fails to proper model lower frequencies and room modes. The waveguide method models physical phenomena better than the geometric methods, but is expensive at high frequencies. All these limitations corresponds into three regions in the Time-Frequency (TF) representation of the RIR. As depicted in Figure 2.19,

- in the time domain, a transition can be identified between the early vs.. late reflection, corresponding to the validity of the deterministic vs.. stochastic models; and
- in the frequency domain, between geometric vs.. wave-based modeling.

By combining three methods, accurate broadband impulse responses can be synthesized, but for a much lower computational cost than would be possible with any individual method. However, this is possible provided that the time- and frequency-domain crossover points are respected and the level of each component is scaled accordingly [Bad19].

THE CROSSOVER POINT in the time domain is called *transition time* or *mixing time*. It identifies the moment after which reflections are so frequent that they form a continuum and, because the sound is partially absorbed by the room surfaces at every reflection, the sound level decays exponentially over time. This point define the cross-fade between the deterministic and the stochastic process¹⁷.

The crossover point in the frequency domain is called *Schroeder's frequency* and it split the spectrum of the RIR into a region with a few isolated modes and one denser, called respectively the *resonant* and *even* behaviors. This point define the cross-fade between the geometrical and wave-based model.

Each simulator available has its own way to compute and implement this crossover points as well as mixing the results of the three methods.

2.3.3 The Method of Images and the Image Source Model

The *Method of Images* is a mathematical tool for solving certain class of differential equations subjected to boundary conditions. By assuming the presence of a “mirrored” source, certain boundary conditions are verified facilitating the solution of the original problem. This methods is widely used in many fields of physics, and interestingly with specific application to Green's functions. Its application to acoustic was originally proposed by Allen and Berkley in [AB79] and it is know as the Image Source Method (ISM). Now ISM is probably the most used technique for deterministic RIR simulation due its conceptual simplicity and its flexibility.

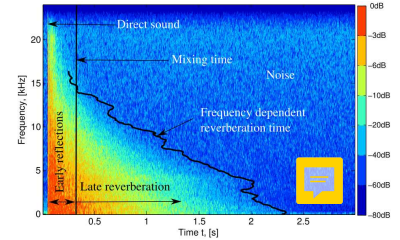


FIGURE 2.17: Time-Frequency profile of a measured RIR.

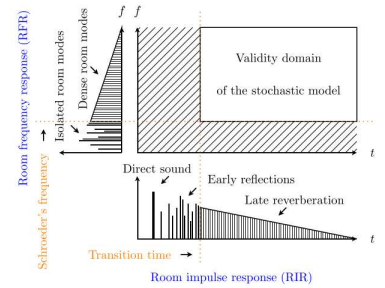


FIGURE 2.18: Schematic of Time-Frequency RIR. Image courtesy of [Bad19].

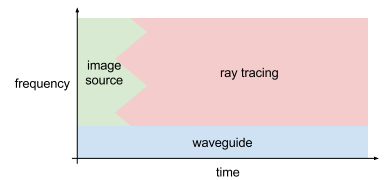


FIGURE 2.19: Time-Frequency regions of RIR associated to the method that better simulate them. Image courtesy of [Tho17].

¹⁷Notice that the stochastic simulator will record both specular and diffuse reflections. Therefore, the mix between the two must be done prudently.

[AB79] Allen and Berkley, “Image method for efficiently simulating small-room acoustics”

The ISM is based on purely specular reflection and it assumes that the sound energy travels around a scene in “rays”.

In the appendix of [AB79], the authors also proved that this method produce a solution the Helmholtz’s equation for rectangular enclosure with rigid boundaries.

- **ON A SINGLE REFLECTOR, THE IMAGE SOURCE** defines the interaction of the propagating sound and the surface. It is based on the observation that when a ray is reflected, it spawns a secondary source “behind” the boundary surface. As show in Figure 2.21, this additional source is located on a line perpendicular to the wall, at the same distance from it as the original source, as if the original source has been “mirrored” in the surface. In this way, the each wavefront that arrives to the receiver from each reflection off the walls as the direct path received from an equivalent (or image) source.

The ISM makes use of the following assumptions:

- sound source and receiver as points in a rectangular cavity
- purely specular reflection paths between a source and a receiver
- This process is simplified by assuming that sound propagates only along straight lines or rays
- Rays are perfectly reflected at boundaries

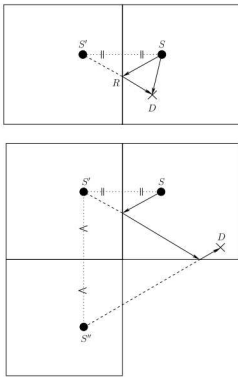


FIGURE 2.20: Path involving one reflection obtained using first-order image (top) and two reflections obtained using two images. It. Image courtesy of [Hab06].

- **FINALLY RIR** is found by summing the contribution from each (image) source, delayed and attenuated appropriately depending on their distance from the receiver. Therefore, in the time domain the RIR associated to the source in position \underline{s} and the receiver in \underline{x} reads

$$\boxed{=} h_{\text{ISM}}^{\text{rir}}(t, \underline{x} | \underline{s}) = \sum_{r=0}^R \frac{1}{4\pi \|\underline{x} - \underline{s}_r\|} \delta\left(t - \frac{\|\underline{x} - \underline{s}_r\|}{c}\right) \quad (2.15)$$

where \underline{s}_r is the r -th image of the source.

The above equation assume perfect rigid and reflective wall. In order to easily incorporate frequency-dependent acoustic impedances (and absorption coefficient) of real surfaces, the Fourier transform of Eq. (2.16) is consider instead, where each reflection term addendum is appropriately scaled

$$H_{\text{ISM}}^{\text{rir}}(f, \underline{x} | \underline{s}) = \sum_{r=0}^R \frac{\alpha_r(f)}{4\pi \|\underline{x} - \underline{s}_r\|} \exp\left(-i2\pi f \frac{\|\underline{x} - \underline{s}_r\|}{c}\right), \quad (2.16)$$

where α_r is the damping coefficient related to the r -th image which in general consider the all the absorption coefficient of the considered surfaces.

2.4 PERCEPTION AND SOME ACOUSTIC PARAMETERS

So far we have analyzed reverberation from a purely mathematical point of view. However in many applications it is important to correlate physical measurements to subjective and perceptual qualities. This will be important in order to define evaluation scenarios later in this thesis.¹⁸

¹⁸ Cite Sacks about perception

2.4.1 The Perception of the RIR's elements

It is commonly accepted that the RIR components defined in § 2.3.1 play rather separate roles in the perception of sound propagation.

- THE DIRECT PATH is the delayed and attenuated version of source signal itself. It coincides with the free-field sound propagation and, as we will see in Chapter 11, it reveals the direction of the source.
- EARLY REFLECTIONS AND ECHOES are reflections which are by nature highly correlated with to the direct sound. They convey a sense of geometry which modify the general perception of the sound:

The Precedence Effect occurs when two correlated sounds are perceived as a single auditory event [Wal+73]. This happens usually when they reach the listener with a delay within 5 ms to 40 ms. However, the perceived spatial location carried by the first-arriving sound is ~~preserved-suppress~~ the perceived location of the lagging sound. This allows human to accurately localize the direction of the main source, even in presence of its strong reflections.

[Wal+73] Wallach et al., “The precedence effect in sound localization (tutorial reprint)”

The Comb Filter Effect indicates the change in timbre of the perceived sound, named *coloration*. This happens when multiples reflections arrive with periodic patterns and some constructive or destructive interferes may arise. Such phenomena can be well modeled with a comb filter [Bar71].

[Bar71] Barron, “The subjective effects of first reflections in concert halls—the need for lateral reflections”

Apparent Source Width is the audible impression of a spatially extended sound source [Gri97]. By the presence of early reflection, the perceived energy increases, providing the impression that a source sounds larger than its optical size.

[Gri97] Griesinger, “The psychoacoustics of apparent source width, spaciousness and envelopment in performance spaces”

Distance and Depth Perception provides to the listener cues about the source location. While the former refers to the spatial range, the latter relates the source to the auditory scene as a whole [Kea+12]. A fundamental cue for distance perception is the *direct-to-reverberant ratio* (DRR)¹⁹, i. e. the ratio between the direct path ration and the remain portion of the RIR. Regarding the depth perception, early reflection are the main responsible. In the context of virtual reality, correct modeling of these quantities is essentials in order to maintain a coherent depth impression [Kea+12].

[Kea+12] Kearney et al., “Distance perception in interactive virtual acoustic environments using first and higher order ambisonic sound fields”

¹⁹See § 2.4.4

- THE LATE REVERBERATION in room acoustics is indicative of the size the environment and the materials within [Väl+16]. It provides the *listener envelopment*, i. e. the degree of immersion in the sound field [Gri97]. This portion of the RIR is mainly characterized by the sound diffusion, which depend on the surfaces roughness.

[Väl+16] Välimäki et al., “More than 50 years of artificial reverberation”

2.4.2 Mixing time

Perceptually, it defines the instant when the reverberation cannot be distinguished from that of any other position of the listener in the room. Analytically, the

mixing time is the instant that divides the early reflections from the late reverberation in a RIR,

And it is represented in Equation 2.47 by the symbol T_m . Due to this, it is an parameters important also in the context of RIRs synthesis as it defines cross-over point for room acoustics simulator using hybrid methods [SS15]²⁰.

²⁰See § 2.3.2

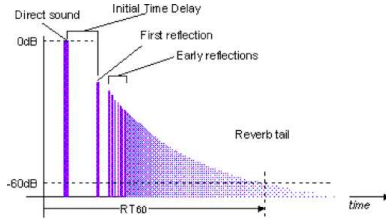



FIGURE 2.21: illustration of the Reverberation Time (RT_{60}) definition. It. Image courtesy of wikipedia.

2.4.3 Reverberation Time

The *reverberation time* measures the time that takes the sound to “fade away” after it ceases. In order to quantify it, ~~acoustics and in audio signal processing~~ use the *Reverberation Time at 60 dB*, i. e.

the RT_{60} , the time after which the sound energy relatively dropped by 60 dB. 

It depends on the size and absorption level of the room (including obstacles), but not on the ~~position of~~ specific position of the source and the receiver. Real measurements of RIRs are affected by noise. As a consequence, it is not always possible to consider a dynamic range of 60 dB, i. e. the energy gap between the direct path and the ground noise level. In this case, the RT_{60} value must be approximated with other methods. A practical approach is presented in Chapter 13.

By knowing the room geometry and the surfaces acoustics profiles, it is possible to use the empirical *Sabine’s equation*:

$$RT_{60} \approx 0.161 \frac{V_{TOT}}{\sum_l \alpha_l S_l} \quad [s], \quad (2.17)$$

where V_{TOT} is the total volume of the room [m^3] and α_l and S_l are the absorption coefficient and the area [m^2] of the l -th surface.

2.4.4 Direct-to-Reverberant ratio and Critical Distance

The direct-to-reverberant ratio (DRR) quantifies the power of direct against indirect sound [Zah02].

[Zah02] Zahorik, “Direct-to-reverberant energy ratio sensitivity”

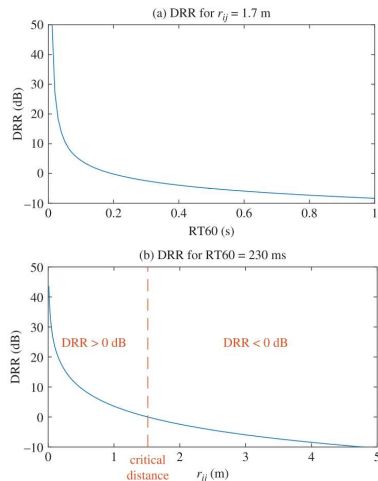


FIGURE 2.22: DRR as a function of the RT_{60} and the source distance r_{ij} based on Eyring’s formula (Gustafsson et al., 2003). These curves assume that there is no obstacle between the source and the microphone, so that

It varies with the size and the absorption of the room, but also with the distance between the source and the receiver according to the curves depicted in Figure 2.22 The distance beyond which the power of indirect sound becomes larger than that of direct sound is called the *critical distance*.

These quantities represent an important parameter to assert the robustness of audio signal processing methods, since they ~~basically~~ measure the validity of the free-field assumption.