Influence of Crossover Frequency on a Hybrid Acoustic Model for Room Impulse Response Synthesis

CHARLOTTE ROUGIER

MSc by Research

ABSTRACT

Room impulse responses (RIRs) of arbitrary enclosures may be synthesized using acoustic modelling methods that all offer advantages and limitations in terms of accuracy and efficiency. The use of a hybrid model, which combines a finite difference time domain (FDTD) approach for the low frequency range, and geometric methods for the mid-high frequency range has been shown to provide appropriate results in the past. This thesis is concerned with the optimisation of such a method by reducing the crossover frequency between the two mixed models to improve its efficiency and cost. To that end, hybrid models are implemented to generate synthesized RIRs with various crossover frequencies. These RIRs are compared with real measurements from two different rooms in the time domain, in the frequency domain, and with the use of acoustic parameters.

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work presented here is my own work. Work done in collaboration with, or with the assistance of, others, is indicated as such.

TABLE OF CONTENTS

INTRODUCTION

The modelling of sound propagation paths within an enclosed space offers an indication of the acoustic properties of non-existent environments, for the purposes of virtual acoustic rendering, or artificial reverberation. Acoustic modelling techniques can further provide an accurate indication of the acoustic properties of non-existent spaces through a process of *auralisation* that consists of reproducing a soundfield via headphones or speakers.

Industrial applications are many and varied. In architecture, the process of auralisation can provide a complete acoustical immersion in an environment prior to its construction, such as in he case of concert hall or other public venue. In the field of virtual reality and video games, it can help to create the illusion of a completely imaginary environment in the eyes and ears of the user.

1.1 Motivation

The acoustic properties of an environment are characterized by its *room impulse response* (RIR) whose accuracy is therefore essential to the quality of the resulting auralisation. This can be generated by measurement or by simulating the sound wave propagation within the enclosure. Over the last 50 years, RIR synthesis techniques have been considerably developed and improved to provide accurate solutions, but full bandwidth RIRs are still computationally demanding to produce, especially for large or complex spaces.

The current established methods for RIR modelling can be divided into two categories : *geometric* methods and *numerical* ones. Geometric approaches, like ray-tracing (RT) or image source methods (ISM), are based on the simplification of sound waves as rays. They present the advan-

tage of not being computationally demanding, but show deficiencies for low frequencies due to the failure to preserve wave characteristics such as diffraction or interaction effects. In recent years, numerical approaches such as the digital waveguide mesh (DWM) or finite difference time domain (FDTD) methods amongst others, have been shown to be more appropriate at low frequencies. Providing a discrete numerical solution to the wave equation, they offer an accurate simulation for the physical wave propagation. However, the computational requirements for such methods can be extremely demanding both in terms of execution time and storage memory for large enclosures.

In order to achieve an optimal balance between computational efficiency and accuracy of the simulation, an alternative approach consists of combining a geometric and a wave-based method for the high and low frequency range respectively to form a hybrid synthesized RIR. A recent implementation of such an acoustic model has provided a valid response using the FDTD method at low frequencies and mixing the beam tracing and acoustic radiance transfer methods for high frequencies [32]. However, such a model is still too computationally intensive to be used for real-time auralisations where model attributes (for example the source and the receiver) are dynamic.

Reducing the execution time and the storage memory of a hybrid acoustic model while maintaining its quality would benefit various areas using real-time auralisation (architecture, virtual reality, video games). Therefore, this thesis will aim to optimize a hybrid model by reducing the crossover frequency between the numerical model and the geometric model that are combined for its implementation.

1.2 Statement of Hypothesis

The limitation of the geometric approach in terms of frequency range is due to wave phenomena that occur for wavelengths that are larger or equal to the size of irregularities on the boundaries of the enclosure. Therefore, an arbitrary choice of a 1kHz crossover frequency for the hybrid model has often been made in the past, stating that the linear dimensions of the walls of the regarded space are generally greater than 34cm, which is equivalent to 1kHz in terms of frequency. If such a crossover frequency usually leads to an efficient and accurate model, it could possibly be reduced on a case-by-case basis depending on the dimensions of the room, in order to improve the computational requirements. Therefore, the hypothesis guiding this thesis can be stated as follows :

"The crossover frequency of a hybrid acoustic model can be reduced to the wavelength equivalent to the smallest distance between two boundaries of the enclosure, without compromising the quality of the model."

To that end, hybrid acoustic room impulse responses will be synthesized in the case of two rooms, with a crossover frequency varying from 20 Hz to 1200 Hz, and compared with real measurements. Results will be analysed in the time domain, in the frequency domain and with the use of acoustic parameters.

1.3 Thesis Structure

This thesis is organized in the following way :

Chapter 2 provides a general background of room acoustics by detailing basic principles of sound propagation in an enclosed space. It also describes how a room can be characterized by its impulse response (RIR) and how this one can be measured.

Chapter 3 discusses the auralisation process which consists of reproducing a soundfield via headphones or loudspeakers. A virtual environment can be simulated by synthesizing its RIR which can be achieved via virtual acoustic modelling methods. These methods, which can be categorized as numerical, geometric or hybrid, are presented and detailed to determine which one is the most adapted to this research.

Chapter 4 describes the implementation of the acoustic model, which is a hybrid approach combining finite difference time domain methods for the low frequency range and ray-tracing / image source methods for the mid-high frequency range. In the context of this research, 2 rooms are modelled, at 3 different source / receiver positions. For each situation, 60 RIRs are synthesized with a crossover frequency ranging from 20 Hz to 1200 Hz.

Chapter 5 details the results provided by the comparison of each synthesized RIR and real measurements. The RIRs are analysed in the time domain, in the frequency domain, and by the use of acoustic parameters like T_{30}, early decay time and C_{80}.

Chapter 6 provides a summary of the main findings in accordance with the hypothesis, and raises new lines of inquiry for further research.

Elements of Room Acoustic Theory

Sound is the consequence of vibrations in a compressible fluid. These vibrations are generated by a source causing a force toward the molecules of the considered media, leading to pressure fluctuations. This results in a sequence of waves of pressure that travel, reaching on its way the human ear which is very sensitive to pressure variations, interpreted by what is commonly called *sound* by the human brain. According to Kuttruff [17], "any complex sound field can be considered as a superposition of numerous simple sound waves, e.g. plane waves".

This chapter will review some elements of acoustic theory related to sound propagation in an enclosed space. In section 2.1, the sound field will be considered as unbounded in every direction to describe sound wave propagation in a general environment. The impact of room boundaries will be detailed in section 2.2 to provide an accurate description of the behaviour of sound waves in an enclosed environment, as it is the case in room acoustics. Such an environment can be characterized by its impulse response from which acoustic parameters are derived, as presented in section 2.3.

2.1 Sound Propagation in air

In this section, the medium for propagation is considered as unbounded in every direction. It is also assumed that the medium is homogeneous, at rest, and free of losses. In this case, the velocity of sound is constant. Therefore, the speed of sound in air can be measured as :

$$c = 331.3\sqrt{1 + \frac{\theta}{273.15}} \text{ m/s} \simeq 331.3 + 0.606\theta \text{ m/s} \tag{2.1}$$

where θ is the temperature in degrees centigrade. The second equation is a Taylor approximation of the first one.

2.1.1 The Wave Equation

As stated previously, a sound wave is the result of vibrations of particles present in a medium. The wave can therefore be described fully by the movements of these particles, thus by their velocity.

According to [17], wave propagation can be modelled using the equation for conservation of momentum (2.2), the equation for conservation of mass (2.3), and the adiabatic equation of state (2.4), respectively defined as:

$$\text{grad } p = -\rho_0 \frac{\partial \mathbf{v}}{\partial t} \tag{2.2}$$

$$\rho_0 \text{ div } \mathbf{v} = -\frac{\partial \rho}{\partial t} \tag{2.3}$$

$$\frac{p}{p_0} = \kappa \frac{\delta \rho}{\rho_0} \tag{2.4}$$

where p denotes sound pressure, $\mathbf{v}$ the vector particle velocity, $\rho = \rho_0 + \delta\rho$ the medium density, t the time, κ the medium adiabatic exponent, p_0 and ρ_0 the static value of the medium pressure and density.

The *wave equation* that describes the propagation of waves in a lossless fluid is obtained combining these three equations, as :

$$c^2 \Delta p = \frac{\partial^2 p}{\partial t^2} \tag{2.5}$$

where $c^2 = \kappa \frac{p_0}{\rho_0}$ and Δ is the Laplace operator defined as $\Delta p = \frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2}$ in Cartesian coordinates.

2.1.2 Plane and Spherical Acoustic Waves

A sound wave is caused by an excitation that can be usually modelled as a *point source* which tends to radiate sound equally in each direction. In this context, wave propagation can be analysed in a single direction (the x axis for example) and equation (2.5) can be rewritten as :

$$c^2 \frac{\partial^2 p}{\partial x^2} = \frac{\partial^2 p}{\partial t^2} \tag{2.6}$$

A solution to (2.6) can therefore be of the form :
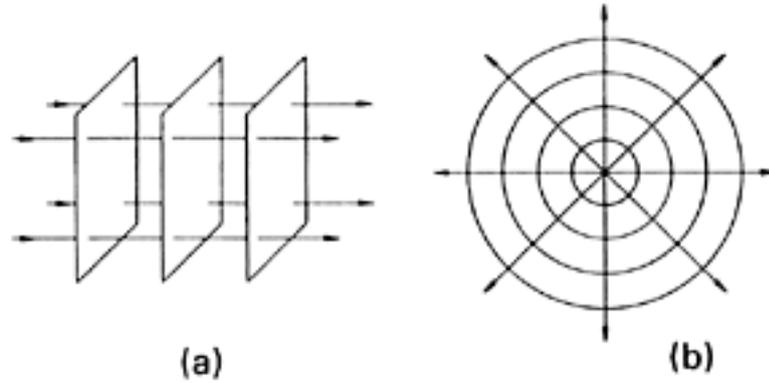
Figure 2.1: (a) Plane wave; (b) Spherical wave. Image sourced directly from [17].

$$p(x,t) = p_0 \exp[i(\omega t - kx)] \tag{2.7}$$

In this equation, k and ω are respectively the wave number and the angular frequency, that rescale the periods of the wave in time and space to 2π. The period in space is the wavelength λ obtained in (2.8), and the temporal period T is obtained in (2.9). These two periods are related by the equation in (2.10), where f is the wave frequency.

$$\lambda = \frac{2\pi}{k} \qquad (2.8) \qquad\qquad T = \frac{2\pi}{\omega} \qquad (2.9) \qquad\qquad c = \frac{\omega}{k} = \frac{\lambda}{T} = \lambda f \qquad (2.10)$$

This type of wave is called a *plane wave*. As shown in Figure 2.1(a) and according to Kuttruff [17], "the sound pressure is constant in any plane perpendicular to the x-axis". These planes are called *wavefronts*, and lines perpendicular to them are called *wave normals*. The wave is said to be *harmonic* if it is represented by an exponential with imaginary arguments. Although it is not possible in practice to have a true plane wave, waves approximate plane wave behaviour at a sufficient distance from the source.

Using (2.2), it is possible to directly determine the particle velocity v from the pressure p via :

$$v = \frac{p}{\rho_0 c} \tag{2.11}$$

This product $\rho_0 c$ is called the *characteristic impedance* of the medium. For air, $\rho_0 c = 416$ Pa.s.m^{-1}.

In practice, the wave fronts are concentric spheres (or circles in 2D) instead of planes, as in Figure 2.1(b). Using spherical coordinates, and taking into account only the radial part due to the symmetries of the situation, the wave equation given in (2.5) can be rewritten as :

7

$$\frac{\partial^2 p}{\partial r^2} + \frac{2}{r}\frac{\partial p}{\partial r} = \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} \tag{2.12}$$

A solution of (2.12) will therefore be of the form (considering the source at the origin of the system, and the wave travelling in the x-direction) :

$$p(x,t) = \frac{p_0}{r}\exp[i(\omega t - kx)] \tag{2.13}$$

At a sufficient distance from the source where r is supposed constant, the spherical wave can therefore be approximated as a plane wave.

2.1.3 Medium Absorption

In the case of a medium with losses, which occurs in reality, the pressure amplitude decreases during the propagation of the wave, following the exponential law, m being defined as the constant attenuation coefficient. Equation (2.7) can therefore be modified as:

$$p(x,t) = p_0\exp[-\frac{mx}{2}]\exp[i(\omega t - kx)] \tag{2.14}$$

Therefore, the wave number k will also be modified as :

$$k = \frac{\omega}{c} - i\frac{m}{2} \tag{2.15}$$

A study conducted by Harris [10] showed that the attenuation coefficient of air was dependant on temperature, humidity and frequency. Thus, for typical values of humidity and temperature, the attenuation coefficient will be very low for frequencies up to 2 kHz and can be neglected. However, for the range of frequencies above 2 kHz, it has to be taken into account when modelling wave propagation.

2.1.4 Sound Intensity

In generating a sound wave, a sound source introduces a quantity of energy to a fluid, which will be carried away by the sound wave [17]. The rate of this energy transfer is described by the instantaneous sound intensity $I(t)$ that corresponds to the flow of energy through a unit area of one square meter at a certain instant of time. However, it is more usual to use the sound intensity I, that is, the time average of $I(t)$:

$$I = \langle I(t)\rangle_T = \langle pv\rangle_T = \frac{1}{T}\int_T pv\,dt \tag{2.16}$$

In general, sound intensity is measured on the scale of *Sound Intensity Level* (SIL) defined as:

$$SIL = 10 \log_{10} \left(\frac{I}{I_{\text{ref}}} \right) \qquad (2.17)$$

where I_{ref} is the reference intensity level at 10^{-12} W.m^{-2} which corresponds to the lower limit of human perception.

For a plane wave, the sound pressure and the particle velocity are related by the equation given in (2.11). Therefore, the intensity reduces to :

$$I = \frac{p^2}{\rho_0 c} \qquad (2.18)$$

As explained previously, the sound intensity corresponds to the flow of energy through a unit area. Therefore, as r increases, the intensity level decreases proportionally to A. In other words, for a given position r, the intensity is :

$$I = \frac{P_s}{4 \pi r^2} \qquad (2.19)$$

where P_s is the power of the source.

For a spherical wavefront, (2.19) provides an indication of the spread of sound energy called the *inverse square law*. Indeed, as the spherical wave front travels away from its source, the area of the spreading wave front is $A = 4 \pi r^2$.

2.1.5 Sound Pressure Level

In room acoustics, the quantity used when describing the amplitude of a sound is usually the amplitude of the pressure or the associated velocity component of the sound wave [11]. Indeed, the human ear is sensitive to pressure and furthermore it is easier to measure than sound intensity.

A unit of pressure (in Pascal) corresponds to the force of 1 Newton applied on a 1 square meter surface. The range of hearing sounds vary from 20 µPa which represents the threshold of hearing to 20 Pa which represents the threshold of pain. Due to the way the ear perceives sound, acoustic pressure is more generally based on a logarithmic scale, commonly called *Sound Pressure Level* (SPL), which is defined as:

$$SPL = 20 \log_{10} \left(\frac{p}{p_{\text{ref}}} \right) \qquad (2.20)$$

where p_{ref} is the reference pressure level at 20 µPa.

2.2 Sound Propagation in an Enclosed Field

Up to this point, the soundfield was considered as unbounded in every direction. However, in the case of room acoustics, sound propagates in an enclosure in which the medium is bounded by walls, floor and ceiling. As the wave encounters this kind of boundary, it is reflected, leading the soundfield to become more complex.

The change of media properties through a space have multiple effects on the propagating sound, leading to phenomena like reflection and absorption, scattering, or diffraction. In such cases, waves are forced to interact with each other, causing *interference*.

2.2.1 Reflection and Absorption

When a plane wave encounters the boundary of a second medium (a wall for example), the sound energy is divided into two parts: the reflected wave and the absorbed wave.

The complex reflection factor R depends on the properties of the wall and defines the change in amplitude $|R|$ and phase χ happening during the reflection :

$$R = |R|\exp(i\chi) \tag{2.21}$$

As expressed in section 2.1.4, the intensity of a sound wave is proportional to the square of the pressure amplitude (2.18). Therefore, the intensity of the reflected wave will be reduced by R^2, setting the loss of energy measurement to $1 - R^2$. This quantity corresponds to the absorption coefficient α, ranging from 0 for a totally reflective surface to 1 for a fully absorbing surface.

$$\alpha = 1 - R^2 \tag{2.22}$$

The introduction of a new parameter called *acoustic impedance* is necessary to describe the behaviour of waves towards a surface. This parameter is defined by the ratio of the normal component of the particle velocity to the boundary, and the pressure at its surface (2.23) and describes the quantity of sound pressure generated at a certain frequency.

$$Z = \frac{p}{v_n} \tag{2.23}$$

It can also be useful to define the *specific acoustic impedance ζ* that is a property of the medium alone, defined by the ratio of the the acoustic impedance and the *characteristic impedance* of air Z_0.

$$\zeta = \frac{Z}{Z_0} \tag{2.24}$$
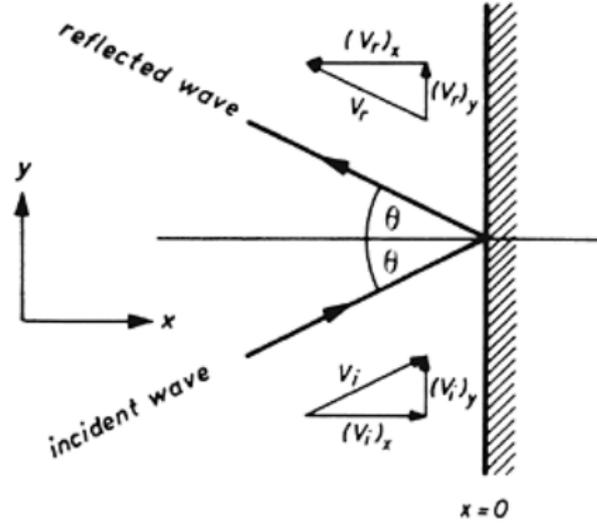
Figure 2.2: Sound reflection at oblique incidence. Image sourced directly from [17].

$$Z_0 = \rho_0 c = 414 \text{ kg m}^{-2} \text{ s}^{-1} \tag{2.25}$$

If the boundary is normal to the direction in which the incident wave is travelling, the reflective wave will be travelling in the opposite direction, with a change of amplitude and phase. However, in a more general case, the angle of incidence θ_i as illustrated in Figure 2.2, varies between 0 and 90° and the angle of reflection θ_r that defines the direction of the reflected wave is determined according to Snell's law :

$$\theta = \theta_i = \theta_r \tag{2.26}$$

In the case of a boundary b and for a normal reflection, the impedance coefficient of the boundary Z_b can be determined in terms of the reflection coefficient as :

$$Z_b = \rho c \frac{1+R}{1-R} \tag{2.27}$$

It is therefore possible to calculate the reflection coefficient based on the specific impedance of the boundary ζ_b dependant on the angle of incidence θ:

$$R = \frac{\zeta_b \cos(\theta) - 1}{\zeta_b \cos(\theta) + 1} \tag{2.28}$$

2.2.2 Scattering

In most real environments, surfaces hit by a sound wave are not completely smooth, and might contain irregularities. If these are smaller than the wavelength, the reflection is not disturbed and occurs as described in 2.2.1. However, if the wavelength is shorter than the wall irregularities (i.e.
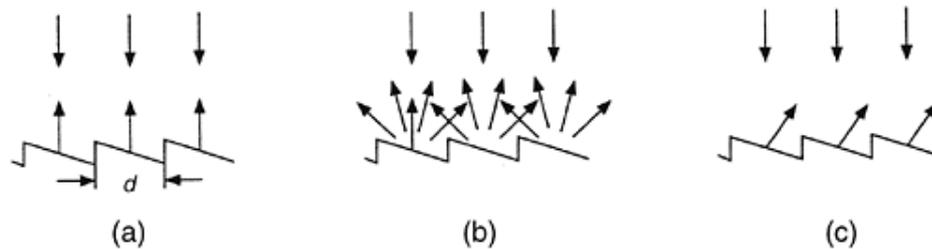
Figure 2.3: Scattering by wall irregularities : (a) $d \ll \lambda$; (b) $d \approx \lambda$; (c) $d \gg \lambda$. Image sourced directly from [17].

high frequencies), the wave will be reflected by the irregularities, changing the angle of reflection, as in Figure 2.3. There is also a particular case for an intermediate range of wavelengths equal to the surface irregularities that have the potential to be greatly diffused. Scattering can also occur for walls with a non-uniform impedance.

2.2.3 Diffraction

According to [37], diffraction occurs when a sound wave encounters free edges, corners or edges of a room, or boundaries between mediums with different impedances. For a certain range of wavelengths, the direction of the emerging wave can be deviated, which corresponds to the phenomenon of diffraction. This is due to Huygen's principle stating that a wave front can be modelled as a series of wavelets that interfere with each other. As they encounter an obstacle, some of them are slowed down, causing destructive interferences leading to a diffraction pattern.

In a real case, diffraction can occur in the case of an aperture between two boundaries as in Figure 2.4, or in the presence of an obstacle between the source and a receiver. It concerns mostly those wavelengths approximately the size of the obstacle. Therefore, diffraction is mostly prevalent for low frequencies in room acoustics.

2.2.4 Interference

When a sound wave is reflected, the pressure and velocity component of the reflected wave is added with those of the incident wave. In this case, considering that these two waves have the same frequency, the amplitude of the pressure of the resulting wave can vary from zero to the sum of the pressure amplitudes of the waves [11]. This is called *interference* and can be categorised as constructive or destructive.

Constructive interference is created if the waves are in phase, with the consequence that the resultant wave is bigger than either of the two originals (see Figure 2.5 (top)). On the contrary,
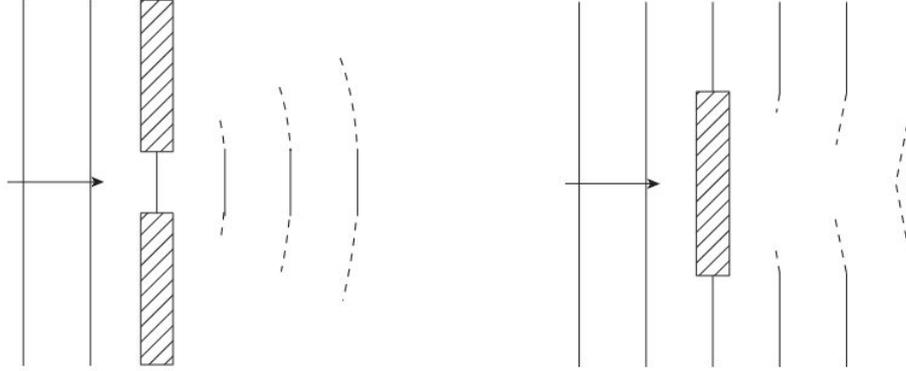
Figure 2.4: Diffraction through an opening (left) and an object (right). Image sourced directly from [11]

.

destructive interference happens when the waves are out of phase (see Figure 2.5 (bottom)), and the resultant wave will have pressure lower than either original wave.

2.2.5 Standing Waves and Resonances

The soundfield in a cavity or enclosure is much more complex than in the free field, due to the reflection of waves from the surrounding boundaries. The superposition of the reflected waves create interference, due to the addition of their amplitudes, as explained in 2.2.4. This can lead to the phenomenon of *standing waves* (see Figure 2.6). These occur at certain frequencies for which the distance between the extremities of the room equals a direct multiple of the wavelength. These frequencies are called *resonances* and can be determined from the wave equation. According to [37], considering the boundary conditions (2.28), the wave equation can be transformed into the Helmholtz equation (stationary case) :

$$\Delta p + k^2 p = -i\omega\rho_0 q(r) \tag{2.29}$$

where q is the harmonic volume source excitation. The boundary conditions can be expressed as:

$$Z_b \frac{\partial p}{\partial n} + i\omega\rho_0 p = 0 \tag{2.30}$$

The solutions of equation (2.29) are a discrete series of eigenvalues $k_{l,m,n}$ related to discrete eigen frequencies $f_{l,m,n}$ that are the resonant frequencies [37] :

$$f_{l,m,n} = \frac{c}{2\pi} k_{l,m,n} \tag{2.31}$$
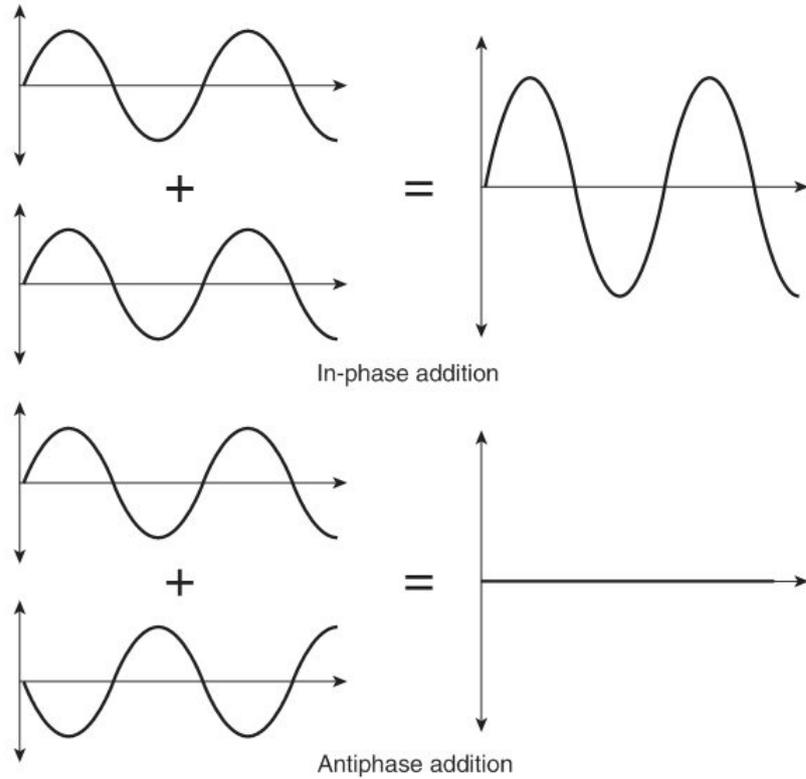
Figure 2.5: Constructive Interference (top) and Destructive Interference (bottom). Image sourced directly from [17].

For example, let's consider a room of dimensions L_x, L_y, L_z (as in Figure 2.7), with fully reflective boundaries ($Z_b = \infty$). The boundary conditions in (2.30) can therefore be rewritten for the x component as :

$$\frac{dp_x}{dx} = 0 \text{ for } x = 0 \text{ and } x = L_x \tag{2.32}$$

Without considering the source, and taking into account the boundary conditions (2.32), the Helmholtz equation (2.29) leads to a solution of the form:

$$p_x(x) = A \cos(k_x x) \tag{2.33}$$

with $k_x = \frac{l\pi}{L_x}$, where A is an arbitrary constant and l a non-negative integer. Similar results are obtained regarding the y and z components.

It can be shown that for a *3-D* situation, the wave number k and the pressure p can be written as:
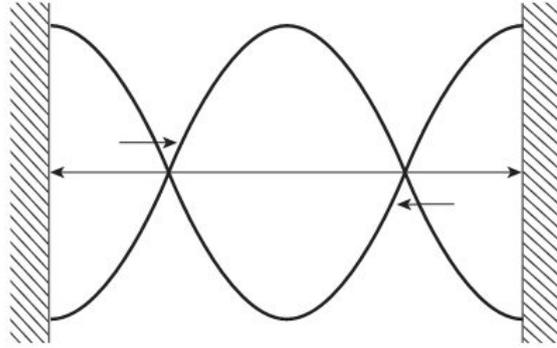
Figure 2.6: The pressure components of a standing wave between two hard boundaries. Image sourced directly from [11]
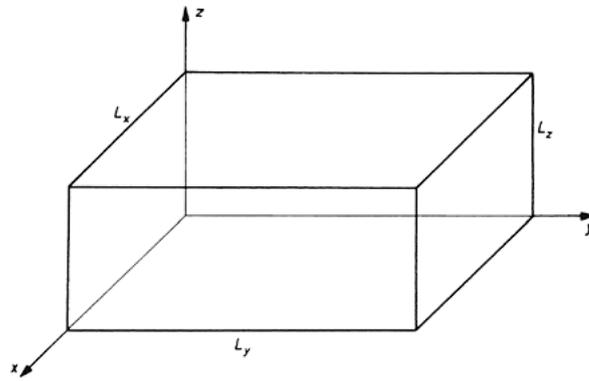
Figure 2.7: Rectangular room. Image sourced directly from [17]

$$p = p_x.p_y.p_z \tag{2.34}$$

$$k^2 = k_x^2 + k_y^2 + k_z^2 \tag{2.35}$$

where p_x and k_x are respectively the pressure and wave number associated to the resulting wave on the x-axis (idem for y and z).

Therefore, combining these two equations with (2.31), it is possible to rewrite the resonant frequencies as :

$$f_{l,m,n} = \frac{c}{2}\sqrt{\left(\frac{l}{L_x}\right)^2 + \left(\frac{m}{L_y}\right)^2 + \left(\frac{n}{L_z}\right)^2} \tag{2.36}$$

The eigenfunctions corresponding to those eigenvalues (resonant frequencies) can also be written as:
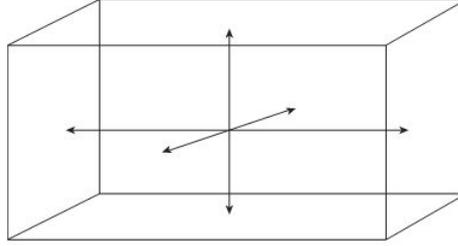
Figure 2.8: Axial mode paths in a room. Image sourced directly from [11].

$$p_{l,m,n}(x,y,z) = A \cos(\frac{l\pi x}{L_x})\cos(\frac{m\pi y}{L_y})\cos(\frac{n\pi z}{L_z}) \qquad (2.37)$$

2.2.6 Room Modes

According to [11], the standing waves have the following properties:

- Only a subset of the surfaces are involved with their reflection

- The reflection occurs at a particular angle of incidence

- A cyclical path is necessary for a coherent return of energy

Regarding these properties, standing waves can be categorised into three basic types called room modes, depending on the number of surfaces involved.

Axial modes : These modes result from a cyclic wave propagation in a single dimension, involving two opposing surfaces (Figure 2.8). The frequencies of an axial mode on the x-axis are given by :

$$f_l = \frac{c}{2}\frac{l}{L_x} \qquad (2.38)$$

Tangential modes : These modes result from a cyclic wave propagation in two dimensions, between four surfaces (Figure 2.9). The frequencies of an tangential mode on the x-axis and y-axis are given by :

$$f_{l,m} = \frac{c}{2}\sqrt{\left(\frac{l}{L_x}\right)^2 + \left(\frac{m}{L_y}\right)^2} \qquad (2.39)$$

Oblique modes : These modes result from a cyclic wave propagation in three dimensions, between six surfaces (Figure 2.11). The frequencies of an oblique mode are given by:

16
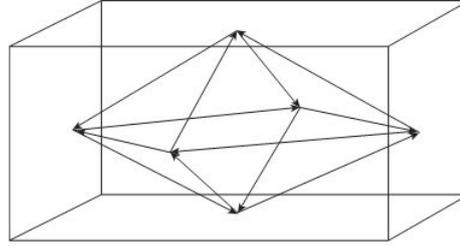
Figure 2.9: Tangential mode paths in a room. Image sourced directly from [11].
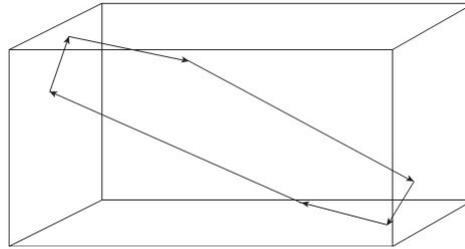
Figure 2.10: Oblique mode paths in a room. Image sourced directly from [11].

$$f_{l,m,n} = \frac{c}{2} \sqrt{\left(\frac{l}{L_x}\right)^2 + \left(\frac{m}{L_y}\right)^2 + \left(\frac{n}{L_z}\right)^2} \qquad (2.40)$$

2.2.7 Frequency Regions

In Room Acoustics, the audio spectrum can be divided up into four regions, each characterized by a specific acoustic property [11].
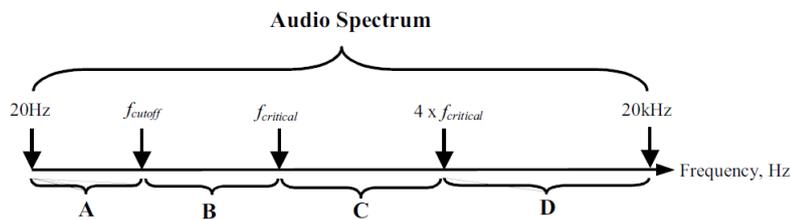
Figure 2.11: Division of the audio spectrum into four regions. Image sourced directly from [23].

Region A is defined as the frequency region under the lowest room axial mode. The cut-off frequency is calculated from equation (2.38) as :

$$f_{cutoff} = \frac{c}{2L} \qquad (2.41)$$

17

where L is the longest distance between two boundaries of the room, and c the velocity of sound. In this frequency region, called the *room cut-off region*, sound propagates in the room and is not amplified by any resonance phenomenon.

In Region B, the sound is fully subject to the behaviour of the room modes defined in Section 2.2.6. However, it is no longer the case in Region D, where the diffuse field predominates over the room modes. In this region, wavelengths are small enough to use geometric acoustic methods (see Chapter 4).

The boundary between regions B and D is not precise and can be defined as region C, which corresponds to a range of frequency from $f_{critical}$ (2.42) to $4f_{critical}$. In this region, wavelengths are too small to use a wave-based approach due to dispersion (detailed in Section 4.4.4) and too long for a ray approximation (see Chapter 4).

$$f_{critical} = \frac{3c}{d} \tag{2.42}$$

where d is the smallest distance between two boundaries of the room.

The critical frequency is also known as the *Schroeder frequency* and can be calculated in terms of the reverberation time RT_{60} (that will be defined and detailed in the next section) and the volume of the room (in m^3) [11] :

$$f_{Schroeder} = 2102\sqrt{\frac{RT_{60}}{V}} \tag{2.43}$$

2.3 The Room Impulse Response

2.3.1 Characterization of Room Acoustics

A room can be considered as a system where the input is a sound excitation at a specific position, resulting in an output at another position. Therefore a room can be characterized by its *impulse response* (RIR) subject to the source and receiver locations which represents the output when presented with an impulsive excitation. This impulse can be modelled as the Dirac delta function (or Kronecker delta for discrete time systems) defined as:

$$\delta(t) = \begin{cases} 1, & \text{if } t = 0 \\ 0, & \text{if } t \neq 0 \end{cases} \tag{2.44}$$

If the system is *linear time invariant* (LTI), the output signal can be obtained through the convolution of the input sound signal with the room impulse response. The input signal will have to be previously recorded in an anechoic chamber in order to preserve it from further reflection effects. If the room does not exist in reality, different techniques can provide a model of the impulse response (that will be reviewed in Chapter 3) and still need to be improved. On the other hand, if the room is an existing one, some methods for measuring the impulse response are very efficient, and will be described in the next subsection.

2.3.2 RIR Measurements

According to [33], the common way of measuring the impulse response of an acoustic system is to "apply a known input signal and to measure the system output". Some requirements upon the equipment and the signal source are necessary for a good measurement : the sound source should be near omni-directional, and the receiving microphone should be small enough to not interfere with the local soundfield.

The input signal should ideally be an impulsive signal of infinitely short duration, infinitely high power and unit energy, in order to cover the whole spectrum. Such a signal is not generatable (due to a hypothetical infinite bandwidth), but there are other types of signal sources with a flat frequency response over a certain bandwidth that can be used. One of them is the Maximum Length Sequence (MLS) signal that is a periodic pseudo-random white noise signal with a flat frequency spectrum over a period of time. However, some distortion peaks can be present due to the non-linearities of the system, and a tedious calibration has to be carried out [33]. Another way to approximate the impulse response is to emit a swept sine, with a frequency range as desired. This technique has been developed by Muller and Massarani [19] and by Farina [8], and allows the separation of the linear and non-linear responses, after a deconvolution with the inverse input signal. Comparing the two mentioned methods [33], it seems like the MLS technique is more appropriate for occupied rooms, or for exterior environments. However, the swept-sine technique is more convenient for empty rooms, and will therefore be applied for live recordings in the future of this work.

The positions of the source and the receiver need to be taken into account. Indeed, they should reflect real positions, and not be located too close to each other. For example, if the measurement is for a concert hall, the source should be typically placed on stage, and the receiver in the audience area.

2.3.3 Parameters derived from Impulse Responses

Apart from providing an accurate representation of the soundfield, due to a certain set of conditions, the room impulse response is also useful to determine some acoustic parameters of the
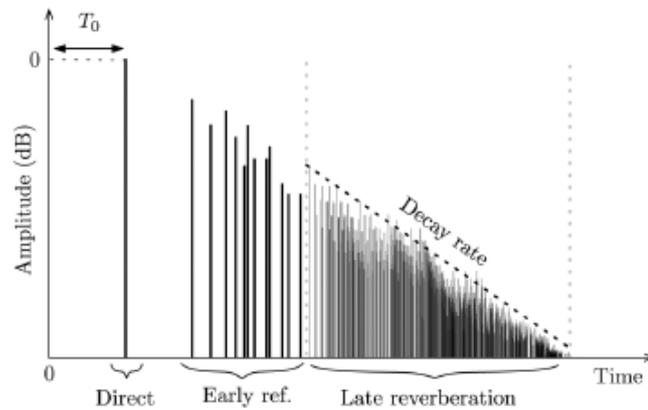
Figure 2.12: Schematic Example of a Generic Impulse Response. Image sourced directly from [34]

.

room, as described in this subsection.

The impulse response indicates the amplitude of pressure at the position of the receiver as a function of time. As an impulsive sound is emitted, the sound propagates into the room until it reaches the receiver (microphone). However, some paths will take more time than others due to reflection or other boundary effects (as mentioned in 2.2.1), resulting in time delay and energy loss which is represented in the impulse response. As can be seen in Figure 2.12, the impulse response can be decomposed into three parts : the *direct sound* which is the first impulse and corresponds to the sound from the source to the receiver without any reflection; the *early reflections* that are delayed due to a longer path of propagation and first reflection orders; the *late reverberation* that are low intensity reflections arriving at the receiver after many varied boundary interactions.

Reverberation times

According to ISO 3382 [3], the *reverberation time* indicates how fast reflection energy decays, and is one of the main indicators as to the acoustic characteristics of an enclosed space. It corresponds to the amount of time taken for the level of the sound pressure to drop by 60dB.

Reverberation time can also be obtained without extracting this measure from the impulse response and is a good way to determine the accuracy of a synthesized or measured impulse response. The Sabine reverberation formula [13] for RT_{60} is given by :

$$RT_{60} = \frac{0.161V}{S\alpha} \qquad (2.45)$$

where S is the total surface area in the room, V its volume, and α the average Sabine absorptivity.

However, equation (2.45) is only accurate in the case of a room where the distance between two surfaces of reflection is approximately the same (such as a cubic room for example), with diffuse soundfields (where the sound pressure level is uniform), and with low absorbing coefficients on the surfaces. Another method of predicting reverberation time is provided by the Norris-Eyring equation [17] :

$$RT_{60} = \frac{-0.161V}{S \ln(1-\alpha)} \tag{2.46}$$

This is valid for the same situations mentioned previously and with more accurate results for high absorption values.

Acoustic parameters can also be derived from RIR measurements. In this context, RT_{60} can be derived from the *energy decay curve* (or Schroeder curve), obtained from the RIR by the following equation :

$$EDC(t) = \int_t^\infty p^2(\tau)d\tau = \int_0^\infty p^2(\tau)d\tau - \int_0^t p^2(\tau)d\tau \tag{2.47}$$

The reverberation time can therefore be obtained directly from the values of EDC converted on a logarithmic scale. However, in practice, RT_{60} cannot be directly extracted from the Schroeder curve due to the presence of prominent noise. Therefore T_{30} is used instead, which corresponds to the amount of time for sound energy to drop from 5 dB to 35 dB and is measured with a linear fit. RT_{60} can be subsequently estimated using the same linear regression as T_{30}.

Another way to measure reverberance is to calculate the *early decay time* (EDT) which is an indicator of the decay between 0 and 10 dB. Compared to the reverberation time, it takes into account the direct field and the very early reflections, as the reverberation time is more an indicator of the late reflective energy. Kuttruff [17] affirms that, "listening tests based on binaural impulse responses recorded in different concert halls confirmed that the perceived reverberance is closely related to EDT".

Energy Ratios

The Early-to-Late index or *clarity* is the ratio of energy in early reflections to that in the late reverberations, expressed in decibels, and describes the perceived definition of the sound. It is given by the following formula :

$$C_{t_e} = 10 \log_{10} \left(\frac{\int_0^{t_e} p^2(t)dt}{\int_{t_e}^{\infty} p^2(t)dt} \right) \text{ dB} \qquad (2.48)$$

where t_e is the temporal limit between early and late RIR components, often corresponding to 50 or 80 ms.

2.4 Summary

This chapter has provided a general background about room acoustic notions that are essential for the next chapters. Sound propagation has been presented in unbounded environments to introduce general concepts about the way in which waves travel, based on the wave equation. Wave behaviour toward boundaries has also been discussed to provide an overall picture of sound propagation in an enclosed field, like it is the case in room acoustics. Finally, it has been mentioned that each room can be characterized by its impulse response, from which acoustic parameters about can be derived, providing a global evaluation of the acoustic quality of the space.

Room impulse responses (RIRs) can be synthesized if they cannot be measured, via numerous techniques that still need to be improved in terms of accuracy and efficiency. Chapter 3 will provide a general overview of these approaches, followed by the implementation and optimization of a hybrid acoustic model.

VIRTUAL ACOUSTIC MODELLING METHODS

The acoustic properties of a room (existent or non-existent) can be provided by the analysis of its impulse response. It is therefore essential to synthesize it as accurately as possible, and computationally efficiently in the case of real-time auralisation that will be defined in section 3.1. The approaches to virtual acoustic modelling are many and varied, with various strengths and weaknesses in each case, and can be categorized as geometric (see section 3.2) or wave-based (see section 3.3). Numerical techniques usually provide more accurate results but are computationally very expensive.

3.1   Auralisation

The room impulse response contains the most objective acoustic parameters of the room. However, a listening experience is necessary in order to simulate an accurate impression of the environment. This process is called auralisation and is defined as, "the technique for creating audible sound files from numerical (simulated, measured, synthesized) data", according to Vorländer [37].

3.1.1   Artificial Reverberation

The process of reverberation consists of the reproduction of a sound environment applied to an input audio signal. Three categories of reverberation algorithms are commonly used [34] :

- **Feedback Delay Networks :** The input signal is delayed, filtered and returned to build the output based on reverberation parameters.

- **Convolution Algorithms:** The output signal is obtained by a convolution of the captured or estimated impulse response with the dry input signal.

- **Computational Acoustics :** Acoustic energy propagation is simulated for a specified geometry.

- **Virtual Analog Models :** Reverberation effects are simulated using electromechanical or electric devices (tape delays, spring reverberation devices, reverberation plates)

Given the computational resources of today, the most suitable process for a room with a simple geometry and without any external interactions would be convolution-based techniques, that simply are the equivalence of FIR filtering in the time domain. Using discrete time linear convolution, the sound signal perceived in a specific sound environment y characterized by its impulse response h will be derived from the dry input signal x via:

$$y[n] = (x * h)[n] = \sum_{m=-\infty}^{\infty} x[n]h[n-m] \tag{3.1}$$

The main drawback of these methods is the computational load that can rapidly become significant. According to [34], fast convolution techniques (using Fast Fourier Transform $\mathscr{F}$ and filtering in the frequency domain) are the most effective, especially for reducing the latency for real-time auralisation. Therefore, the linear convolution described in (3.1) will be carried out in the frequency domain as :

$$y = \mathscr{F}^{-1}\left\{\mathscr{F}\{x\} \cdot \mathscr{F}\{h\}\right\} \tag{3.2}$$

3.1.2   Spatial Sound Reproduction

The outcome of the auralisation process is to enable the subject to experience a complete listening immersion in a room in which they are not currently present. Therefore, a procedure of spatial encoding is necessary to capture the sound field and reproduce it in a suitable way for auralisation. The room can be a real one, where measurements have been made, or a virtual one, for which the whole process will be performed by simulation. The immersion can be created via two categories of approaches : Fully Computed Auralisation and Computed Multiple-Loudspeaker Auralisation [37].

In the first case, the subject will experience the immersion via a pair of headphones. This process begins with the measurement of the *Head Related Transfer Function* (HRTF) which is specific to each listener's anatomy. A HRTF is derived from the combination of an impulse response recorded (or synthesized) at the position corresponding to the center of the subject's

head, and a *Binaural Impulse Response* (BRIR) provided (via measurement or simulation) at the exact position of the left and right ears of the subject. HRTFs are then processed with an audio signal that has been previously recorded or via a real-time input, which effectively becomes the source.

The computed multiple-speaker auralisation technique employs a large number of loudspeakers to physically reproduce the desired sound field. The previous approach can only be applied on a single listener, and this one is a multi-listener solution. This process is based on the simple source approach which states that, "the acoustic field generated by events outside a volume can also be generated by a continuous distribution of secondary simple sources enclosing the respective volume" [34]. In this case, a large number of RIRs over multiple points is necessary.

This is mostly done by using *Ambisonics* which is a method based on spherical decomposition of a soundfield. In first-order ambisonics, each transmission channel carries a speaker-independent representation of the soundfield called B-format. The W channel is the non-directional mono component of the signal, and the X, Y, Z channels are the directional components, each derived from a figure-of-eight microphone. The main advantage of ambisonics is that it is not a fixed representation of the soundfield, which means that the number and layout of speakers used for playback is independent from the source panning. The quality of the representation can be improved by adding more directional components, leading to higher-order ambisonics.

Once the spatial encoding is complete, a procedure of spatial decoding using loudspeakers is needed to recreate the soundfield. Various methods can be used, like *decoding through projection* where each speaker receives its weighted sum of all ambisonic channels, or *decoding through pseudoinverse* which is based on a system of equations to describe the encoding and decoding.

Ambisonics presents the advantage of being flexible on the number and layout of loudspeakers. It also allows signals to be manipulated after the encoding through various operations, like the rotation of the soundfield around the solid angle. However, it only operates at a specific listening location and requires expensive material.

Another computed multi-speaker auralisation technique is Wave Field Synthesis (WFS) that uses a microphone and loudspeaker array such as that shown in Figure 3.1. The sound is captured by the microphones and instantly transmitted by the loudspeakers. The main advantage of this technique is the representation of the soundfield that is accurate on a large volume rather than a centralised listening position. However, the size of the required equipment makes it quite difficult to operate.
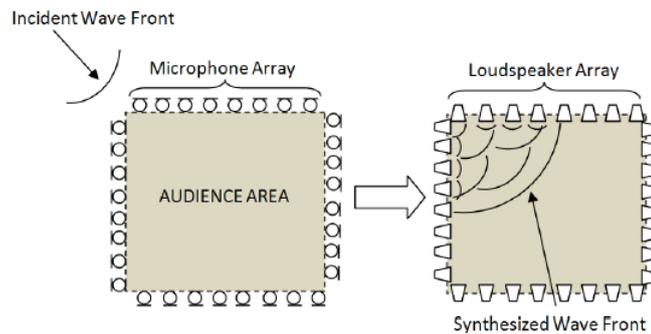
Figure 3.1: Wave Field Synthesis as applied to a rectangular portion of a soundfield. Image sourced directly from [24]
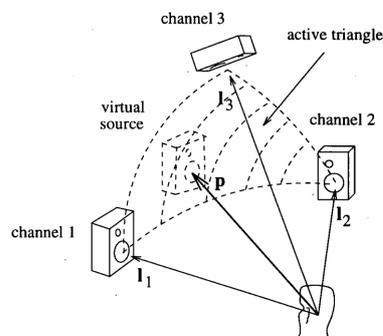
Figure 3.2: Sample configuration for 3D amplitude panning. Loudspeakers form a triangle in which the virtual source can be placed. Image sourced directly from [26]

Another way of reproducing the soundfield is to use a *Vector Based Amplitude Panning* (VBAP) that balances the gain of each loudspeaker to model a virtual source, like in Figure 3.2. This technique was developed by Pulkki [26] to create recordings independent of loudspeaker placement. It does not require as much equipment as WFS and allows recordings for any possible speaker configuration. However, it is not suitable for rendering spatial auralisations as reflections would seem to come from the direction of the virtual source.

3.2   Geometric Acoustic Models

In geometrical room acoustics, a sound wave is simplified as a sound *"ray"*, for which the total energy remains constant (but the intensity decreases as a spherical wave), and all the wave phenomena are neglected [17]. This approximation can be made for waves with smaller wavelengths (high frequencies) than the dimensions of the walls, considering diffraction requirements

(see 2.3.1). Geometric acoustic based modelling techniques are varied, with strengths that differ. Hence, the choice of the method can be adapted regarding the shape of the room, as stated in [27]. The main and more useful approaches for this project will be described in the following paragraphs.

3.2.1 Image Source Method (ISM)
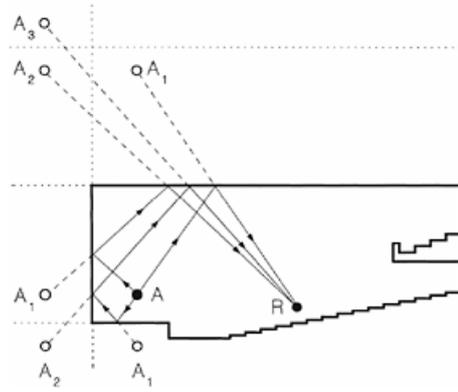
Figure 3.3: 2D Image Sources. A = Sound source, Ai = i-order image sources. Image sourced directly from [17].

As described in [27], the principle of the Image Source method is to use an image source solution, which is given by the reflection of the sound source against the considered surface. As the sound source is reflected against all surfaces, it leads to a set of image sources that are reflected again, resulting in more image sources. The process is repeated and ceases when a termination condition is met, such as response length or reflection order. The RIR is then obtained by superposing the contributions of these image sources.

For example, in Figure 3.3, which represents a basic concert hall, A_1 are the reflected images of the sound source A produced by the ceiling, the back wall and the floor of the stage. A_2 correspond to the second-order image of the source, given by the reflection of first order images A_1 towards the boundaries. The received signal s_R is given in equation (3.3), where s_A is the sound source, A_i is the amplitude coefficient for each image source determined by the absorption coefficients of the boundaries, and t_i the travelling time calculated from the distance between each image source and the receiver. If s_A is a dirac function, s_R will correspond to the RIR at these source and receiver positions.

$$s_R(t) = \sum_i A_i s_A(t - t_i) \qquad (3.3)$$

27

Another approach consists of selecting only the image sources that provide valid reflection paths [27]. This leads to reduce the amount of space that the image source method requires before the validation procedure.

3.2.2 Ray Tracing (RT)
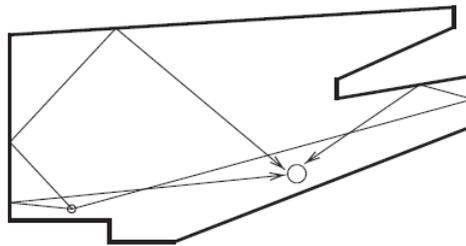
Figure 3.4: Ray Tracing Illustration. Image sourced directly from [27].

The Ray Tracing method consists of projecting ray paths originating from a sound source, and validating them if they join the position of a volumetric receiver [27], shown as a sphere in Figure 3.4 with the sphere. The rays can be emitted from the source randomly or using distributions taking into account the shape of the room, as studied in [16]. Acoustic modelling programs like CATT-Acoustic or ODEON provide efficient results using the ray tracing method and combining it with ISM.

This method differs from the image source method as it can consider arbitrary reflection properties, like diffuse reflections or scattering. Considering only limited specular reflections, the ray tracing solution will only be an approximation to the image source solution, that converges by increasing the number of rays used. The minimum number of rays for an accurate result is, however, difficult to determine and depends on the geometry and materials of the room. Nevertheless, with current computational resources, this technique provides accurate results and presents the advantage of taking into account more complex reflection properties.

3.2.3 Beam Tracing (BT)

The Beam Tracing method is similar to the Ray Tracing method, but differs by expanding the straight lines (rays) to volumetric objects (beams) [27], as shown in Figure 3.5. In this case, the receiver can be represented as a point instead of a volume. The shape of the beams can be varied with pyramidal beams covering the entire sphere around the source, and conical beams covering all path directions, as shown in Figure 3.6.
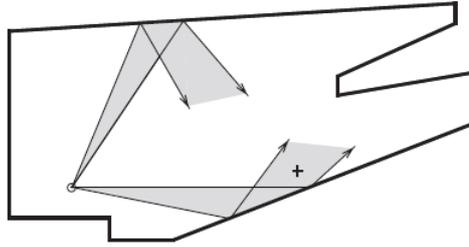
Figure 3.5: Beam Tracing Illustration. Image sourced directly from [27]

.
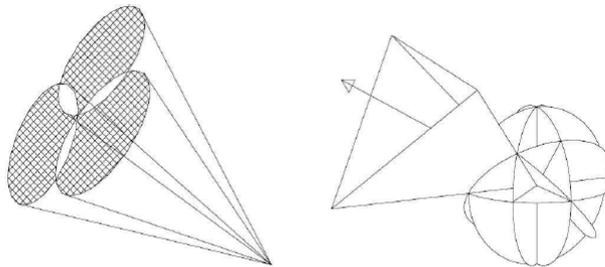
Figure 3.6: Visual comparison of conical and pyramidal beams projected from an omni-directional sound source. Image sourced directly from [27]

.

3.2.4 Radiosity

The radiosity method differs from those discussed previously as it is a surface-based method that assumes ideally diffuse reflections [27]. It is based on the principle of energy conservation during reflections, taking into account the absorption of the boundaries. This technique is therefore suited for diffuse sound fields such as modelling the later stages of the RIR.

3.3 Numerical Acoustic Models

Numerical techniques are based on a solution to the wave equation (2.5), considering physical wave motion. Unlike geometrical methods, they are valid over the entire audible bandwidth and present the advantage of taking into account wave phenomena such as diffusion, diffraction or scattering (see Section 2.2). However, they are very computationally intensive if a solution is required across the whole audible bandwidth.

One such example is the finite element method (FEM), which consists of dividing a room into non-uniform elements for which the sound pressure will be determined. The boundary element method (BEM) is based on the same process but only discretizes the boundaries of the room.
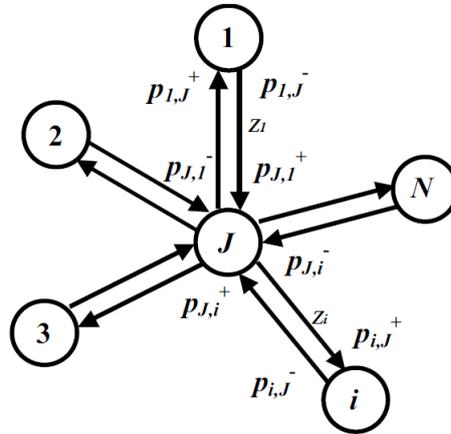
Figure 3.7: A general junction J with N connected junctions. Image sourced directly from [23] and edited.

However, due to their complexity and computational cost, these techniques will not be used in this research and will therefore not be developed further here. Algorithms based on the discretization of the wave equation are the only ones that will be considered in this section.

3.3.1 Digital Waveguide Mesh (DWM)

The one-dimensional digital waveguide is based on a time and space discretization of the d'Alembert solution to the 1-D wave equation as shown in 3.4. Van Duyne and Smith [35] proposed an extension to this model as a 2D digital waveguide mesh via an arrangement of bi-directional delay units and scattering junctions.

$$p(x,t) = p^+(ct - x) + p^-(ct + x) \tag{3.4}$$

The model is based on Kirchhoff's laws, stating that for a lossless junction J, the sum of the input velocities is equal to the sum of the output velocities, and that the sound pressures in all crossing waveguides are equal [20],[23]. For a junction J with N connected waveguides at pressure $p_{J,i}$ and with impedance Z_i (as represented on Figure 3.7), the pressure at the junction can be expressed as :

$$p_J = \frac{2\sum_{i=1}^{N} \frac{p_{J,i}^+}{Z_i}}{\sum_{i=1}^{N} \frac{1}{Z_i}} \tag{3.5}$$

The total pressure at the junction can also be defined as the sum of the input and output pressures :

$$p_J = p_{J,i}^+ + p_{J,i}^-$$

(3.6)

Updating temporally the pressure passing from a junction to its neighbour leads in the z-domain to :

$$p_{J,i}^+ = z^{-1} p_{i,J}^-$$

(3.7)

Combining the previous equations (3.5, 3.6, 3.7), it is possible to update temporally the pressure value at each junction from previous values with :

$$p_J = \frac{2 \sum_{i=1}^{N} \frac{p_i}{Z_i} z^{-1}}{\sum_{i=1}^{N} \frac{1}{Z_i}} - p_J z^{-2}$$

(3.8)

The digital waveguide mesh presents the advantage that impedance values associated with scattering junctions may be altered dynamically in the case of changes in the acoustic field. A limitation of the model is the dispersion error [23], due to the fact that the velocity of the wave depends on its frequency and direction of travel, leading to wave propagation errors and bias in the resonant modes.

3.3.2 Finite Difference Time Domain (FDTD)

Wave propagation can be described with partial differential expressions, combined to form the wave equation represented in (3.9) in a 3D cartesian coordinate system. The Finite Difference Time Domain (FDTD) approach consists of approximating those expressions, using the Taylor series, and discretizes them in time and space, rendering the simulation possible. It was developed by Yee [38] in 1966 in order to solve Maxwell's equations for electromagnetic applications, and then applied to acoustical simulations by Botteldooren [7].

Using Taylor's series, the first derivative of an arbitrary function $f(x)$ can be approximated as a first-order forward (3.10) or backward (3.11) difference :

$$c^2 \Delta p = \frac{\partial^2 p}{\partial t^2} \Leftrightarrow \frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2} = \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2}$$

(3.9)

$$f'_+(x) = \frac{f(x + \triangle x) - f(x)}{\triangle x} + O(\triangle x^2) \tag{3.10}$$

$$f'_-(x) = \frac{f(x) - f(x - \triangle x)}{\triangle x} + O(\triangle x^2) \tag{3.11}$$

Similarly, the second derivative of this function can be approximated as a second-order difference :

$$f''(x) = \frac{f(x + \triangle x) - 2f(x) + f(x - \triangle x)}{\triangle x^2} + O(\triangle x^3) \tag{3.12}$$

The finite difference operators are usually applied in an environment characterized by a regular spatial grid and sampling frequency. In the context of a non-staggered rectilinear standard leapfrog mesh, the 3D wave equation (3.9) can be rewritten as :

$$\frac{p_{l,m,j}^{n+1} - 2p_{l,m,j}^n + p_{l,m,j}^{n-1}}{T^2} = c^2 \Big(\frac{p_{l+1,m,j}^n - 2p_{l,m,j}^n + p_{l-1,m,j}^n}{X^2} + \frac{p_{l,m+1,j}^n - 2p_{l,m,j}^n + p_{l,m-1,j}^n}{Y^2} + \frac{p_{l,m,j+1}^n - 2p_{l,m,j}^n + p_{l,m,j-1}^n}{Z^2} \Big) \tag{3.13}$$

where T is the temporal sampling period, (X, Y, Z) the spatial sampling instance, n describes the position of the system in discretized time, and (l, m, j) are the space coordinates.

Hence, considering a regular grid of pressure nodes in the room (with $X = Y = Z$), the FDTD algorithm can provide the pressure value at the next time step at a certain space position (l, m, j), as :

$$p_{l,m,j}^{n+1} = \frac{c^2 T^2}{X^2} (p_{l+1,m,j}^n + p_{l-1,m,j}^n + p_{l,m+1,j}^n + p_{l,m-1,j}^n + p_{l,m,j+1}^n + p_{l,m,j-1}^n) + 2(1 - 3\frac{c^2 T^2}{X^2}) p_{l,m,j}^n - p_{l,m,j}^{n-1} \tag{3.14}$$

The FDTD method can alternatively be formulated on a staggered Yee's grid [38], using ordinary differential operators rather than partial ones. In this case, the update of the pressure value will be updated regarding the particle velocity components. However, the staggered scheme and the non-staggered one have been shown to have the same stability condition and the same dispersion effects (that will be described in the next chapter). It is then equivalent to use each of these models in the context of the RIR synthesis.

The 3D FDTD method can also be adapted into a 2D multiplane FDTD method, in order to increase computational efficiency and reduce the associated costs. This work has been done by Oxnard [24] in 2016 and combined with a geometric method (as it will be seen in the next subsection), it provides accurate results for low frequencies, with a better agreement to the 3D FDTD method than with geometric ones.

3.3.3 Finite Volume Time Domain (FVTD)

This method is similar to the FDTD one, using pressure cells instead of nodes. It has been first developed by Botteldooren [6], in the context of quasi-Cartesian grids, in order to simulate wave propagation in rooms with tilted and curved boundaries. The main idea is to decompose the space into adjacent cells, called *Voronoi cells*, and apply energy methods to update the average pressure in each cell. The implementation of this algorithm can be found in [5] for more details. It is more accurate than the FDTD method for rooms with complex geometries, but is also much more computationally intensive.

3.4 Hybrid Acoustic Models : Influence of the crossover frequency

Numerical methods are generally more accurate than geometric methods, especially at low frequencies. This is due to the fact that geometrical methods do not take into account wave phenomena, such as diffraction, diffusion or scattering. However, numerical approaches are computationally intensive, especially for large volumes or wide frequency bandwidth. One approach to this issue is to combine them, by applying numerical methods to low frequencies and geometrical ones to mid-high ones.

An application of this idea is the *RenderAir* system [21] developed by Murphy et al that combines a digital waveguide mesh approach for low frequencies and a ray-tracing approach for mid-high ones. As the early-part of the RIR is more perceptually relevant to the listener, it is simulated with a 3D mesh, and the later part with a 2D mesh, allowing to reduce the execution time. Reverberation time measurements from the synthesized RIR match those obtained by RIR measurements. However, variations can still be observed for measurements at certain positions of the receiver. It could also be interesting to validate the model by performing listening tests to confirm the results.

Another approach to synthesize the RIR has been developed in [30] and validated in [32] conducted by Southern et al, applying the 3D FDTD method for low frequencies. The mid-high

frequency range of the RIR is obtained by combining a beam tracing method for the early reflections and a radiosity method for the late reflections. The final response is obtained by using octave band filters, with an experimental calibration parameter to balance their contribution. The agreement between the synthesized and the measured RIR is correct, but as in the previous study, testing the model by listening experiments could be an asset for its validation. The 3D FDTD method has been restricted to multiplane 2D FDTD method in Oxnard's work [24], reducing considerably the run-time of the simulation, with acoustic results close to those from the 3D model.

In both cases, the final RIR is built combining the numerical model and the geometric one by taking into account their robustness on the appropriate frequency range. Therefore, the numerical model is low-pass filtered at a certain cut-off frequency, and the geometrical one is high-pass filtered at the same cut-off frequency, with the same filter order, to preserve the whole frequency range. After a process of calibration that will be described in the next chapter, the resulting fragments of RIRs are summed and aligned in the temporal domain, to form the final RIR.

The cut-off frequency of the filters, which is defined as the *crossover frequency* of the hybrid model, is usually set at an arbitrary value of 1 kHz as noted in prior work [17]. Indeed, wave phenomena occur for wavelengths that are superior or equal to the size of irregularities on the borders. The choice of a crossover frequency at 1kHz corresponding to a 34 cm wavelength is therefore justified due to the fact that linear dimensions of the walls and the ceiling are generally greater than 34 cm. However, the hybrid model could benefit from a reduction of the crossover frequency in terms of execution time and storage memory, due to the reduced computational requirements of the geometrical model compared to the numerical one.

For the case of ideal rooms without any irregularities on the walls, it can therefore be hypothesized to reduce the crossover frequency to the wavelength corresponding to the smallest distance between two boundaries. This assumption will be the heart of this research project which will aim to determine if the crossover frequency of a hybrid model can be reduced to that extent without compromising its quality. To that end, a 3D FDTD model will be built entirely in MATLAB, and will be combined with a geometric model generated via ODEON Auditorium to form a hybrid model as it will be described in Chapter 5. This model will be tested on 2 simple rooms, with various crossover frequencies and compared to real RIR measurements in terms of various results detailed in Chapter 5.

3.5 Summary

This chapter has provided a general review of virtual acoustic modelling methods to select an appropriate model for this research. First of all, the auralisation process has been presented, which consists of reproducing a soundfield via headphones or loudspeakers.

Virtually acoustic modelling methods provide the RIR synthesis of virtual enclosures, and can be categorized as geometric or numerical. Geometric approaches, like ray-tracing (RT) or image source methods (ISM), are based on the simplification of sound waves as rays. Numerical approaches, such as the digital waveguide mesh (DWM) or finite difference time domain (FDTD) methods, provide a discrete numerical solution to the wave equation, and offer an accurate simulation for the physical wave propagation. Each method has been listed and described to provide an adequate RIR synthesis method for this research.

Eventually, a hybrid model combining finite difference time domain method for the low frequency range and a ray tracing / image source method for the mid-high frequency range has been selected. The cut-off frequency of the filters used to create this hybrid response is called the *crossover frequency* of the model and is the key parameter of this reseach.

IMPLEMENTATION OF A HYBRID ACOUSTIC MODEL

As described in Chapter 3, this research project will focus on a virtual hybrid acoustic model, combining a geometric model for the mid-high frequency range that will be detailed in Section 4.3 and a numerical model for the low frequency range that will be detailed in Section 4.4. To determinate the influence of the crossover frequency on the model, several RIRs will emanate from the combination of the FDTD model and the geometric one, each with a different crossover frequency. This will aim to estimate the best compromise between the accuracy and the computational requirements of the RIR synthesis.

4.1   Presentation of the Study

The main goal of this study is to investigate the correlation between the crossover frequency of a hybrid response and the quality of the model. To that end, RIRs need at first to be recorded in existing enclosures, to be subsequently compared with the synthesized RIRs.

The rooms that were used for the project were chosen based on several criteria. They had to be empty and to have a simple geometry in order to be not too complicated to model. All of the materials composing the room had to be known in order to estimate absorption coefficients. For each room, various positions for the source and receiver were selected to improve the accuracy of the results.

The first room is an empty bedroom from a house based in York. It is a $(3.28 \times 2.23 \times 2.66)m$ cuboid shape, with a door, a window, a small closet and a hatch. The room can be seen on Figure

4.1 and will be referred as "Room 1" in the rest of this research.
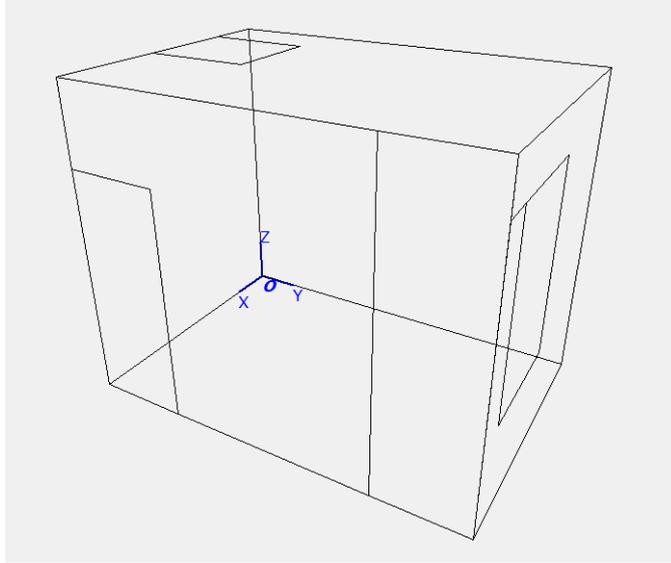
Figure 4.1: Perspective view of Room 1.

The second room is another empty bedroom from the same house. It is a $(3.61 \times 3.10 \times 2.66)m$ cuboid shape, with a chimney area, a door, a window area. It is a bit more complex than the previous one due to the chimney and the window frame. The room can be seen on Figure 4.2 and will be called "Room 2" in the rest of this research.

Considering the audible human range between 20 Hz and 20 kHz, the Nyquist-Shannon sampling theorem [18] gives a direct condition on the sampling temporal frequency in (4.1). To provide a margin for anti-aliasing filters, the sampling frequency will be set at $F_s = 44100$ Hz.

$$F_s = \frac{1}{T} \geq 40\text{kHz} \tag{4.1}$$
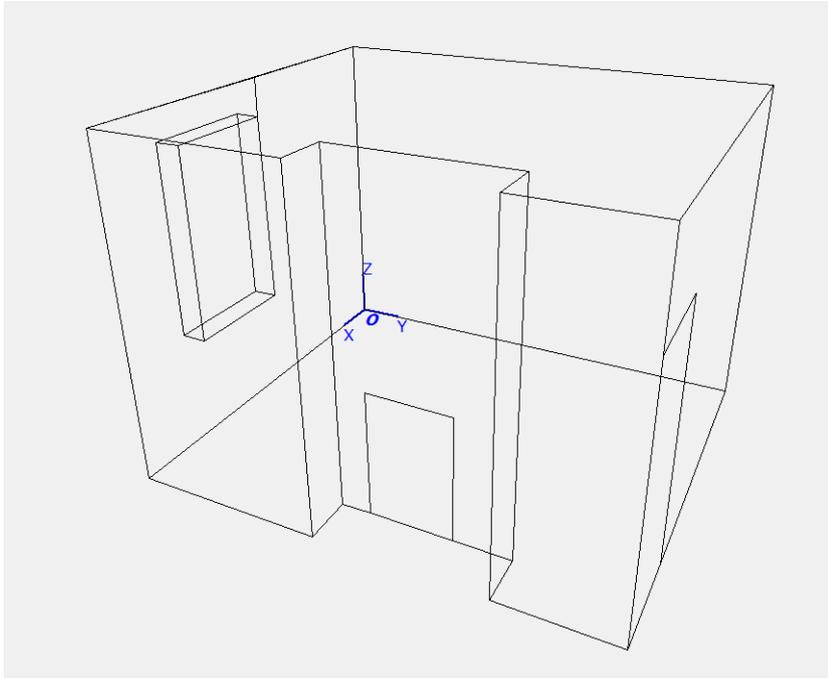
Figure 4.2: Perspective view of Room 2.

4.2 Live RIR Measurements

RIR measurements were taken in each room for comparison with the acoustic models. Three source and receiver placements enable the extension of this investigation to different cases. Their positions are listed in Table 4.1.

The exponential swept-sine technique developed by Farina [8] was used to provide a practical RIR, as described in Section 2.3. 30 second exponential sine sweeps were emitted via a Genelec 8130A loudspeaker rotating around the azimuthal plane at increments of 90°, to approximate an omnidirectional sound source. For each case, responses were captured by a ST450 Soundfield B-Format microphone and converted into a mono RIR by summing their W-channel at each orientation. Three RIRs were obtained for each room, each corresponding to a source/receiver position, as described in Table 4.1. In total, 6 RIRs were measured.

Table 4.1: *Source and receiver placements for each case study*

CASE STUDY 1

	Source (x,y,z) (m)	Receiver (x,y,z) (m)
1	(1.64, 0.56, 1.30)	(1.64, 1.90, 1.30)
2	(0.34, 1.11, 1.30)	(2.72, 1.13, 1.30)
3	(2.73, 1.78, 0.98)	(0.25, 0.31, 1.49)

CASE STUDY 2

	Source (x,y,z) (m)	Receiver (x,y,z) (m)
1	(1.55, 0.36, 1.25)	(1.55, 3.29, 1.25)
2	(2.66, 0.42, 0.98)	(0.28, 3.28, 1.49)
3	(0.40, 1.80, 1.25)	(2.33, 1.80, 1.25)

4.3 The Geometric Model

The geometric model is generated via ODEON Auditorium 13.04 acoustic prediction software [2]. ODEON combines ray-tracing methods that project ray paths originating from the sound source, and image-source methods (ISM) that are based on the reflection of the source against surfaces, to provide an appropriate RIR synthesis of the enclosure.

Room dimensions and source/receiver positions were replicated as described in the previous section. Each room geometry was built using SketchUp 2017 [1], a 3D modelling software. The room was exported in a *.par* file to be open in ODEON Auditorium.

The absorption coefficients as described in Chapter 2 are essential to the accuracy of the model, as they will affect the magnitude of the room modes. Each surface was allocated a material type from the ODEON library [2] to associate it with random-angle-incidence absorption coefficients for each octave band. As it was likely that the real absorption coefficients would not exactly correspond to those from the library, a calibration process was done, following Postma's procedure [25] :

- Reverberation Parameters (T_{30}) are calculated from the RIR measurements and estimated from the geometric model, on each octave band.

- Absorption coefficients are adjusted to bring the reverberation parameters to within 1 just noticeable difference (JND) of the measured value, which corresponds to 5 % of T_{30}. The

first materials to be adjusted are those with the largest surface areas to have considerable effects on the value.

- The scattering coefficient is adjusted the same way based on clarity (C_{80}).

Eventually, the absorption coefficients are registered in Table 4.2. As the rooms are from the same house, they are the same for each enclosure.

Table 4.2: *Random-angle-of-incidence Absorption coefficients for each surface type as applied in the geometric model for Room 1 and 2.*

Surface	α_{63}	α_{125}	α_{250}	α_{500}	α_{1000}	α_{2000}	α_{4000}	α_{8000}
Floor	0.10789	0.02404	0.01663	0.01409	0.03633	0.07060	0.05130	0.06020
Ceiling	0.01051	0.01893	0.04534	0.00899	0.04588	0.01595	0.04269	0.02178
Walls	0.00643	0.00380	0.00802	0.01961	0.00258	0.00946	0.01505	0.01354
Window	0.05903	0.06377	0.09825	0.07458	0.03216	0.09473	0.02725	0.03079
Doors, Closet and Hatch	0.17224	0.04147	0.02407	0.01971	0.03439	0.02273	0.01409	0.03825
Chimney	0.04197	0.40015	0.00035	0.96493	0.98517	0.98860	0.97617	0.00092

Image source methods are used up to a transition order that can be manually controlled, and ray-tracing methods are used for the upper reflection orders and diffuse fields. This transition order was experimentally set at the 5th order. 50000 rays are used to optimise the RIR, resulting in a B-Format ".wav" output, from which the W-channel is extracted for post-simulation.

A gap was noticed between the theoretical direct sound and the one extracted from the geometric model, due to the RIR simulation not starting at the beginning of the file. This issue was solved by aligning the direct sound of the geometric model with the FDTD one, which corresponded to the theoretical one.

4.4 The Numerical Model

Among all of the numerical methods, the hybrid acoustic model will be based on a 3D FDTD model for low frequencies. Element methods (FEM and BEM) present the advantage of being effective for challenging structures, but at a high computational cost due to their complexity, and will therefore not be relevant in this study. According to Smith [30], it can be shown that digital waveguide and FDTD methods are related by a change of coordinates, and can therefore be considered as equivalent. However, the digital waveguide method is not as efficient as the FDTD one for multidimensions, which justifies the choice of this method. It will be implemented in MATLAB.

As explained in Chapter 3, the FDTD method is based on the decomposition of the space into a grid of pressure nodes, and on the finite difference approximations to the wave equation. The choice of an adequate scheme is decisive for the accuracy and efficiency of the model.

4.4.1 Formulation of the Standard Rectilinear FDTD scheme

The scheme can be considered as *explicit* or *implicit*. Explicit schemes are based on equation (3.14) for a standard rectilinear Cartesian grid, where the pressure at time $n + 1$ depends explicitly on the pressure at time n. They are well-suited for real-time auralisations, but present the main drawbacks of anisotropy and numerical dispersion errors, which will be discussed in the next subsections.

Implicit schemes are based on the finite difference approximation of the wave equation at the next time-step, evaluating the spatial derivatives Δp at time $n + 1$. In other words, expression (3.14) in the case of a standard rectilinear Cartesian grid will be rewritten as :

$$
\begin{aligned}
p_{l,m,j}^{n+1} - \frac{c^2 T^2}{X^2} ((p_{l+1,m,j}^{n+1} - p_{l,m,j}^{n+1}) + (p_{l-1,m,j}^{n+1} - p_{l,m,j}^{n+1}) \\
+ (p_{l,m+1,j}^{n+1} - p_{l,m,j}^{n+1}) + (p_{l,m-1,j}^{n+1} - p_{l,m,j}^{n+1}) \\
+ (p_{l,m,j+1}^{n+1} - p_{l,m,j}^{n+1}) + (p_{l,m,j-1}^{n+1} - p_{l,m,j}^{n+1})) = 2p_{l,m,j}^n - p_{l,m,j}^{n-1}
\end{aligned}
\tag{4.2}
$$

They can be effective for reducing numerical dispersion, but require a linear system to be solved at each time-step, which is very computationally intensive [15]. In [9], Hamilton and Bilbao suggest a solution employing the Jacobi iterative method that will alleviate the calculation. However, this process would be adequate for situations on a limited frequency bandwidth, and would also increase the execution time. Therefore, this study will be limited to an explicit scheme.

The choice of the position of the pressure nodes in space is important for the accuracy and effectiveness of the results. The simplest is the standard leapfrog (SLF) scheme that consists of calculating the pressure at the next time step of a pressure node, based on its 6 neighbours on a 3D rectilinear grid, as in Figure 4.3.

In this context, the expression of pressure at the next time-step $n + 1$ is given in (4.3) based on the FDTD algorithm (see Subsection 3.3.2) for a spatial sampling period X and a temporal sampling period T.

$$
p_{l,m,j}^{n+1} = \lambda^2 S_{l,m,j}^n + 2(1 - 3\lambda^2)p_{l,m,j}^n - p_{l,m,j}^{n-1}
\tag{4.3}
$$

where
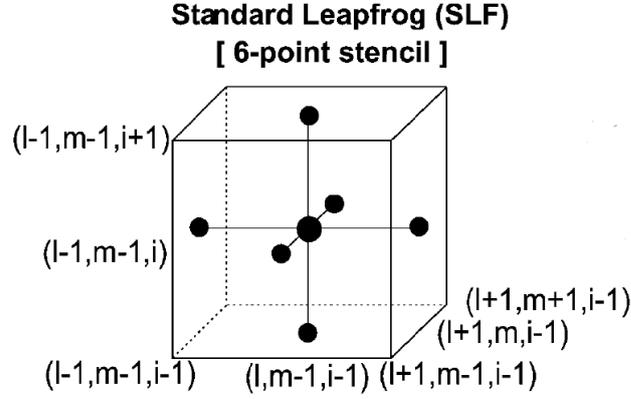
Figure 4.3: Standard leapfrog scheme. Image sourced directly from [15].

$$S^n_{l,m,j} = p^n_{l+1,m,j} + p^n_{l-1,m,j} + p^n_{l,m+1,j} + p^n_{l,m-1,j} + p^n_{l,m,j+1} + p^n_{l,m,j-1} \qquad (4.4)$$

and λ is defined as the Courant number :

$$\lambda = \frac{cT}{X} \qquad (4.5)$$

Applying another approximation to the spatial Laplacian operator can lead to different kinds of schemes, such as those reviewed in [15]. In this article Kowalczyk presents different types of meshes, such as the octahedral (OCTA), the cubic close-packed (CCP) and interpolated schemes, and compares them in terms of stability, accuracy, and computational efficiency. It can, for example, be noticed that interpolated isotropic meshes have a significantly higher efficiency, and would be the most suitable for this study. However, they are a lot harder to implement than the SLF scheme. Therefore, as this study is mostly focused on the crossover frequency, the SLF scheme will be the one implemented.

Absorption coefficients used in the numerical model are the ones estimated for the geometric model in Table 4.2 in Section 4.3. They are random-angle-of-incidence ones, which mean that they are correct for sound rays coming from a random direction, as in the case of the geometric model. However, in the case of the FDTD model, due to the shape of the grid, normal-angle-of-incidence will be used and can be derived from the previous ones following Siltanen's formula [31] (4.6). The normal-angle-of-incidence reflection coefficients used for the numerical model are displayed on Table 4.3.

43

$$R_{\mathrm{norm}} = -0.22897R_{\mathrm{rand}}^2 + 0.99298R_{\mathrm{rand}} + 0.23769 \qquad (4.6)$$

Table 4.3: *Normal-angle-of incidence Reflection coefficients for each surface type as applied in the FDTD model for Room 1 and 2.*

Surface	R_{63}	R_{125}	R_{250}	R_{500}	R_{1000}	R_{2000}	R_{4000}	R_{8000}
Floor	0.9561	0.9812	0.9833	0.9979	0.9918	0.9822	0.9876	0.9851
Ceiling	0.9850	0.9826	0.9750	0.9993	0.9892	0.9974	0.9900	0.9958
Walls	0.9862	0.9869	0.9857	0.9964	1.0010	0.9992	0.9976	0.9981
Window	0.9710	0.9696	0.9591	0.9810	0.9930	0.9752	0.9943	0.9933
Doors, Closet and Hatch	0.9353	0.9761	0.9812	0.9964	0.9924	0.9956	0.9979	0.9913
Chimney	0.9902	0.8694	1.0016	0.4156	0.3552	0.3411	0.3855	1.0015

4.4.2 Locally Reacting Surface (LRS) Boundaries

The correct reflection at boundaries is essential for the quality of the acoustic model. As seen in section 2.2, the boundary is characterized by its acoustic impedance Z, which is frequency dependent in realistic cases. However, such values are quite complicated to obtain and necessitate an accurate knowledge of the materials used to build the boundary. One simpler way to approach frequency-dependent boundaries, is to use acoustic impedances that are constant on each octave band. In this case, the RIR is simulated 8 times, 8 being the total number of octave bands, with values of reflection coefficients corresponding to each octave band, as displayed in Table 4.3. The 8 resulting RIRs are then filtered according to each octave bandwidth and summed to form the final RIR [28]. This approach is the one that will be used in this research project, but only the formulation of locally reacting surfaces with frequency-independent boundaries will be described, as it is the one that will be used for each octave band.

The FDTD formulation of locally reacting surfaces (LRS) has been extensively discussed by Kowalczyk [14, 15] and is based on the assumption that wave propagation can be locally assumed to be one-dimensional at a boundary. In this context, if we consider a sound wave travelling in a positive x-direction and a boundary normal to the x-direction, the equation of momentum (2.2) as seen in Chapter 2 can be derived as:

$$\frac{\partial p}{\partial x} = -\rho \frac{\partial v_x}{\partial t} \qquad (4.7)$$

As seen in Chapter 2, the specific boundary impedance relates the sound pressure to the flow normal to the boundary as:

$$\zeta_x = \frac{1}{\rho c} \frac{p}{v_x} \qquad (4.8)$$
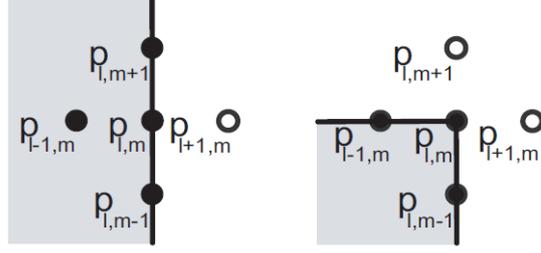
Figure 4.4: Example of a leapfrog stencil in a 2D rectilinear mesh at a boundary and at a corner. Ghost-point nodes are indicated with white-coloured circles and the room interior by grey shading. Image sourced directly from [15].

Differentiating (4.7) combined with equation (4.8) leads to the boundary condition in terms of pressure only that will be subsequently implemented in the FDTD model :

$$\frac{\partial p}{\partial t} = -c\zeta_x \frac{\partial p}{\partial x} \tag{4.9}$$

Using an SLF scheme, the pressure nodes can be updated using (4.3) on the totality of the mesh, except on the boundaries, where one pressure node called a *ghost point* (or several for edges and corner) will fall outside of the space domain, as illustrated in Figure 4.4. Using centered finite difference operators, (4.3) can be approximated as (4.10) that leads to the formulation of the ghost node depending on its neighbours as presented in equation (4.11).

$$\frac{p_{l,m,j}^{n+1} - p_{l,m,j}^{n-1}}{2T} = -c\zeta_x \frac{p_{l+1,m,j}^{n} - p_{l-1,m,j}^{n}}{2X} \tag{4.10}$$

$$p_{l+1,m,j}^{n} = p_{l-1,m,j}^{n} + \frac{1}{\lambda\zeta_x}(p_{l,m,j}^{n-1} - p_{l,m,j}^{n+1}) \tag{4.11}$$

Replacing the ghost point in the general pressure update equation (4.3) yields the following equation for a right boundary in a 3D formulation:

$$p_{l,m,j}^{n+1} = \frac{2(1-3\lambda^2)}{1+\frac{\lambda}{\zeta_x}} p_{l,m,j}^{n+1} + \frac{\lambda^2}{1+\frac{\lambda}{\zeta_x}}(2p_{l-1,m,j}^{n} + p_{l,m+1,j}^{n} + p_{l,m-1,j}^{n} + p_{l,m,j+1}^{n} + p_{l,m,j-1}^{n}) + \frac{\frac{\lambda}{\zeta_x}-1}{1+\frac{\lambda}{\zeta_x}} p_{l,m,j}^{n-1} \tag{4.12}$$

The same process is used for edges and corners where there are respectively two and three ghost nodes. For example, with a right-front edge and with a right-front-top corner, the general

pressure update equation (4.3) yields the following equations:

$$p_{l,m,j}^{n+1} = \frac{2(1-3\lambda^2)}{1+\frac{\lambda}{\zeta_x}+\frac{\lambda}{\zeta_y}}p_{l,m,j}^{n+1} + \frac{\lambda^2}{1+\frac{\lambda}{\zeta_x}+\frac{\lambda}{\zeta_y}}(2p_{l-1,m,j}^n+2p_{l,m-1,j}^n+p_{l,m,j+1}^n+p_{l,m,j-1}^n) + \frac{\frac{\lambda}{\zeta_x+\frac{\lambda}{\zeta_y}}-1}{1+\frac{\lambda}{\zeta_x}+\frac{\lambda}{\zeta_y}}p_{l,m,j}^{n-1}$$
(4.13)

$$p_{l,m,j}^{n+1} = \frac{2(1-3\lambda^2)}{1+\frac{\lambda}{\zeta_x}+\frac{\lambda}{\zeta_y}+\frac{\lambda}{\zeta_z}}p_{l,m,j}^{n+1} + \frac{\lambda^2}{1+\frac{\lambda}{\zeta_x}+\frac{\lambda}{\zeta_y}+\frac{\lambda}{\zeta_z}}(2p_{l-1,m,j}^n+2p_{l,m-1,j}^n+2p_{l,m,j-1}^n) + \frac{\frac{\lambda}{\zeta_x+\frac{\lambda}{\zeta_y}+\frac{\lambda}{\zeta_z}}-1}{1+\frac{\lambda}{\zeta_x}+\frac{\lambda}{\zeta_y}+\frac{\lambda}{\zeta_z}}p_{l,m,j}^{n-1}$$
(4.14)

4.4.3 Numerical Stability

Due to time and space discretization, conditions for stability need to be applied on the scheme, in order to avoid exponential growing or damping solutions of the system. Numerical stability can therefore be achieved applying von Neumann analysis to the scheme which is based on the decomposition of the errors into Fourier series [4].

Consider the SLF scheme in a 3D coordinate system x-y-z, where T and X are respectively the temporal and spatial frequency of the system. Based on the expression (2.7) generalised to a 3D case and discretized, the pressure value of a plane wave at a given point of the grid is :

$$p_{l,m,j}^n = p_0 e^{i\omega nT} e^{-i(k_x lX + k_y mX + k_z jX)}$$
(4.15)

Inserting this expression in (4.3) gives :

$$e^{i\omega T}p_{l,m,j}^n = \lambda^2\left(e^{-ik_xX}+e^{ik_xX}+e^{-ik_yX}+e^{ik_yX}+e^{-ik_zX}+e^{ik_zX}\right)p_{l,m,j}^n$$
$$+2(1-3\lambda^2)p_{l,m,j}^n - e^{-i\omega T}p_{l,m,j}^n$$
(4.16)

$$\Longleftrightarrow$$

$$e^{i\omega T}-2+e^{-i\omega T} = \lambda^2\left(e^{-ik_xX}-2+e^{ik_xX}+e^{-ik_yX}-2+e^{ik_yX}+e^{-ik_zX}-2+e^{ik_zX}\right)$$

which can be rewritten also as :

$$4\sin^2\left(\frac{\omega T}{2}\right) = \lambda^2\left[4\sin^2\left(\frac{k_xX}{2}\right)+4\sin^2\left(\frac{k_yX}{2}\right)+4\sin^2\left(\frac{k_zX}{2}\right)\right]$$
(4.17)

In the z-domain where $z = e^{i\omega T}$, equation 4.16 can be simplified as :

$$z + 2B(\hat{k}, \lambda) + z^{-1} = 0 \tag{4.18}$$

where $\hat{k}$ is the wave vector which corresponds to the number of waves per unit distance in 3-D, and B is defined as :

$$B(\hat{k}, \lambda) = 2\lambda^2 \left(\sin^2\left(\frac{k_x X}{2}\right) + \sin^2\left(\frac{k_y X}{2}\right) + \sin^2\left(\frac{k_z X}{2}\right) \right) - 1 \tag{4.19}$$

In the z-domain, the necessary stability condition is defined as $|z| \leq 1$, which from equation 4.18 is equivalent to :

$$\left| B(\hat{k}, \lambda) \right| \leq 1 \tag{4.20}$$

and gives a sufficient constraint on the Courant number as :

$$\lambda = \frac{cT}{X} \leq \frac{1}{\sqrt{3}} \tag{4.21}$$

As the sampling frequency is set to 44 100 Hz (see Section 4.1), the spatial period can be directly derived from the Courant limit (4.21), with the smallest value as possible to minimise the dispersion error and enhance the accuracy of the model. Therefore, $X = 13.5 \cdot 10^{-3}$ m.

4.4.4 Dispersion Error

An unwanted phenomenon called *dispersion* can appear in this model at high frequencies, with waves travelling at lower speed than in a continuous case [4]. This happens when the numerical phase velocity is different than the theoretical speed of sound as defined in the model, and is due to the application of finite difference operators. This error is commonly expressed using the *relative phase velocity* defined as the ratio of the effective numerical wave speed $\hat{c}$ over the real wave speed c :

$$v(\hat{k}, \omega) = \frac{\hat{c}}{c} = \frac{\omega}{\hat{k} c} \tag{4.22}$$

From equation 4.17, the angular frequency ω can be expressed depending on the wave number $\hat{k}$ as :

$$\omega = \frac{2}{T} \sin^{-1}\left[\lambda \sqrt{\sin^2\left(\frac{k_x X}{2}\right) + \sin^2\left(\frac{k_y X}{2}\right) + \sin^2\left(\frac{k_z X}{2}\right)} \right] \tag{4.23}$$

which directly gives the relative phase velocity :

$$v(\hat{k}, \omega) = \frac{\omega}{\hat{k}c} = \frac{2\sin^{-1}\left[\lambda \sqrt{\sin^2\left(\frac{k_x X}{2}\right) + \sin^2\left(\frac{k_y X}{2}\right) + \sin^2\left(\frac{k_z X}{2}\right)} \right]}{\lambda \sqrt{(k_x X)^2 + (k_y X)^2 + (k_z X)^2}} \tag{4.24}$$

To enhance the precision of the model, the grid spacing needs to be at its smallest value, respecting the stability condition in (4.21), which sets the Courant number to $1/\sqrt{3}$. In [14], the criterion for the admissible isotropy error is limited to 2%. In this case, the constraint on the relative velocity given in (4.24) is given as $0.98 \leq v \leq 1.02$ which leads to a limitation on the angular frequency ω. Therefore, by the calculations from Kowalczyk [15], this will give a constraint on the frequency bandwidth to $0.196 F_s$ to limit dispersion error.

4.4.5 Source Excitation Strategies

In room acoustics, as seen in 2.3, the RIR is defined as the evolution of the pressure through time at a specific position in an enclosed room, following an impulse excitation at another specific position. However, in concrete situations for simulations and measurement, a perfect impulse excitation is not achievable. Therefore, other strategies can be implemented in order to obtain a similar result, as described for real measurements in 2.3.3 with the sine-sweep method.

In the case of FDTD simulations, it is impossible to generate an exact Dirac delta function as in equation (2.44), due to time discretization. However, time-varying sources can be used. According to [22], [12], the main constraints on the source excitation to make it acceptable for FDTD simulations, are that its propagation should be omnidirectional and that its frequency response should be flat over a defined bandwidth. The excitation signal will also have to be sampled by the same sampling frequency as the FDTD scheme. Three models of sources are commonly referred in the literature, with advantages and drawbacks, that make them properly suitable for a situation.

Hard Source : The hard source type is described by imposing a driving function to a pressure node that will not be influenced by the FDTD update. It presents the advantage to be easy to implement but can lead to fluctuations for low frequencies according to [22].

Soft Source : With the soft source, the driving function is added to the pressure value at the source grid point [22]. It is easy to implement, like the hard source, but the pressure values

at the source node will not correspond to the applied driving function. Therefore, it will be necessary to normalize the pressure values on the whole grid by the pressure value at the position of the source.

Transparent Source : The transparent source drives the same signal as the hard source but with the advantage of avoiding scattering. However, according to [22], a compensation filter is required prior to the simulation, which reduces the computational efficiency.

From the results of the test performed in [22] that compares the three different types of sources for 3D and 2D FDTD simulations, the soft source will be the one that will be used for the purpose of this study. Indeed, the hard source is shown to be the least favourable one, due to the appearance of wave scattering effects, and the transparent source is harder to implement and more computationally intensive.

One significant property of the soft source relates to the inclusion of additional wave equation solutions due to source-boundary interaction effects. This results in a DC resonance that will need to be filtered post simulation.

4.5   Hybridisation Process

The hybrid model is built by combining the numerical and geometrical RIRs respectively in the low and mid-high frequency range. This synthesis is achieved via the procedure that was developed by Southern et al. [31] to obtain a valid Hybrid Acoustic Model (HAM).

To conserve the frequency dependency of the boundaries, the FDTD model is generated 8 times, with absorption coefficients corresponding to each octave band, as displayed in Table 4.3. The final RIR is obtained by filtering the 8 responses with a 5th order Butterworth filter on each octave band, and summing them.

Due to the use of a soft source and the phase preserving nature of the boundary conditions applied in the FDTD model, a drifting effect appears on the numerical RIR, resulting in a DC resonance, as mentioned in the previous chapter. Therefore, a DC blocking filter is applied as a 2nd order Butterworth filter. A low-pass filter is also required to limit the RIR to the valid bandwidth ($0.196F_s$) due to the dispersion effects mentioned in 4.4.5.

For a correct hybridisation, the energy levels of both RIRs have to be adjusted. To that end, a calibration process is necessary, based on the principle that the total energy received at at a

certain distance from the source should be the same for each RIR. This calibration parameter can be easily extracted for the geometric model since the RIR is automatically normalized by the software. Therefore, the calibration parameter $K_{geometric}$ for the geometric model is given by :

$$K_{geometric} = \frac{1}{p_{dir}r} \tag{4.25}$$

where p_{dir} corresponds to the amplitude of the direct sound component and r is the distance between the source and the receiver for each case.

The calibration process is more complex for the FDTD model since the spread of energy depends on the inter-nodal distance. It can still be estimated following [31], leading to the coefficient :

$$K_{FDTD} = 5.437 \times 10^{-3} F_s - 3.6347 \tag{4.26}$$

where F_s is the sampling frequency of the FDTD model.

The hybrid model is created by applying a low-pass filter to the FDTD response (Butterworth, 11th order) and a high-pass filter to the geometric response (Butterworth, 11th order). Indeed, the frequency response of a Butterworth filter is maximally flat, which leads to no distortion on the spectrum of the FDTD and geometric responses outside of the cut-off frequency zone. The result is summed after having aligned the two responses in the time domain based on their direct sound component, to form the final hybrid model.

The cut-off frequency of the two filters is the crossover frequency of the hybrid model which is the key parameter for this study. Therefore, 60 hybrid responses are generated with a crossover frequency varying from 20 Hz to 1200 Hz, with a 20 Hz step. It should be noted that the orders of the two filters are adapted with the reduction of the crossover frequency to maintain stability.

4.6 Summary

This chapter has detailed the implementation of the hybrid responses that will be evaluated in the next chapter to assess the influence of their crossover frequencies. The models are based on real measurements recorded in two different rooms in an empty house, at 3 source/receiver positions.

The geometric model, which is used for the mid-high frequency range, is generated via ODEON [2], a software combining ray-tracing and image source methods. The absorption coefficients are approximated via ODEON's library and adjusted with acoustic parameters extracted

from real measurements. The response is outputted to a B-format '.wav' file, from which the W channel is extracted.

The numerical model, which is used for the low frequency range, is generated using the FDTD method, which is based on the discretization of the wave equation in time and space. The explicit standard leapfrog (SLF) scheme is selected for the implementation of the model, being efficient and simpler to set. The absorption coefficients are adapted from the ones estimated for the geometric model. To preserve the frequency dependency of the boundaries for the FDTD model, RIRs are generated 8 times, each time with a different set of absorption coefficients corresponding to each octave band. The final RIR is obtained by filtering each of these responses on the corresponding octave bandwidth and by summing them. Particular attention was given on numerical dispersion effects that limit the upper frequency of the valid bandwidth to $0.196F_s$. A soft sound source is used, resulting in a DC resonance that will be filtered post simulation.

The hybrid model is built by combining the numerical RIR and the geometric RIR after a calibration process. The numerical RIR is filtered with a low-pass filter and the geometric one with a high-pass filter. The cut-off frequency of these two filters is the same and is defined as the crossover frequency of the hybrid model, which is the focus of this study. Therefore, for each room and each source/receiver position, 60 hybrid responses are generated with a crossover frequency varying from 20 Hz to 1200 Hz.

ANALYSIS OF THE SYNTHESIZED HYBRID RESPONSES

The main goal of this study is to investigate the correlation between the crossover frequency of a hybrid response and the quality of the model. In theory, reducing the crossover frequency of a hybrid model would decrease its computational requirements since the numerical model would only be generated on a smaller bandwidth. However, it might also reduce the quality of the model since the FDTD model is more precise at low frequencies, taking into account wave phenomena (see Chapter 2). If it can be proved that it is possible to reduce the crosssover frequency of a hybrid model without compromising its quality, it would be beneficial for many purposes requiring real-time auralisation.

To provide a more informative analysis, the study was conducted on 2 different rooms (as described in 4.1) from which RIRs were recorded at 3 different positions. For each source/receiver position, 60 hybrid responses were generated from the same FDTD and geometric RIRs that were synthesized respectively via MATLAB and ODEON (see Chapter 4). The specificity of each of these hybrid responses is their crossover frequency, varying from 20 Hz to 1200 Hz, with a 20 Hz step. The hybrid responses and the real measurements will then be compared to provide an estimation of the minimal crossover frequency not influencing the quality of the model.

This chapter is divided in 3 sections. The first two are dedicated to the results in each room, combining time domain analysis, frequency domain analysis, and the use of objective acoustic parameters. The last one provides a global conclusion on the analysis conducted in each room.
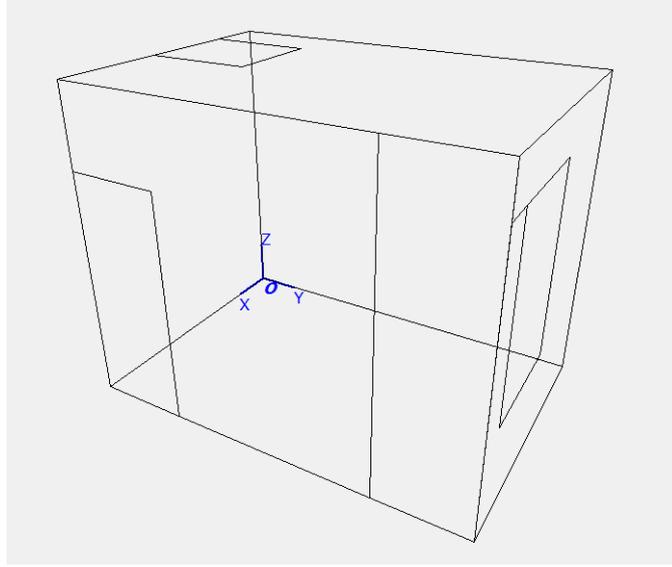
Figure 5.1: Perspective view of Room 1.

5.1   Case Study 1

The first room to be considered is an empty bedroom from a house based in York. It is a $(3.28 \times 2.23 \times 2.66)m$ cuboid shape, with a door, a window, a small closet and a hatch, as represented on Figure 5.1. To permit validation with real measurements, the source and receiver positions displayed in Table 5.1 match those used for the acoustic measurement positions presented in Table 4.1.

Table 5.1: *Source and receiver placements for Case Study 1*

CASE	Source (x,y,z) (m)	Receiver (x,y,z) (m)
1	(1.64, 0.56, 1.30)	(1.64, 1.90, 1.30)
2	(0.34, 1.11, 1.30)	(2.72, 1.13, 1.30)
3	(2.73, 1.78, 0.98)	(0.25, 0.31, 1.49)

This research is conducted to prove the hypothesis that the crossover frequency of the hybrid model can be reduced down to the smallest distance between two boundaries in the room. Regarding the room dimensions, it can be stated that the minimum wavelength should be 0.92 m which is the window width, meaning that considering the sound velocity as 343 m.s^{-1}, the crossover frequency could be reduced down to 373 Hz.

As mentioned in Section 2.2, the theoretical Schroeder frequency can be calculated from the

room dimensions. For case study 1, its value is $f_{Schroeder} = 1185 Hz$.

5.1.1 Analysis in the Time Domain

As mentioned in Section 2.3, the early rays of a RIR offer a first indication of the space. Those can be visualised for a wide range (20 Hz, 300 Hz, 700 Hz and 1000 Hz) of crossover frequencies and compared with the real RIR (orange) for the first source/receiver position on Figure 5.2. The amplitude of the RIR has been squared for a better visualisation. In order to validate the model, it has been checked that the main peaks correspond to the theoretical time positions of the direct sound and early rays.
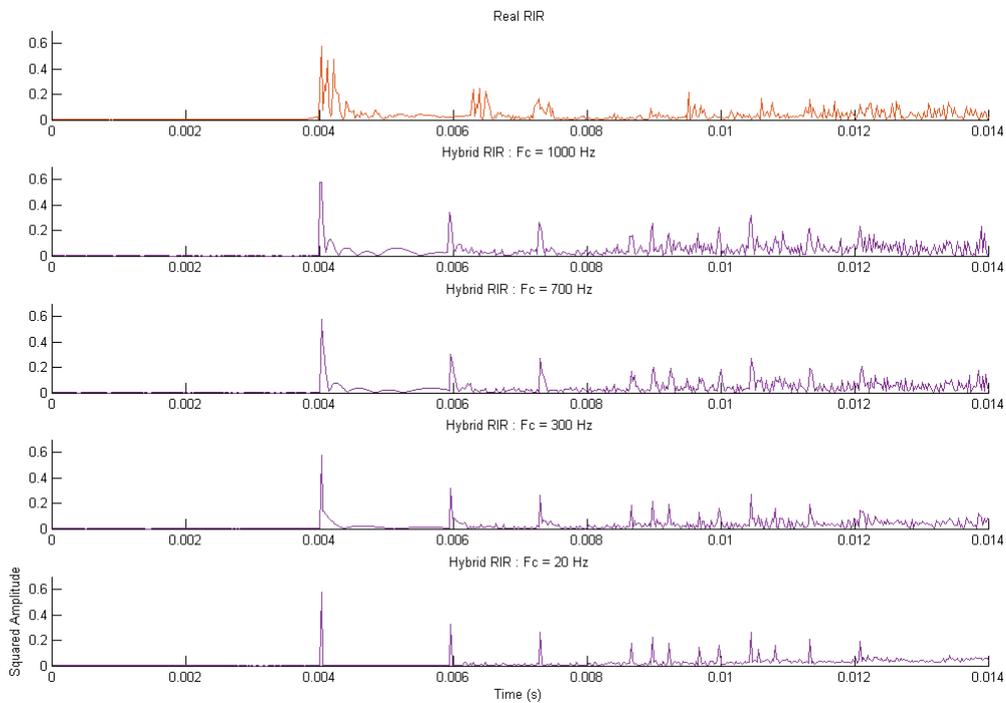
Figure 5.2: Case Study 1 : Real and hybrid impulse responses (early rays) in the Time Domain at Source/Receiver position 1. Crossover frequency values are 20 Hz, 300 Hz, 700 Hz, and 1000 Hz.

The Figures corresponding to the other source/receiver positions displayed in Figures 5.3 and 5.4. It can be noticed that in the third source/receiver position, the hybrid model provides much more accurate results in comparison with the real measurement, regardless of its crossover frequency value. This is probably due to the loudspeaker and microphone positions, since there is very little space between them and the walls and they are placed on an axis perpendicular to these walls (see Table 5.1). In that case, the early rays are strongly affected by the physical

55

presence of the loudspeaker and microphone.
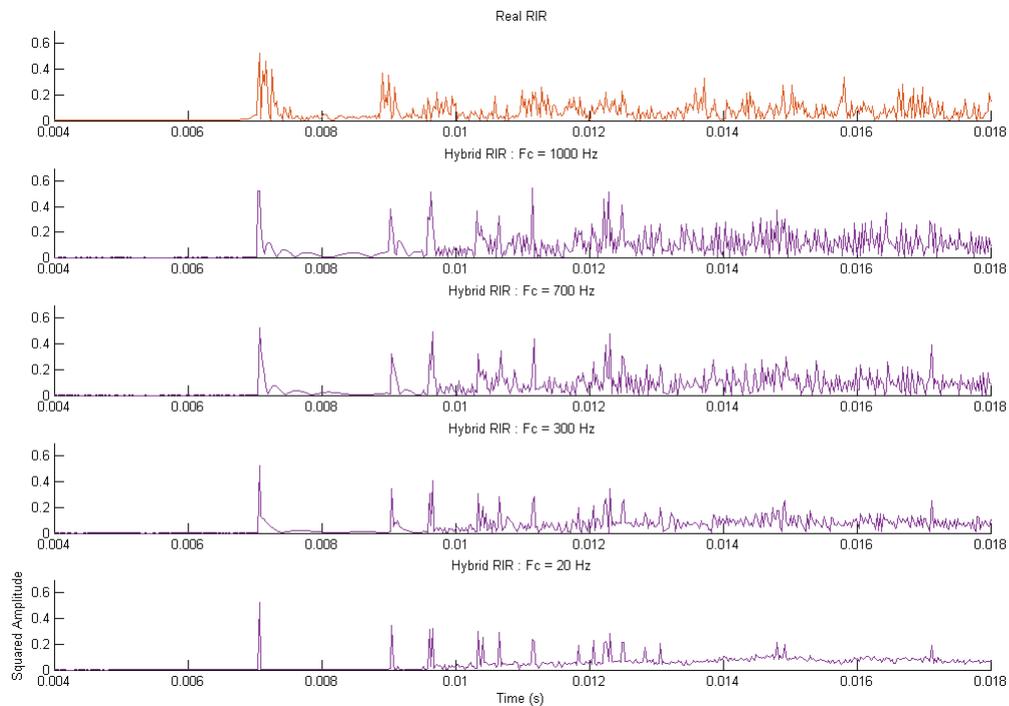
Figure 5.3: Case Study 1 : Real and hybrid impulse responses (early rays) in the Time Domain at Source/Receiver position 2. Crossover frequency values are 20 Hz, 300 Hz, 700 Hz, and 1000 Hz.

It can be shown from each of these figures that a change of the crossover frequency modifies distinguishably the shape of the peaks. This is due to the low frequency components that are better represented in the numerical model than the geometric one. This is the case with a higher crossover frequency, like 1kHz for example.

Whilst the shape of the peaks is modified with the crossover frequency, their position is not. Indeed, the peak apex of each distinguishable peak corresponding to the direct sound and the early rays is at the exact same time position. For purposes of auralisation, the time position of the direct sound and early rays is important. As it appears in each of these figures, it can be noticed that reducing the crossover frequency down to 20 Hz does not affect the time position of the main pressure peaks, especially for the early rays.

On the contrary, it visually looks like reducing the crossover frequency improves the accuracy of the global shape of the RIR, especially for source/receiver cases 2 and 3. Indeed, for hybrid
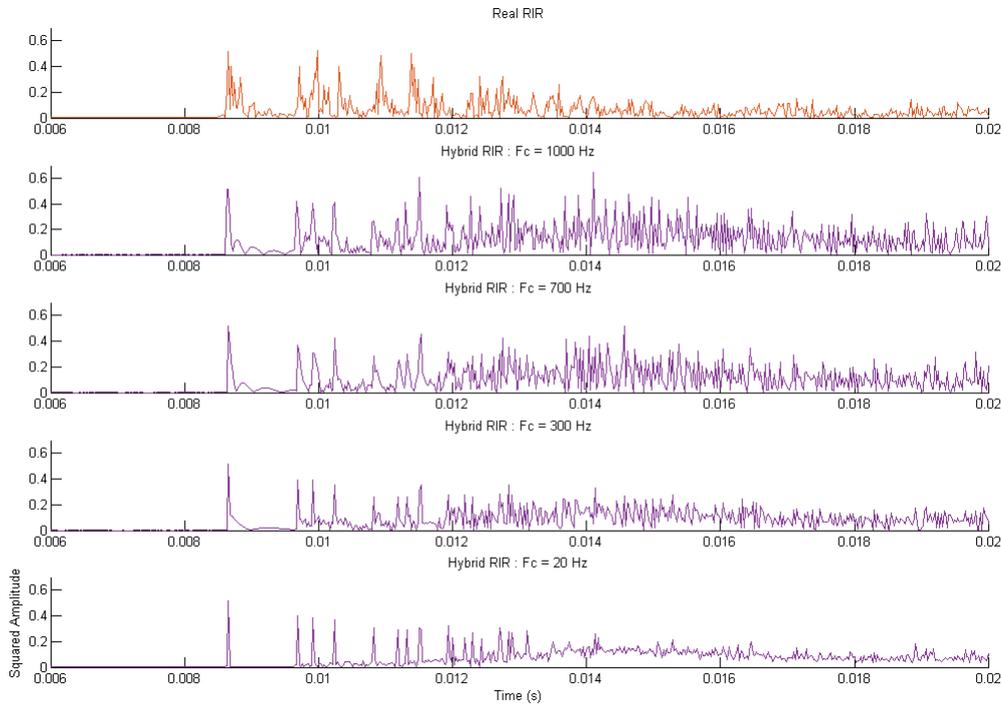
Figure 5.4: Case Study 1 : Real and hybrid impulse responses (early rays) in the Time Domain at Source/Receiver position 3. Crossover frequency values are 20 Hz, 300 Hz, 700 Hz, and 1000 Hz.

models with a crossover frequency higher than 30 Hz, the amplitude of later rays is much higher than in the real measurements, which will have an direct impact on the decay rate.

These results could be confirmed by objective measurements like *Mean Squared Error*, *Pearson Coefficient*, or *Magnitude Index* [29], but these are not relevant due to too much disparity between the signals. Indeed, as the position of the rays in the real response do not coincide exactly with those in the models, it is complicated to estimate via a numerical comparison tool the similarity between them.

5.1.2 Acoustic Parameter Analysis

As mentioned in Section 2.3, it is possible to extract acoustic parameters, providing objective characteristics of the room. The calculation of T_{30}, C_{80} and EDT are therefore good indicators of the quality of the model. These parameters are usually measured in octave frequency bands, since the absorption coefficients are frequency dependent, which is the case in this study. It means that the analysed RIR is filtered with the cut-off frequencies displayed on Table 5.2.

57

Table 5.2: *Center, lower and upper frequencies for standard set of octave bands under 1 kHz*

Lower Band Limit (Hz)	Center Frequency (Hz	Upper Band Limit (Hz)
11	16	22
22	31.5	44
44	63	88
88	125	177
177	250	355
355	500	710
710	1000	1420

Acoustic parameters are extracted from hybrid RIRs and from the real measured RIR for each octave-band. As hybrid RIRs are generated from the same numerical and geometrical models, the set of acoustic parameters is the same among hybrid RIRs with similar crossover frequencies. To that end, only the acoustic parameters calculated from the octave band containing the crossover frequency of the hybrid RIR will be considered. It will then be compared with the value extracted from the real RIR on the same octave band in terms of *absolute error ε* or *relative error η*, as defined in (5.1) and (5.2) where θ is the acoustic parameter value.

$$\varepsilon(\theta) = \left| \theta_{\text{real}} - \theta_{\text{hybrid}} \right| \tag{5.1}$$

$$\eta(\theta) = \frac{\left| \theta_{\text{real}} - \theta_{\text{hybrid}} \right|}{\theta_{\text{real}}} \tag{5.2}$$

The error between the real value and the one extracted from the hybrid model should be less than 1 *Just Noticeable Difference* (JND), which corresponds to the extension of the value without being detected by the human auditory system.

Reverberation times

The reverberation time indicates how fast reflection energy decays. T_{30} is defined as the energy decay from -5 to -35 dB and is related to the physical properties of the room. 1 JND for T_{30} is currently stated as 5 % from previous extensive work [36].

Figure 5.5 shows the T_{30} values for the hybrid model and the real measurement and Figure 5.6 shows the relative error between the two values for each source/receiver position, using the octave-band process described previously. It can be noticed that the error values are all very much above the JND, 64.78% being the smallest value for position 1. This is due to
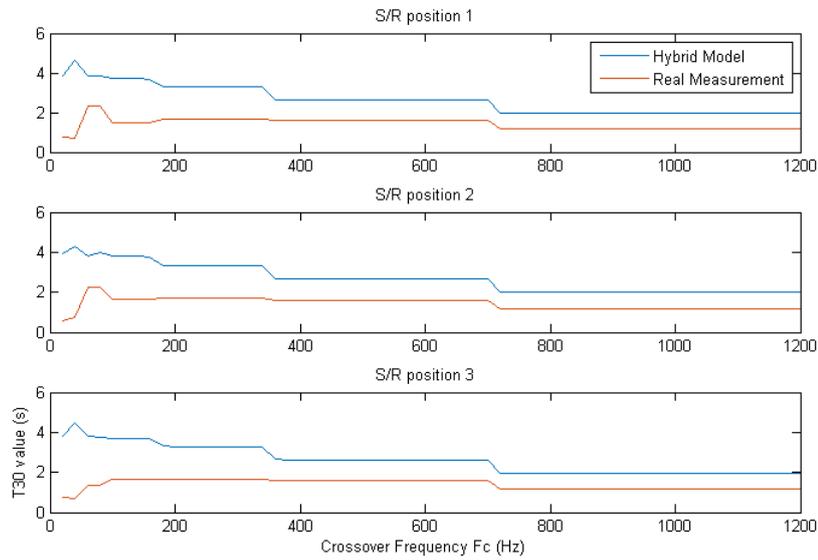
Figure 5.5: Case Study 1 : T_{30} values from hybrid and real responses calculated at each crossover frequency value for each source/receiver (S/R) position. Each value was extracted for the octave band containing the crossover frequency.
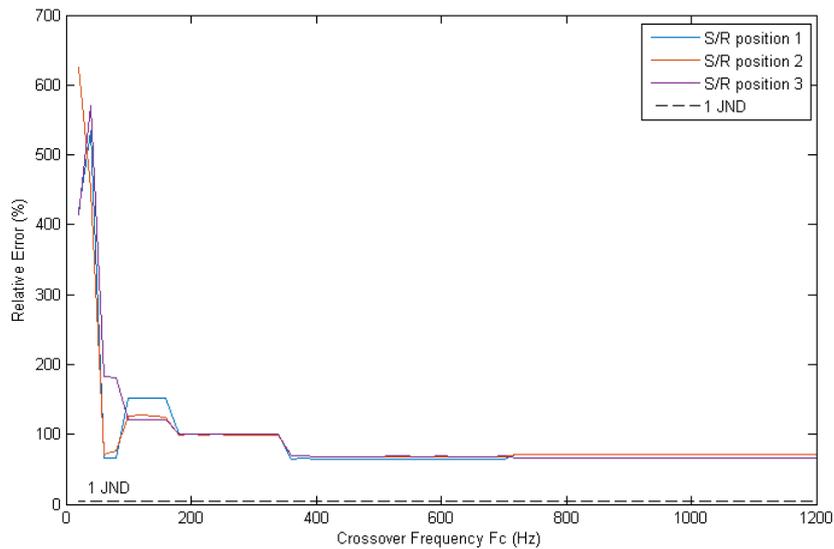
Figure 5.6: Case Study 1 : Relative errors between T_{30} values from hybrid and real responses calculated at each crossover frequency value for each source/receiver (S/R) position. Each value was extracted for the octave band containing the crossover frequency.

59

the quality of the model, which could still be improved by optimizing the geometry of the room and the absorption coefficients. A calibration process was achieved in Section 4.3 to bring the reverberation parameters (T_{30}) to 1 JND of the measured value. Therefore, it is very likely that the gap between the hybrid model and the real response, comes from the quality of the FDTD model.

For each source/receiver position, it is observed on Figure 5.6 that the relative error starts to strongly increase when the crossover frequency is inferior to 360 Hz. It could be due to the T_{30} values increasing with the reducing crossover frequencies. However, it can be seen on Figure 5.5 that it is not the case, since the T_{30} values diminish and the absolute error with the hybrid model decreases.

Figure 5.6 shows that when the crossover frequency is superior to 360 Hz, the error is approximately the same, meaning that the crossover frequency can be reduced down to that value without compromising the T_{30} value. The same trend emerges for each source/receiver position. This confirms the hypothesis stating that the crossover frequency could be reduced to 373 Hz.

According to ISO 3382 [3], Early Decay Time (EDT) is impacted by early reflections and is perceptually more relevant than T_{30} since it rather affects higher frequencies. 1 JND for EDT is currently stated as 5 % [36]. Considering Figures 5.7 and 5.8 showing EDT values and relative errors, the same observations than for T_{30} can be done, resulting in the reduction of the crossover frequency to 360 Hz.
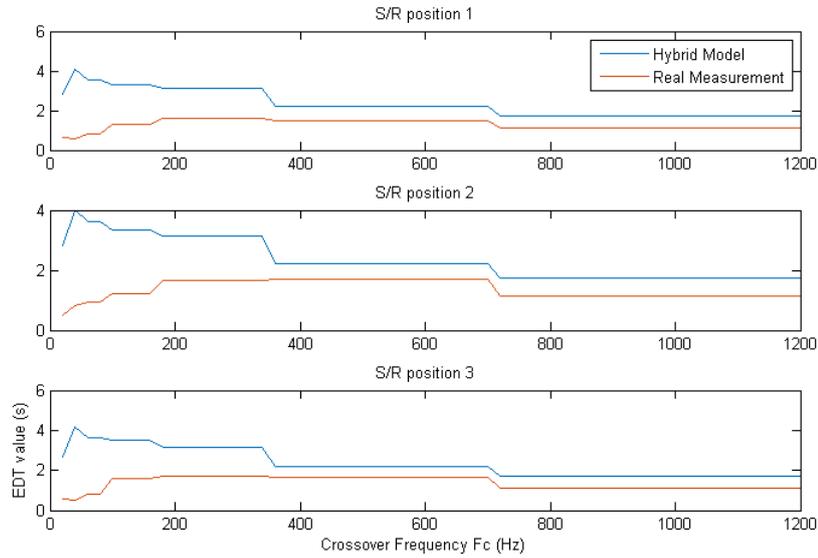
Figure 5.7: Case Study 1 : Early Decay Time (EDT) values from hybrid and real responses calculated at each crossover frequency value for each source/receiver (S/R) position. Each value was extracted for the octave band containing the crossover frequency.
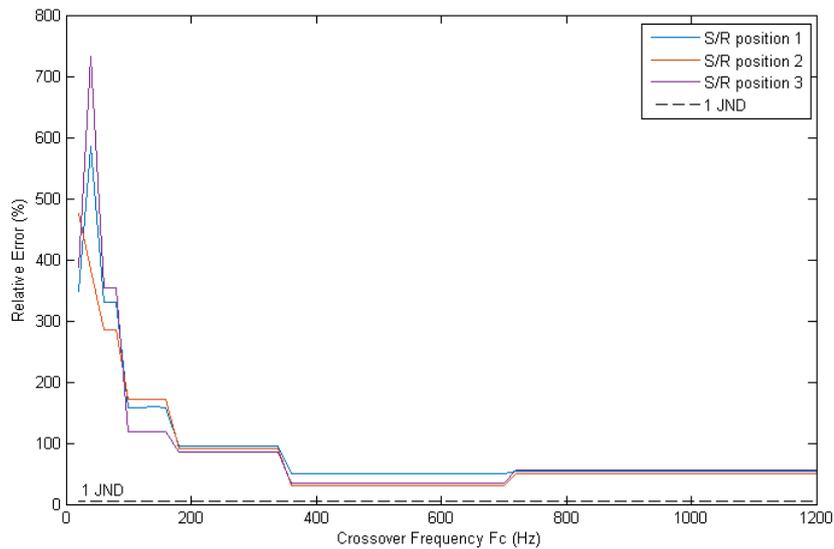
Figure 5.8: Case Study 1 : Relative errors between EDT values from hybrid and real responses calculated at each crossover frequency value for each source/receiver (S/R) position. Each value was extracted for the octave band containing the crossover frequency.

61

Energy ratios

According to ISO 3382 [3], clarity is the ratio of energy in early reflections to that in the late reverberations, expressed in decibels, and describes the perceived definition of the sound. 1 JND of C_{80} is reported as 1 dB [36].
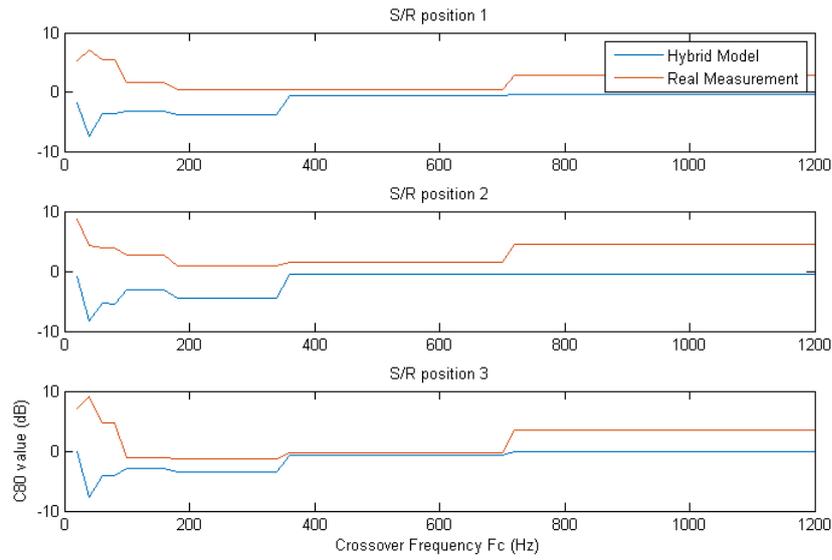
Figure 5.9: Case Study 1 : C_{80} values from hybrid and real responses calculated at each crossover frequency value for each source/receiver (S/R) position. Each value was extracted for the octave band containing the crossover frequency.

Figure 5.9 shows the C_{80} values for the hybrid model and the real measurement and Figure 5.10 shows the absolute error between the two values for each source/receiver position, using the octave-band process described above. It can be noticed that the error values are all above the JND, 2 dB being the smallest value. This is due to the quality of the model, which could still be improved by optimizing the geometry of the room and the absorption coefficients.

Figure 5.10 shows that the frequency region for which the absolute error is the lowest is between 360 Hz and 700 Hz, meaning that the crossover frequency could be reduced down to 360 Hz without compromising the C_{80} value. The same trend emerges for each source/receiver position. This confirms the hypothesis stating that the crossover frequency could be reduced to 373 Hz.
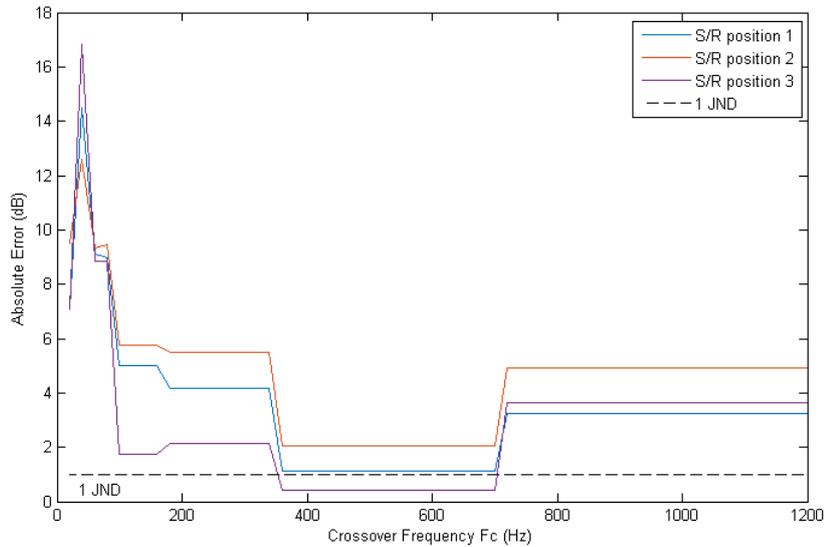
Figure 5.10: Case Study 1 : Absolute errors between C_{80} values from hybrid and real responses calculated at each crossover frequency value for each source/receiver (S/R) position. Each value was extracted for the octave band containing the crossover frequency.

5.1.3 Frequency Response Analysis

The magnitude spectra of each model is calculated via the Fourier transform after applying a Hanning window on the RIR as an apodization function. At first, only the frequency responses of the FDTD and geometric models will be compared with the real frequency response for a better understanding.

For the first source/receiver position, Figure 5.11 shows the magnitude in decibels of the frequency responses of the FDTD model, the geometric model and the real measurement. Theoretical room modes are calculated as described in Section 2.2.6 but only the axial (black) and tangential modes (grey) are displayed since they have more impact on the frequency response than oblique modes.

In Section 2.2.7, it has been mentioned that the sound is fully subject to the behaviour of the room modes in the frequency region between f_{cutoff} and $f_{critical}$ (also known as the Schroeder frequency), which values are respectively $52.3Hz$ and $1185Hz$ in the case of this first room. This means that the acoustic model can only be valid if it preserves the room modes between these two values.

On Figure 5.11, it can be observed that theoretical room modes coincide precisely with resonant frequencies from the FDTD model and the real measurement at low frequency, which is not

63
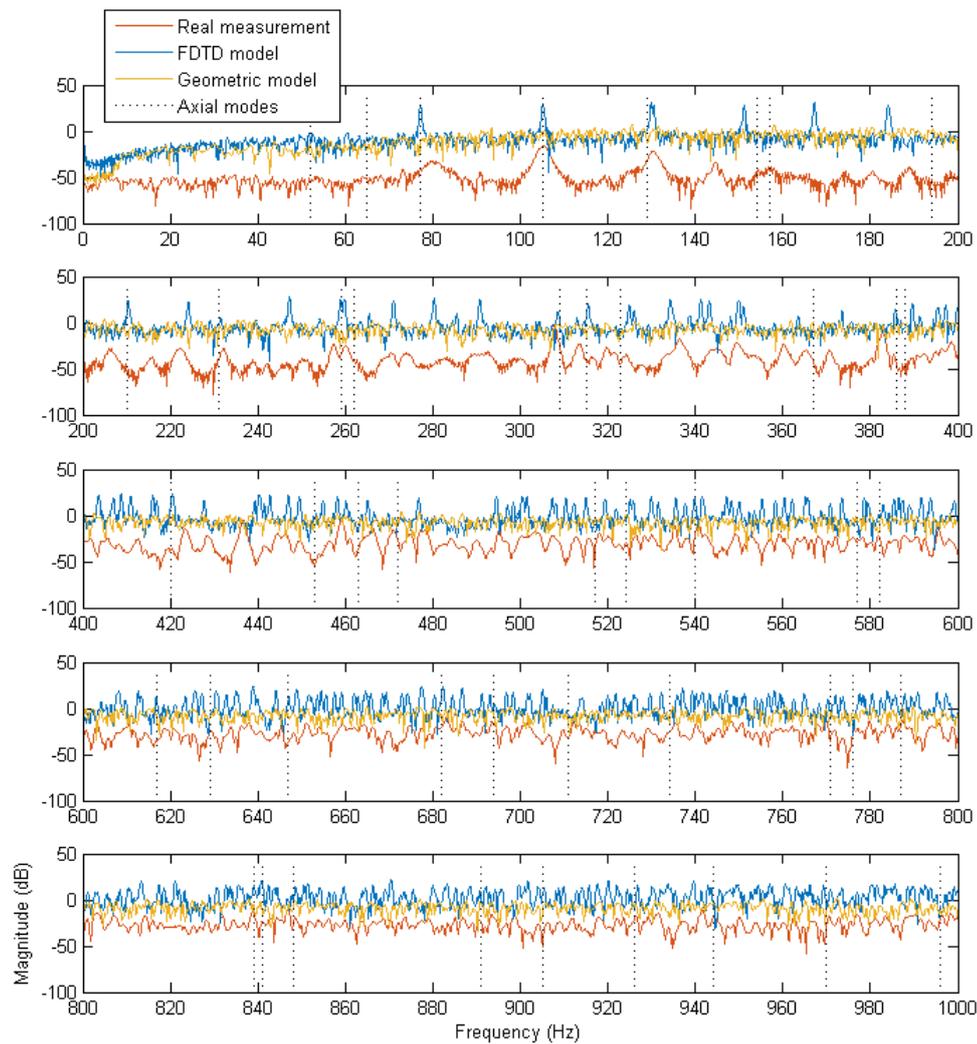
Figure 5.11: Case Study 1 : Frequency spectra of the real, FDTD and geometric RIRs at source/receiver (S/R) position 1. Theoretical axial and tangential room modes are also shown by the vertical dashed lines.

the case with the geometric model. It can be noticed that the first two modes do not appear for each model, which is due to the DC blocking filter (as mentioned in Section 4.5) used to remove the drifting effect. However, the room modes correctly coincide with the theoretical and real ones from 70 Hz to at least 320 Hz. It must be mentioned that the tangential modes have not been displayed for a better visibility and correspond to the other frequency peaks.

This confirms that the numerical model is much more precise at low frequencies than the geometric one to preserve room modes. However, for frequencies higher than 320 Hz, room modes

do not appear clearly on the real response. This could mean that the crossover frequency could be reduced to approximately 320 Hz but this needs to be confirmed by other results in the frequency domain.
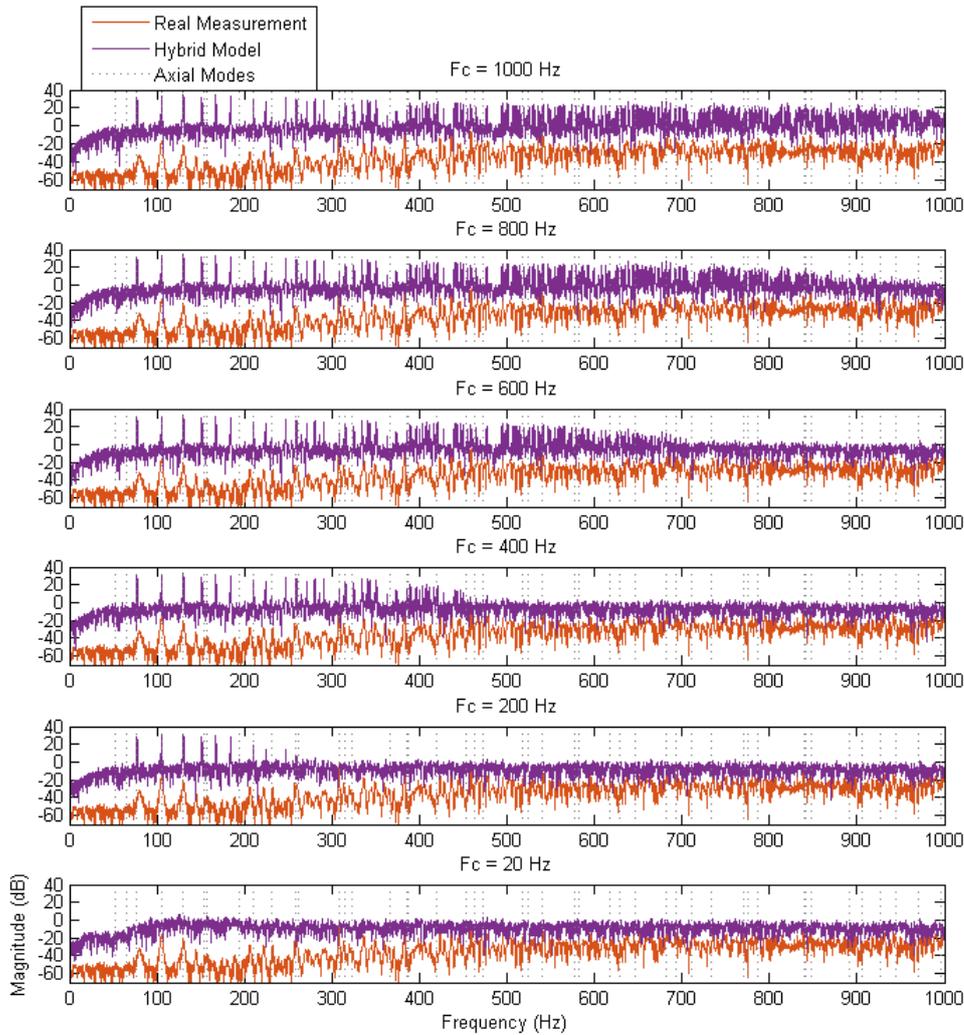
Figure 5.12: Case Study 1 : Frequency spectra of the real and hybrid RIRs at source/receiver (S/R) position 1 for several crossover frequency values (Fc = 20 Hz, 200 Hz, 400 Hz, 600 Hz, 800 Hz, and 1000 Hz). Theoretical axial room modes are also shown by the vertical dashed lines.

Figure 5.12 shows frequency responses at the first source/receiver position for the real measurement and hybrid models with a crossover frequency being 20 Hz, 200 Hz, 400 Hz, 600 Hz, 800 Hz, and 1000 Hz. It can be noticed that, for each hybrid model, resonant frequencies do not appear clearly on frequencies superior to the crossover frequency, due to the predominance of

the geometric model. This means that the crossover frequency could be reduced to the highest frequency for which room modes are indistinguishable in the real measurement. However, it is not possible to define such a limit visually speaking.

These results could be confirmed by objective measurements like *Mean Squared Error*, *Frequency Response Assurance Criterion*, or *Magnitude Index* [29], but these are not relevant due to the magnitude offset between the numerical frequency response and the real one as a result of the calibration, leading to a major increase of disparity between the hybrid model and the real one as the crossover frequency diminishes. An attempt to list all the peaks corresponding to the room modes and comparing them in terms of amplitude and frequency has also been performed, but has not led to any concrete result. The use of an average filter before doing the comparisons has also been tested.

In terms of analysis in the frequency domain, no objective criterion can be used to validate a minimum crossover frequency. However, visually speaking, the crossover frequency could be reduced to a value between 300 and 500 Hz without compromising the quality of the model, but this is very approximate. Similar results are obtained with source/receiver positions 2 and 3 with Figures 5.13, 5.14, 5.15, 5.16 displayed below.
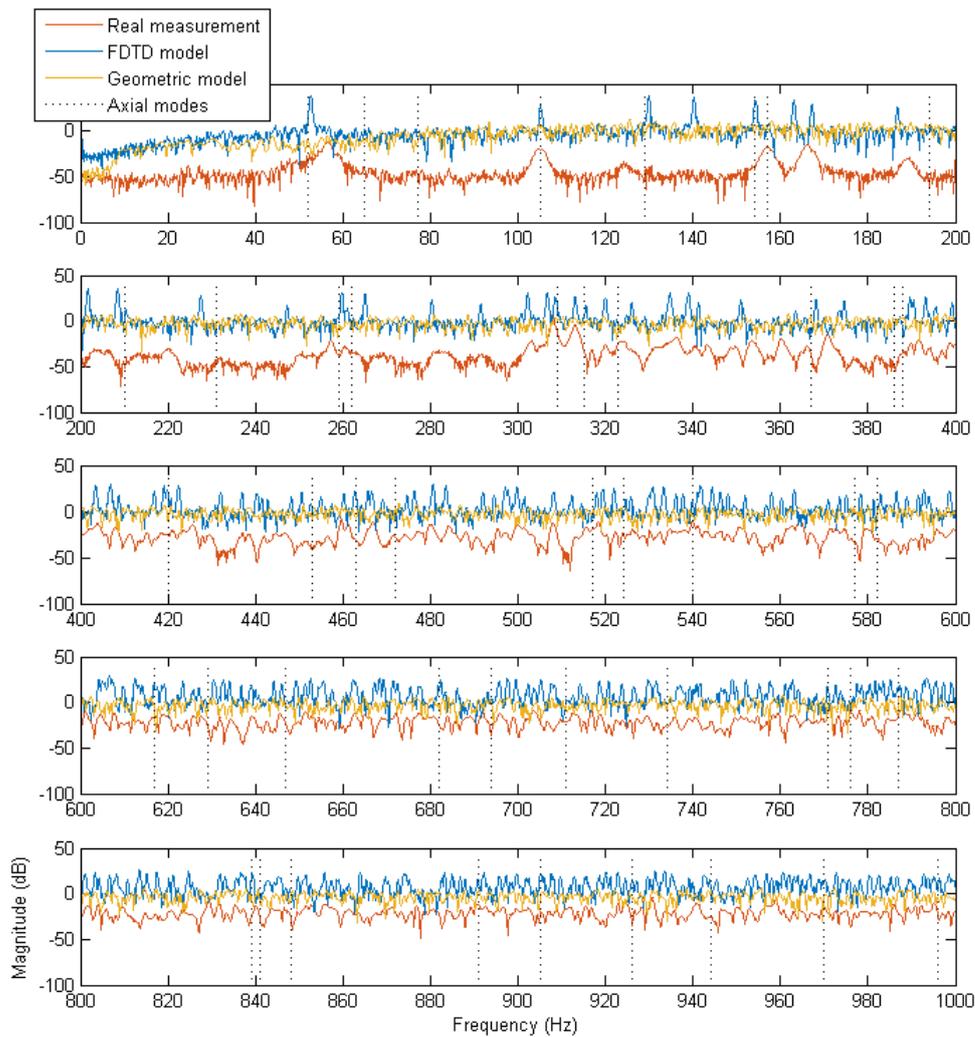
Figure 5.13: Case Study 1 : Frequency spectra of the real, FDTD and geometric RIRs at source/receiver (S/R) position 2. Theoretical axial and tangential room modes are also shown by the vertical dashed lines.
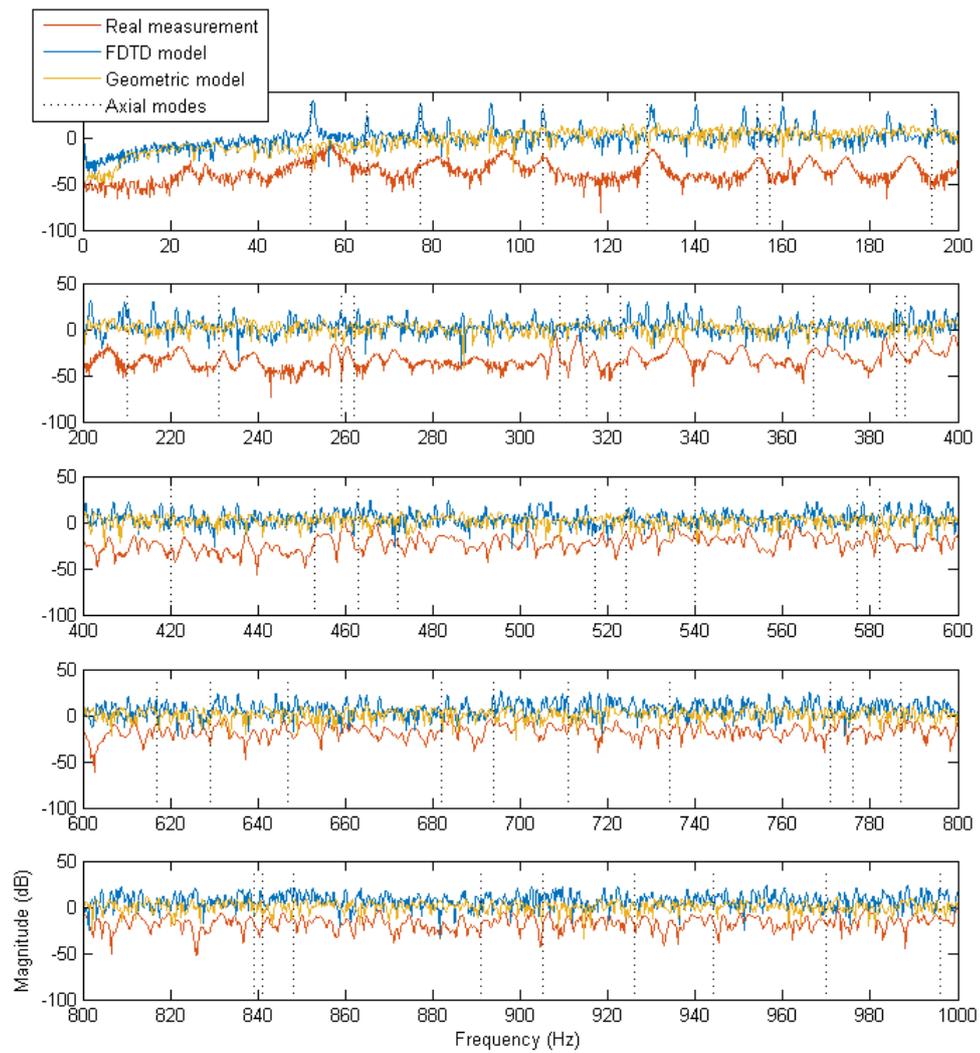
Figure 5.14: Case Study 1 : Frequency spectra of the real, FDTD and geometric RIRs at source/receiver (S/R) position 3. Theoretical axial and tangential room modes are also shown by the vertical dashed lines.
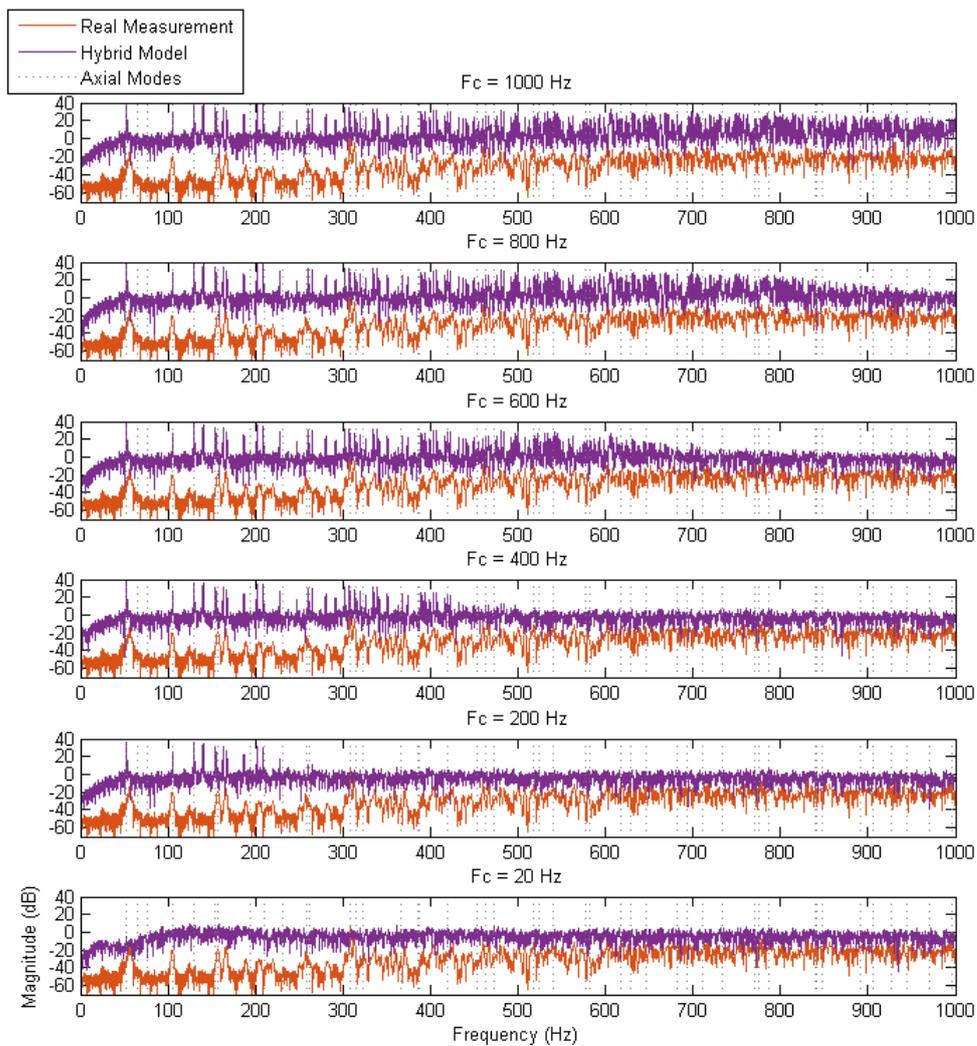
Figure 5.15: Case Study 1 : Frequency spectra of the real and hybrid RIRs at source/receiver (S/R) position 2 for several crossover frequency values (Fc = 20 Hz, 200 Hz, 400 Hz, 600 Hz, 800 Hz, and 1000 Hz). Theoretical axial room modes are also shown by the vertical dashed lines.
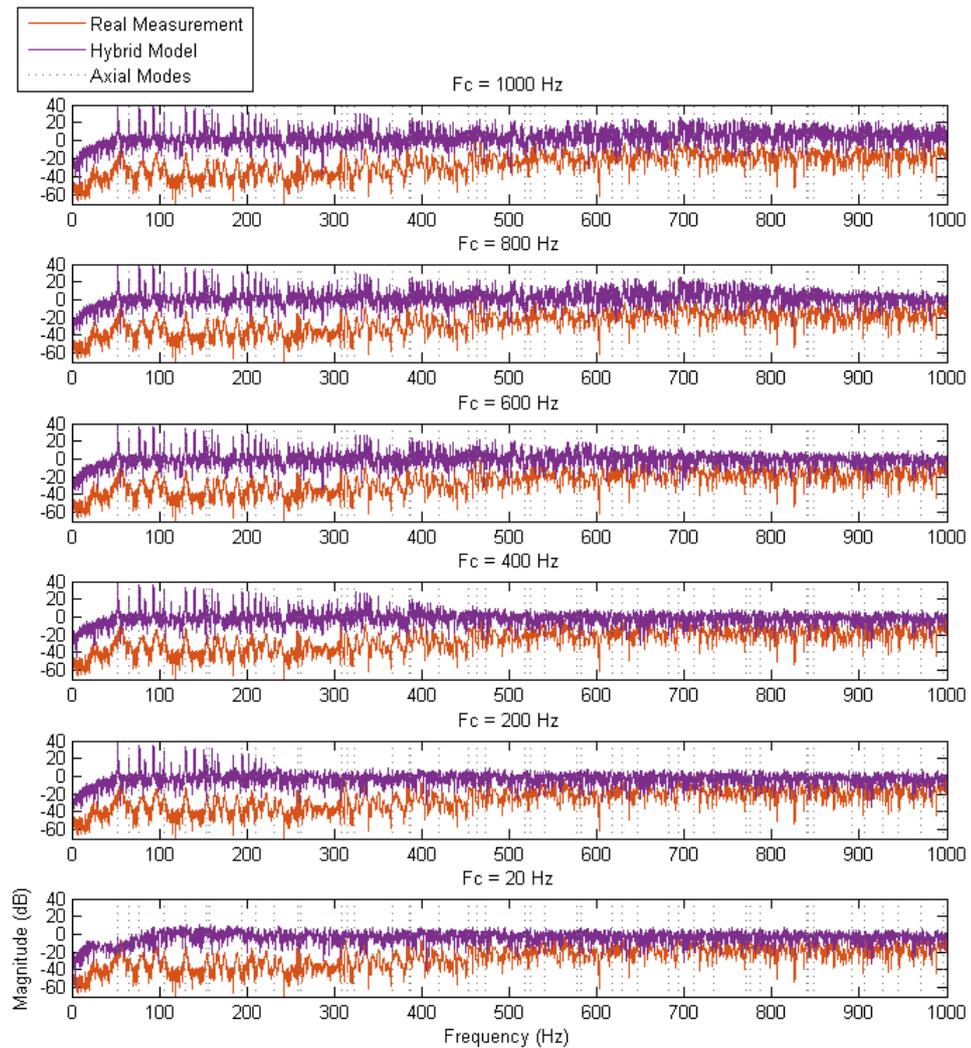
Figure 5.16: Case Study 1 : Frequency spectra of the real and hybrid RIRs at source/receiver (S/R) position 3 for several crossover frequency values (Fc = 20 Hz, 200 Hz, 400 Hz, 600 Hz, 800 Hz, and 1000 Hz). Theoretical axial room modes are also shown by the vertical dashed lines.

5.2 Case Study 2

The second room is in the shape of a $(3.61 \times 3.10 \times 2.66)m$ cube, with a chimney area, a door, a window area. It is a bit more complex than the previous one due to the chimney and the window frame, as represented on Figure 5.17. The source and receiver positions are displayed on Table 5.3.
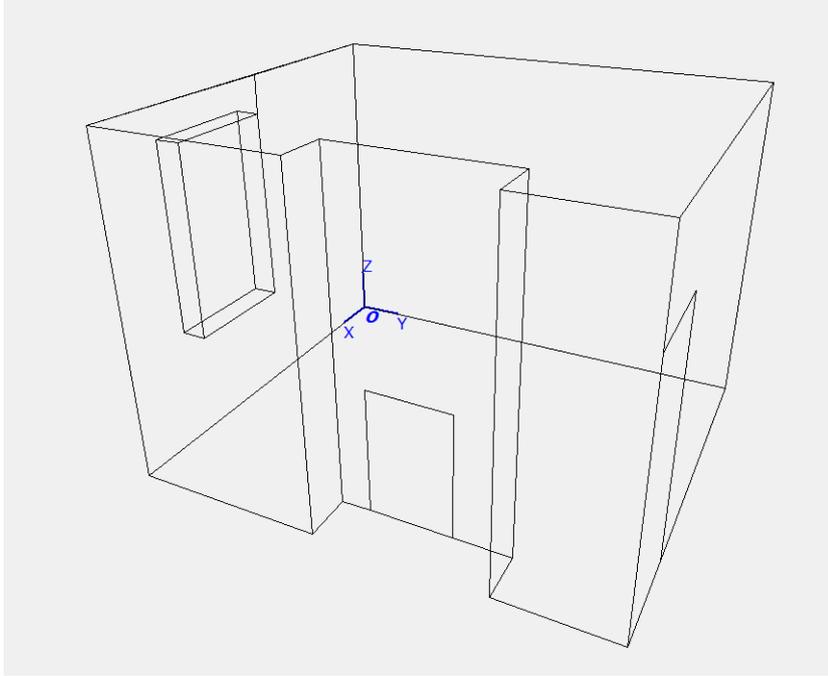
Figure 5.17: Perspective view of Room 2.

Table 5.3: *Source and receiver placements for Case Study 2*

	Source (x,y,z) (m)	Receiver (x,y,z) (m)
1	(1.55, 0.36, 1.25)	(1.55, 3.29, 1.25)
2	(2.66, 0.42, 0.98)	(0.28, 3.28, 1.49)
3	(0.40, 1.80, 1.25)	(2.33, 1.80, 1.25)

Regarding the room dimensions, it can be stated that the minimum wavelength should be 0.64 m which is the chimney width, meaning that considering the sound velocity as 343 m.s^{-1}, the crossover frequency could be reduced down to 536 Hz. As mentioned in Section 2.2, the theoretical Schroeder frequency can be calculated from the room dimensions. For case study 2, its value is $f_{Schroeder} = 1608Hz$.

71

5.2.1 Analysis in the Time Domain

As for Case Study 1, a time domain analysis is conducted at first, with a visualisation of the RIRs. They are displayed for a wide range (20 Hz, 300 Hz, 700 Hz, and 1000 Hz) of crossover frequencies on Figure 5.18 for the first source/receiver position. The amplitude of the RIR has been squared for a better visualisation. It can be noticed that, due to the complexity of the room, the early rays are more numerous than in the previous case study.
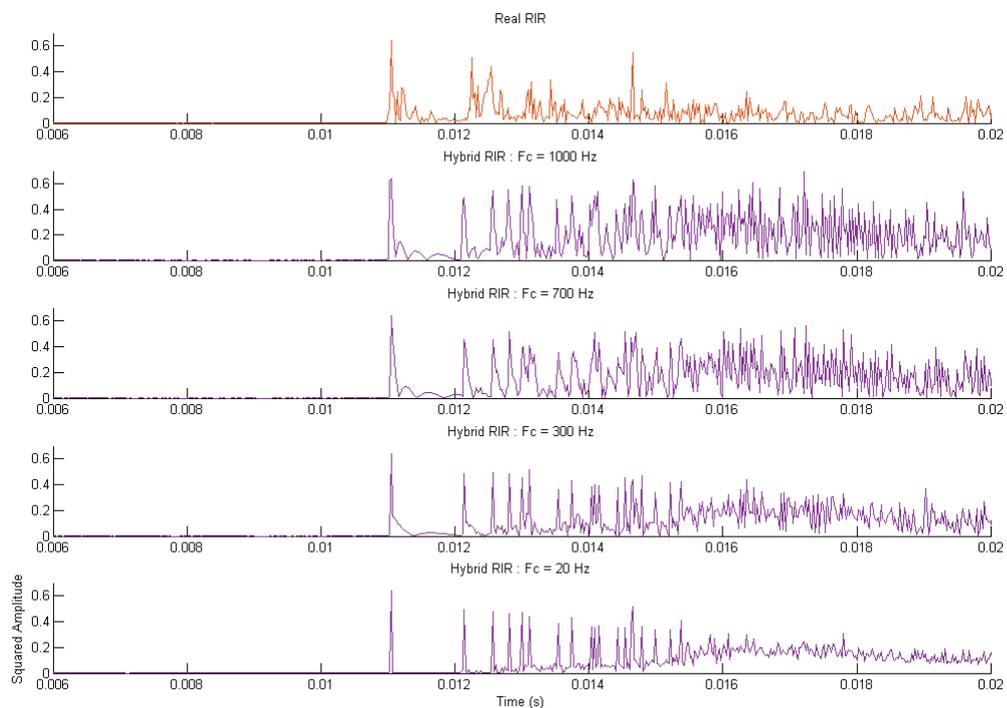
Figure 5.18: Case Study 2 : Real and hybrid impulse responses (early rays) in the Time Domain at Source/Receiver position 1. Crossover frequency values are 20 Hz, 300 Hz, 700 Hz, and 1000 Hz.

The Figures corresponding to the other source/receiver positions are displayed in Figures 5.19 and 5.20.

Similarly to case study 1, a change of the crossover frequency modifies distinguishably the shape of the peaks, due to the low frequency components that are better represented in the numerical model than the geometric one.

Comparing the cases of a 1000 Hz and a 20 Hz crossover frequency, it can be observed on
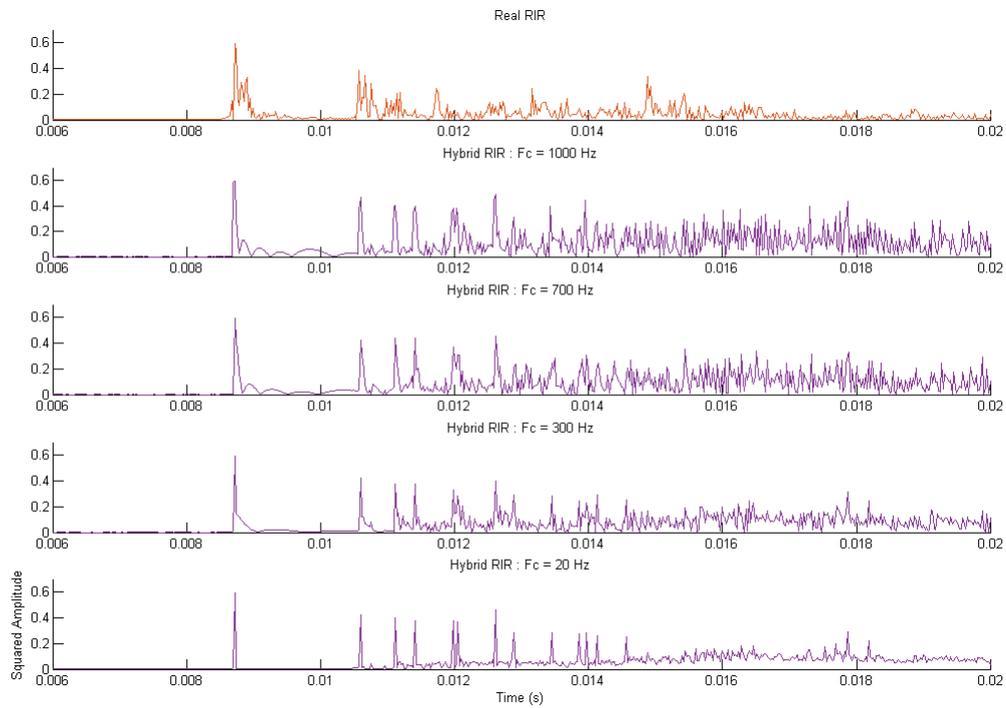
Figure 5.19: Case Study 2 : Real and hybrid impulse responses (early rays) in the Time Domain at Source/Receiver position 2. Crossover frequency values are 20 Hz, 300 Hz, 700 Hz, and 1000 Hz.

each figure that there is no significant dissimilarity within the first 0.005 s (starting from the direct sound). However, the late reverberation (starting from 0.016 s on Figure 5.18) is further attenuated at lower crossover frequencies. This global shape of the hybrid RIR corresponds better to the real measurements for crossover frequencies of 300 Hz and lower.
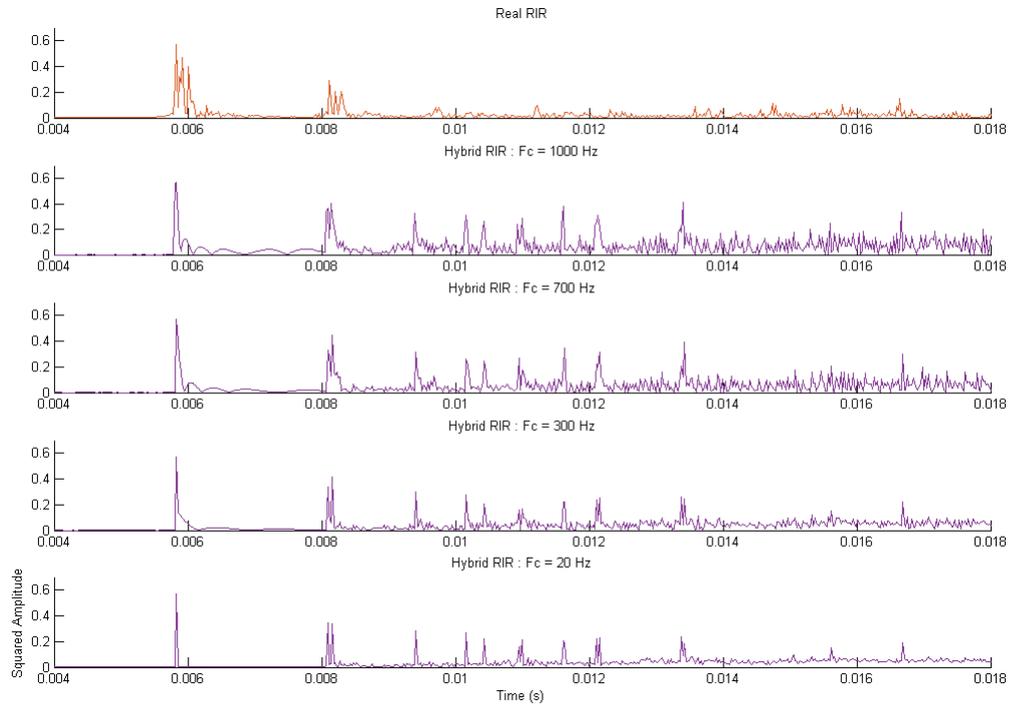
Figure 5.20: Case Study 2 : Real and hybrid impulse responses (early rays) in the Time Domain at Source/Receiver position 3. Crossover frequency values are 20 Hz, 300 Hz, 700 Hz, and 1000 Hz.

5.2.2 Acoustic Parameter Analysis

As in case study 1, acoustic parameters like T_{30}, C_{80} and EDT, are extracted from the hybrid RIRs to provide objective characteristics of the room. They are calculated and compared in terms of octave bands, as described in Subsection 5.1.2. Acoustic parameters are compared with the values extracted from the real RIR on the same octave band in terms of *absolute error* or *relative error*, as described in 5.1.2.

Reverberation times

Figure 5.21 shows the T_{30} values for the hybrid model and the real measurement and Figure 5.22 shows the relative error between the two values for each source/receiver position, using the octave-band process described previously. It can be noticed that the error values are much smaller than in case study 1, which, combined with the results in the time domain, shows that the model in this case study is of better quality.

For each source/receiver position, it is observed on Figure 5.22 that, as the crossover fre-
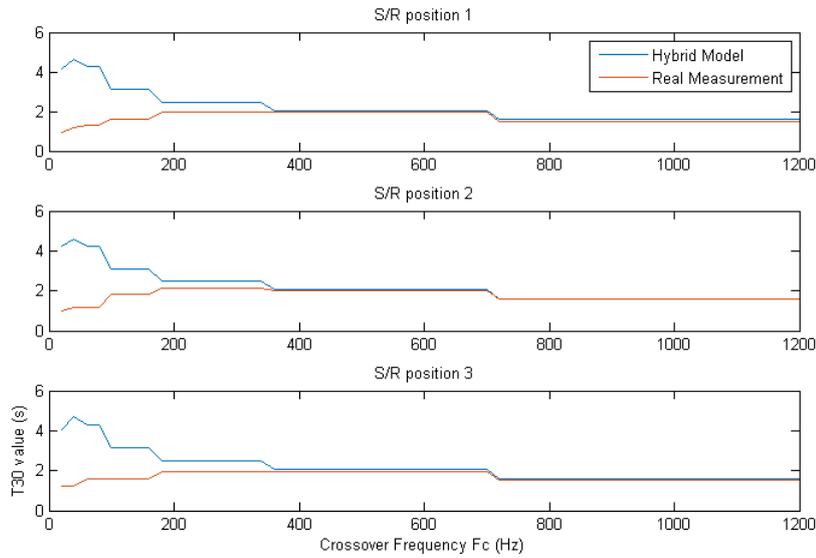
Figure 5.21: Case Study 2 : T_{30} values from hybrid and real responses calculated at each crossover frequency value for each source/receiver (S/R) position. Each value was extracted for the octave band containing the crossover frequency.

quency increases, the relative error between the T_{30} values of the hybrid model and the real measurement diminishes down to approximately 4% from 360 Hz. As 1 JND is stated as 5% [36], it means that the model is valid in terms of T_{30} values with a crossover frequency of 360 Hz or higher.

The same trend emerges from Early Decay Time values (see Figures 5.23 and 5.24) as relative errors are very slightly above 1 JND (5%) when the crossover frequency is between 360 and 700 Hz, and under this threshold when it gets higher. This means that the model is valid in term of EDT values when the crossover frequency is worth at least 700 Hz, but it could also be acceptable between 360 and 700 Hz.
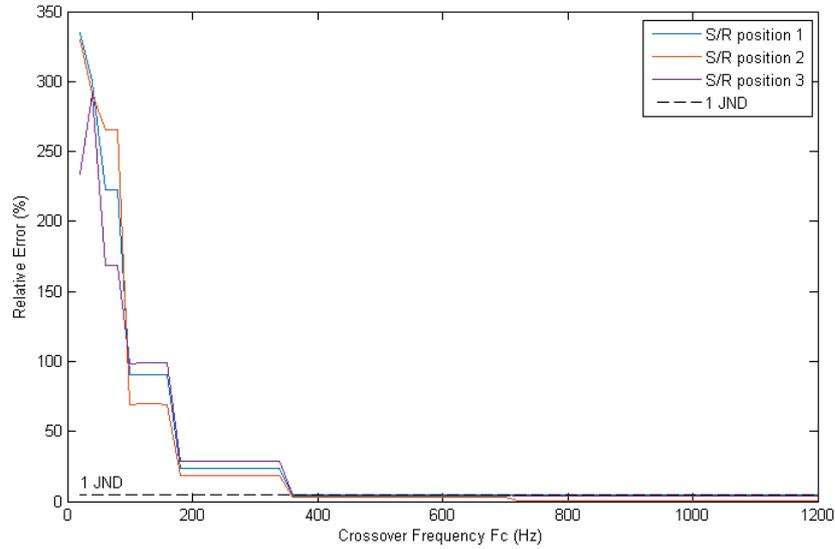
Figure 5.22: Case Study 2 : Relative errors between T_{30} values from hybrid and real responses calculated at each crossover frequency value for each source/receiver (S/R) position. Each value was extracted for the octave band containing the crossover frequency.
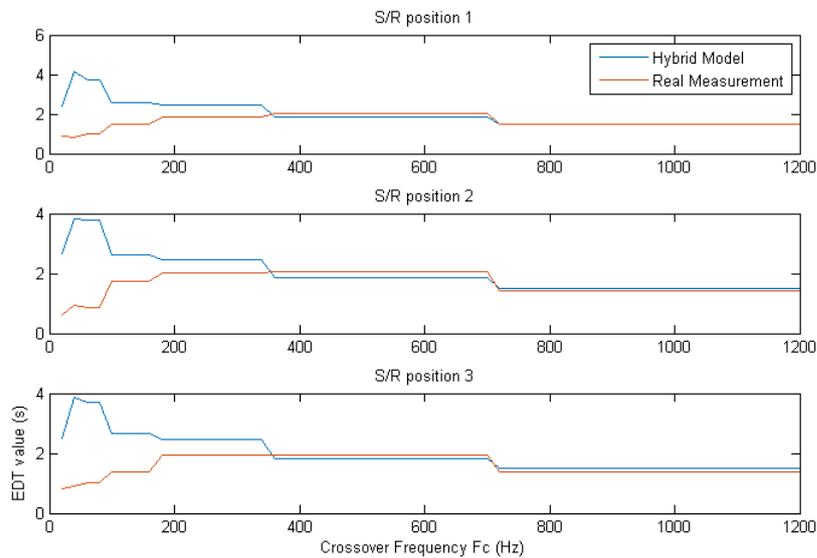
Figure 5.23: Case Study 2 : Early Decay Time (EDT) values from hybrid and real responses calculated at each crossover frequency value for each source/receiver (S/R) position. Each value was extracted for the octave band containing the crossover frequency.
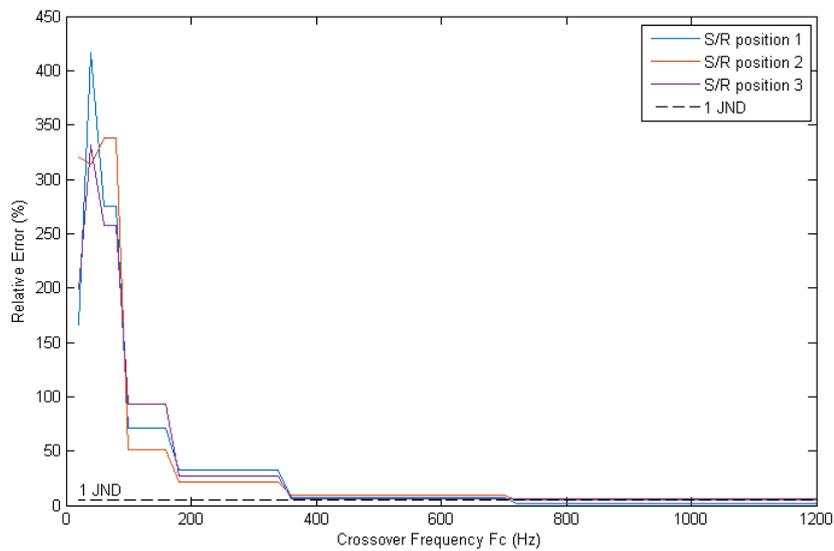
Figure 5.24: Case Study 2 : Relative errors between EDT values from hybrid and real responses calculated at each crossover frequency value for each source/receiver (S/R) position. Each value was extracted for the octave band containing the crossover frequency.
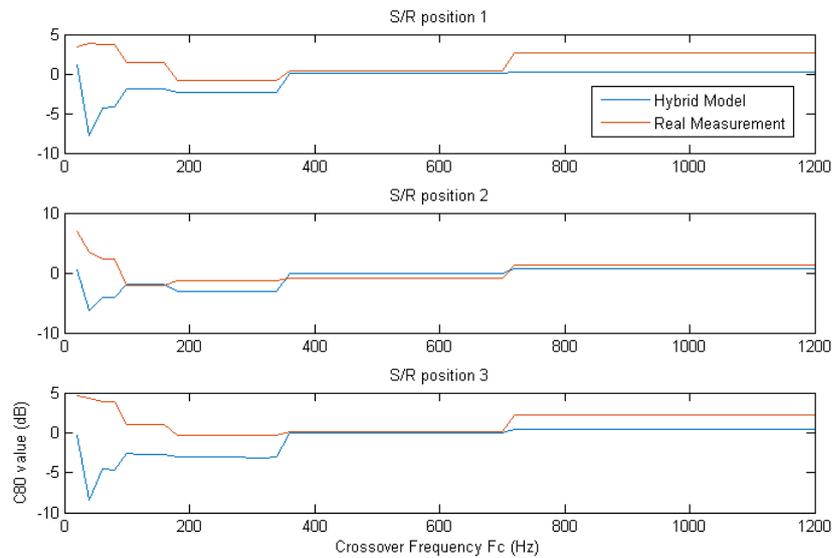
Figure 5.25: Case Study 2 : C_{80} values from hybrid and real responses calculated at each crossover frequency value for each source/receiver (S/R) position. Each value was extracted for the octave band containing the crossover frequency.

Energy ratios

Figure 5.25 shows the C_{80} values for the hybrid model and the real measurement and Figure 5.26 shows the absolute error between the two values for each source/receiver position, using the octave-band process described above. It can be noticed that the error values are all smaller than in case study 1, showing again that this modelling is of better quality.

Figure 5.26 shows that when the crossover frequency is between 360 Hz and 700 Hz, the error gets under 1 JND for each source/receiver position, which corresponds to 1 dB [36]. It can be noticed that the error increases as the crossover frequency gets higher than 700 Hz for source/receiver positions 1 and 3, which is due to the evolution of C_{80} values on Figure 5.25. This means that considering C_{80} values, the model is only valid for crossover frequencies between 360 and 700 Hz. However, it is still acceptable for crossover frequencies higher than 700 Hz.
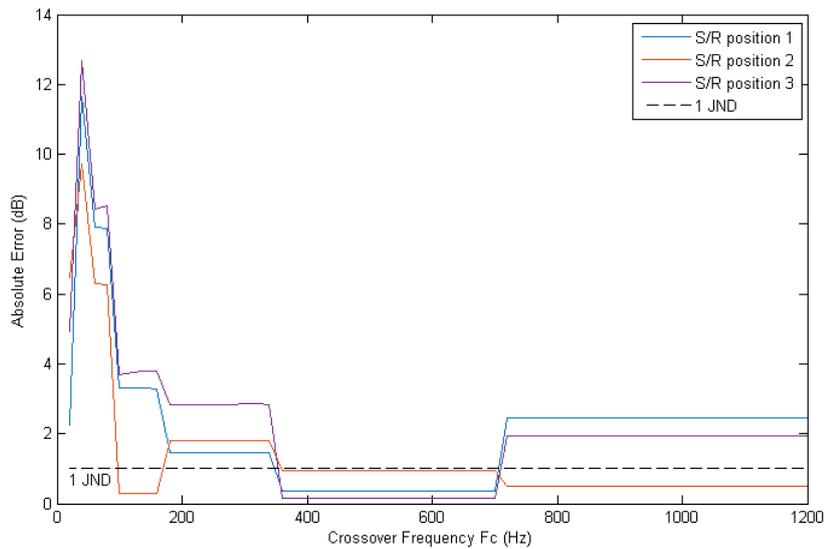
Figure 5.26: Case Study 2 : Absolute errors between C$_{80}$ values from hybrid and real responses calculated at each crossover frequency value for each source/receiver (S/R) position. Each value was extracted for the octave band containing the crossover frequency.

5.2.3 Frequency Response Analysis

As for study case 1, the magnitude spectra of each model is calculated via the Fourier transform after applying a Hanning window on the RIR as an apodization function.

For the first source/receiver position, Figure 5.27 shows the magnitude in decibels of the frequency responses of the FDTD model, the geometric model and the real measurement. Theoretical room modes are calculated as described in Section 2.2.6 but only the axial and tangential modes are displayed since they have more impact on the frequency response than oblique modes.

In Section 2.2.7, it has been mentioned that the sound is fully subject to the behaviour of the room modes in the frequency region between f_{cutoff} and $f_{critical}$ (also known as the Schroeder frequency), which values are respectively $47.5Hz$ and $1608Hz$ in the case of this second room. This means that the acoustic model can only be valid if it preserves the room modes between these two values.

It can be observed on Figure 5.27 that room modes coincide between the real and the FDTD RIR between approximately 40 and 550 Hz. However, room modes from the FDTD model can be slightly shifted from the theoretical ones, which is due to dispersion error that increases with increasing frequencies. Also, a room mode can be noticed on the FDTD model at 5.7 Hz, which could be due to an error in the simulation or a remaining of the drifting effects due to the use
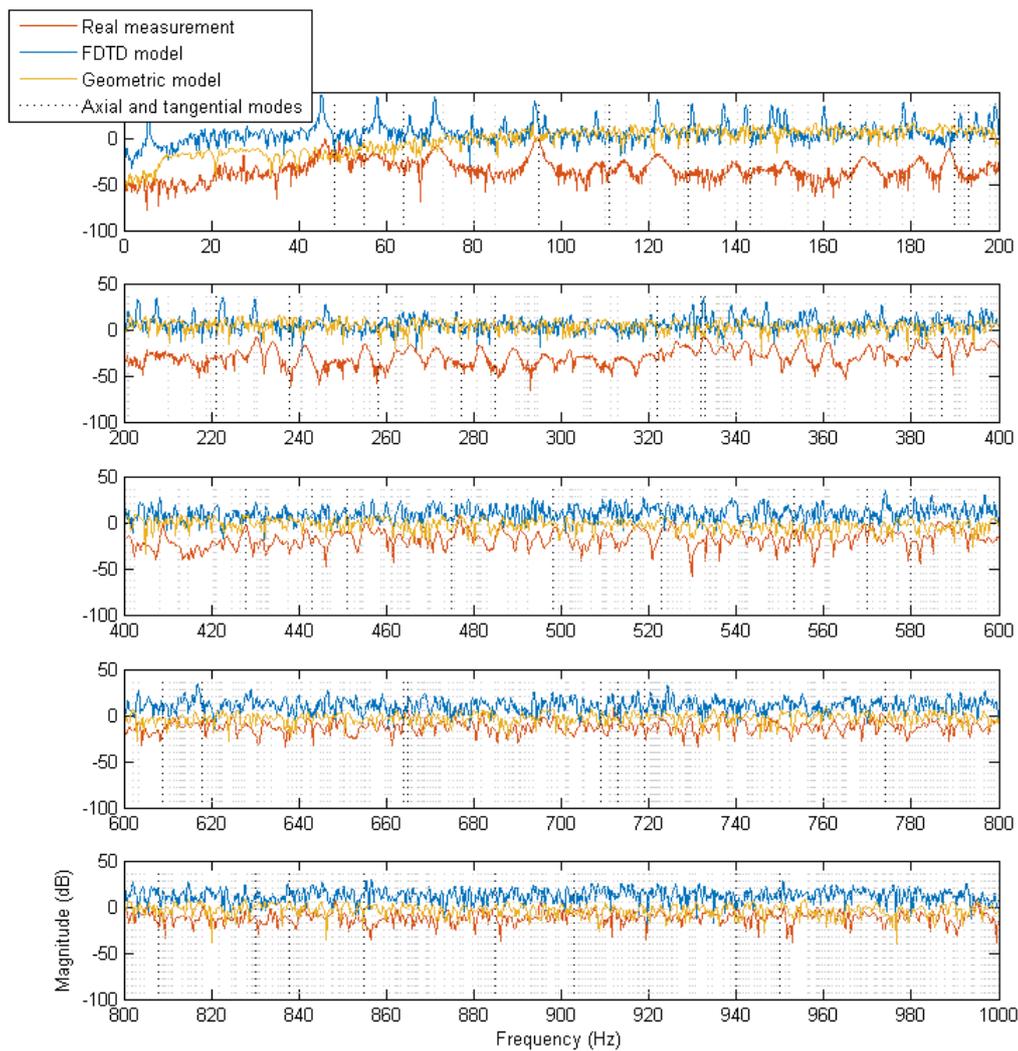
Figure 5.27: Case Study 2 : Frequency spectra of the real, FDTD and geometric RIRs at source/receiver (S/R) position 1. Theoretical axial room modes are also shown by the vertical dashed lines.

of the soft source. There are also a few room modes that are present in the real RIR and not in the FDTD RIR, and the other way around, which is due to to the complex geometry of the room, making the model more complicated to be absolutely accurate.

Even with the errors regarding the room modes positions in the FDTD RIR, Figure 5.27 shows that the numerical model is much more precise at low frequencies than the geometric one to preserve room modes. However, for frequencies higher than 550 Hz, room modes do not appear clearly on the real response. This could mean that the crossover frequency could be reduced to

approximately 550 Hz but this needs to be confirmed by other results in the frequency domain.
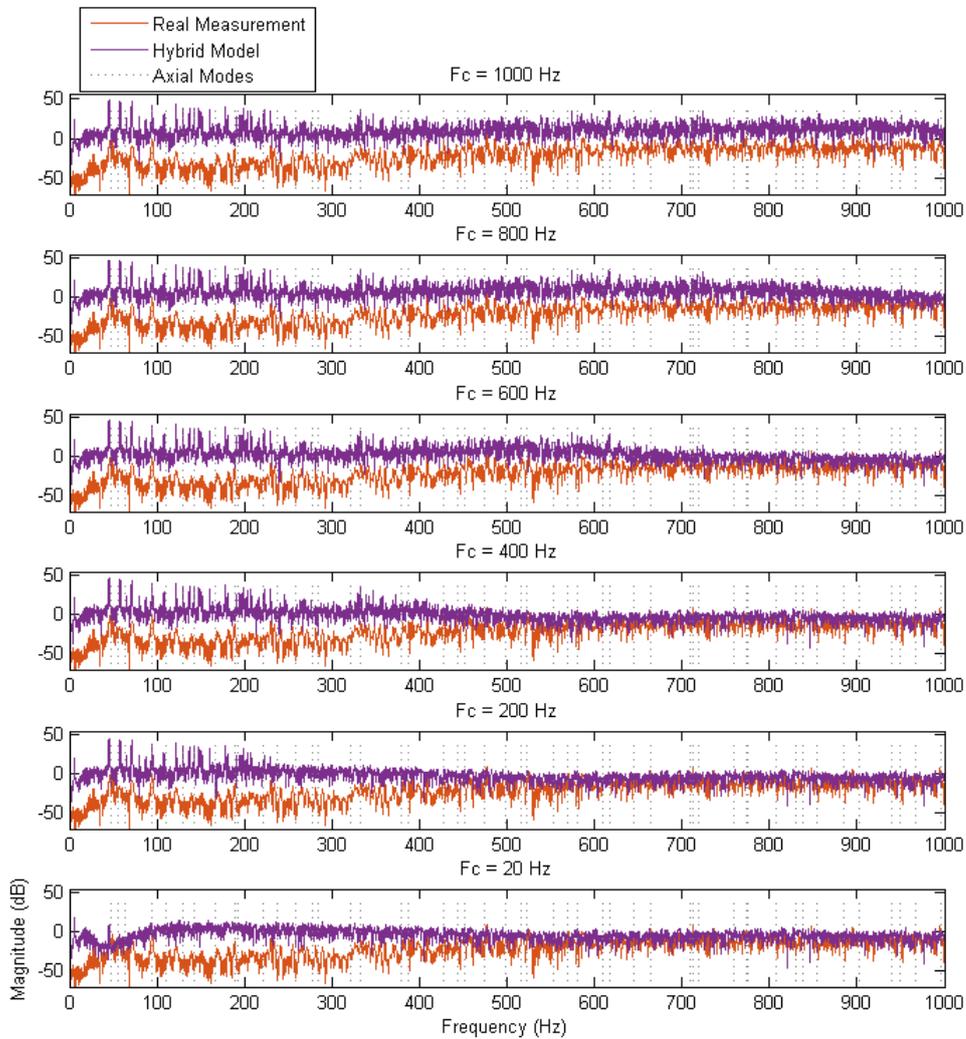
Figure 5.28: Case Study 2 : Frequency spectra of the real and hybrid RIRs at source/receiver (S/R) position 1 for several crossover frequency values (Fc = 20 Hz, 200 Hz, 400 Hz, 600 Hz, 800 Hz, and 1000 Hz). Theoretical axial room modes are also shown by the vertical dashed lines.

Figure 5.28 shows frequency responses at the first source/receiver position for the real measurement and hybrid models with a crossover frequency being 20 Hz, 200 Hz, 400 Hz, 600 Hz, 800 Hz, and 1000 Hz. It can be noticed that, for each hybrid model, resonant frequencies do not appear clearly on frequencies superior to the crossover frequency, due to the predominance of the geometric model. This means that the crossover frequency could be reduced to the highest frequency for which room modes are indistinguishable in the real measurement. However, it is

81

not possible to define such a limit visually speaking.

As with case study 1, these results could be confirmed by objective measurements like *Mean Squared Error*, *Frequency Response Assurance Criterion*, or *Magnitude Index* [29], but these are not relevant due to the magnitude offset between the numerical frequency response and the real one as a result of the calibration, leading to a major increase of disparity between the hybrid model and the real one as the crossover frequency diminishes.

In terms of analysis in the frequency domain, no objective criterion can be used to validate a minimal crossover frequency. However, visually speaking, the crossover frequency could be reduced to a value between 400 and 600 Hz without compromising the quality of the model, but this is very approximate. Similar results are obtained with source/receiver positions 2 and 3 with Figures 5.29, 5.30, 5.31, 5.32 displayed below.
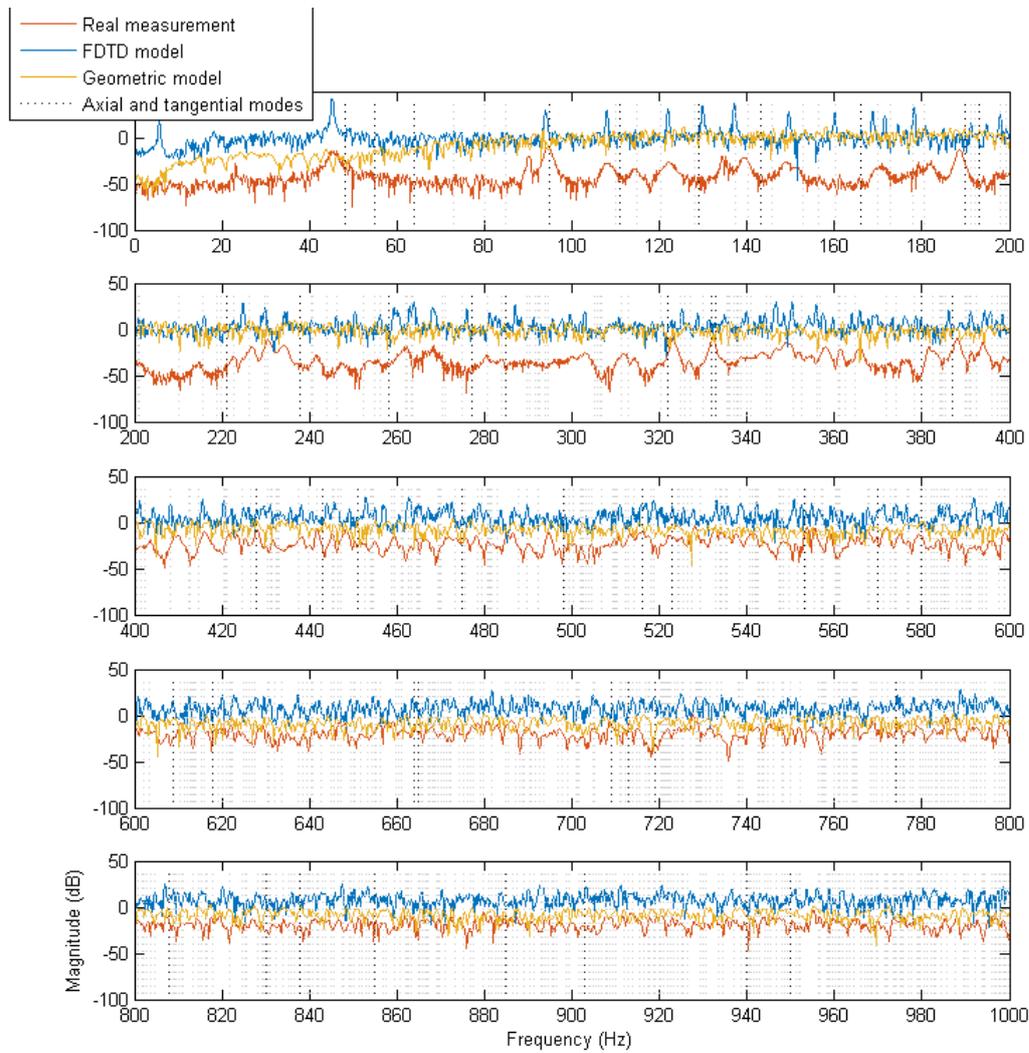
Figure 5.29: Case Study 2 : Frequency spectra of the real, FDTD and geometric RIRs at source/receiver (S/R) position 2. Theoretical axial room modes are also shown by the vertical dashed lines.
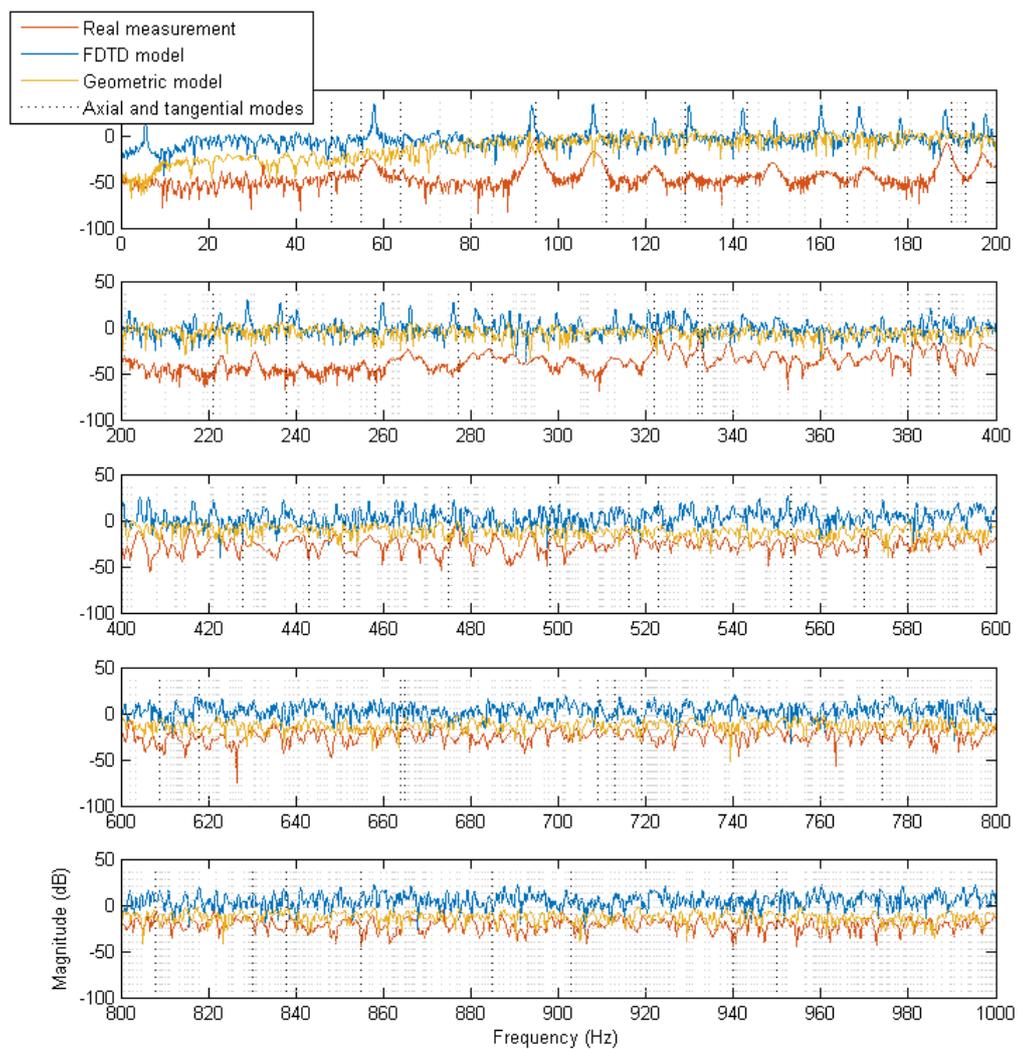
Figure 5.30: Case Study 2 : Frequency spectra of the real, FDTD and geometric RIRs at source/receiver (S/R) position 3. Theoretical axial room modes are also shown by the vertical dashed lines.
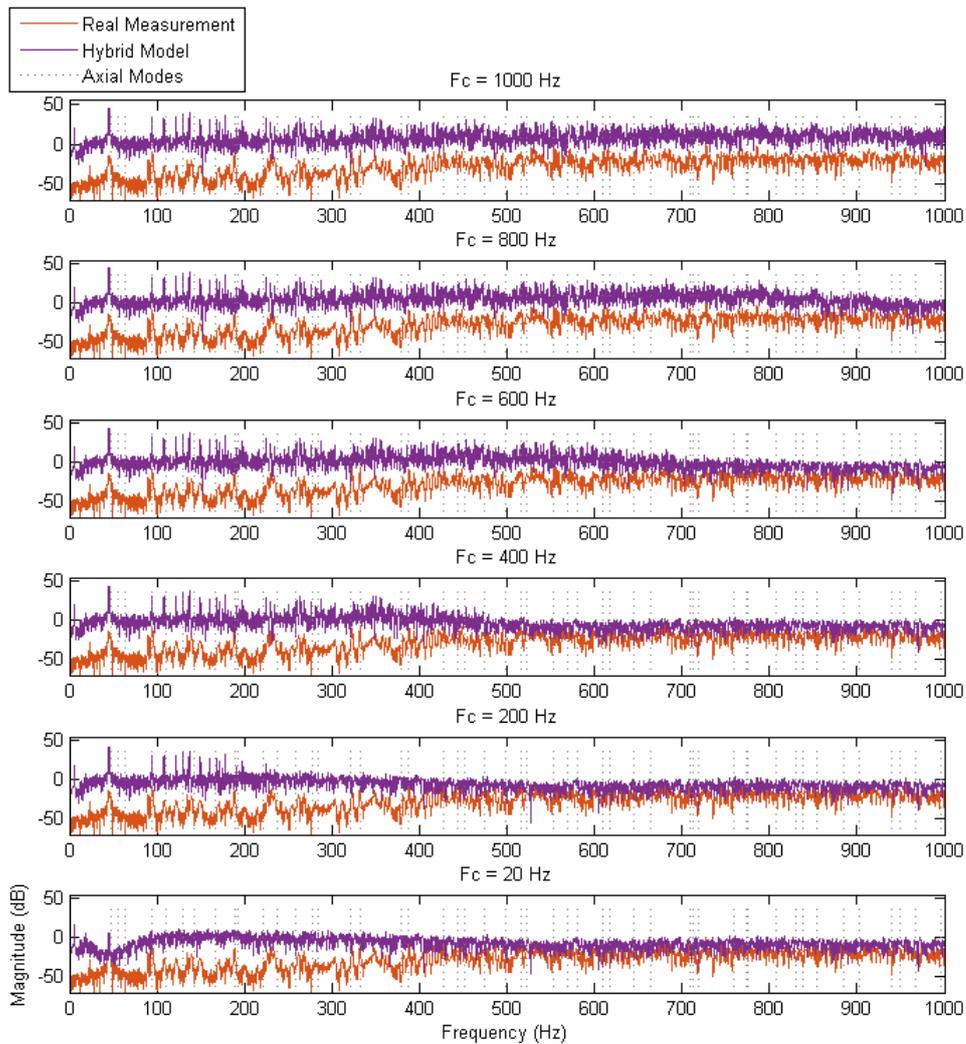
Figure 5.31: Case Study 2 : Frequency spectra of the real and hybrid RIRs at source/receiver (S/R) position 2 for several crossover frequency values (Fc = 20 Hz, 200 Hz, 400 Hz, 600 Hz, 800 Hz, and 1000 Hz). Theoretical axial room modes are also shown by the vertical dashed lines.
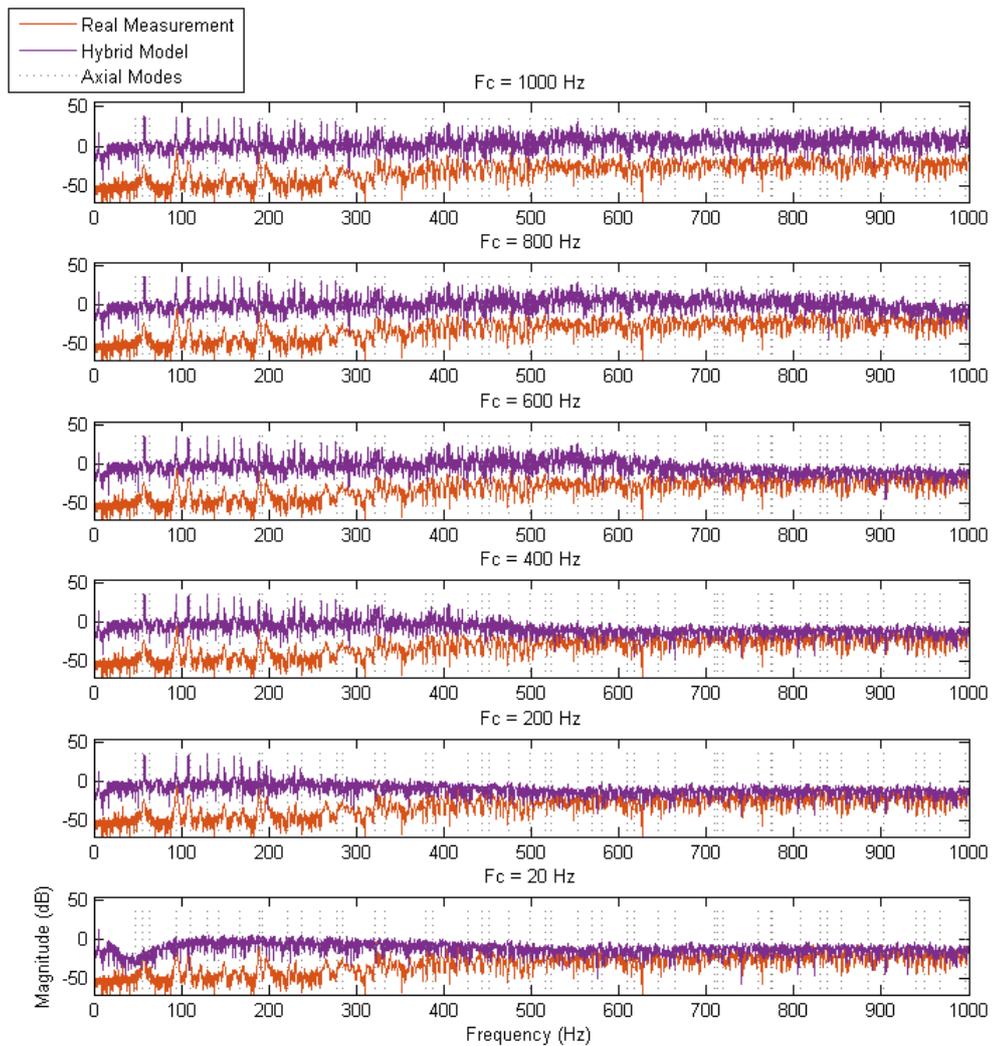
Figure 5.32: Case Study 2 : Frequency spectra of the real and hybrid RIRs at source/receiver (S/R) position 3 for several crossover frequency values (Fc = 20 Hz, 200 Hz, 400 Hz, 600 Hz, 800 Hz, and 1000 Hz). Theoretical axial room modes are also shown by the vertical dashed lines.

5.3 Summary of the Results and Discussion

In this chapter, two case studies have been presented and analysed, each corresponding to an existing room. Each room has been modelled with a hybrid model presented in Section 4.5 with 60 crossover frequencies and 3 source/receiver positions. Real measurements have been recorded and compared with each model in terms of analysis in the time domain, of acoustic parameters, and in terms of analysis in the frequency domain.

The main hypothesis of this research states that the crossover frequency of the hybrid model can be reduced down to the smallest distance between two boundaries in the room. Regarding dimensions for two rooms, the crossover frequency could therefore be reduced to 373 Hz for the first case study, and to 536 Hz for the second one.

In the time domain, hybrid models with varied crossover frequencies differ from each other by the shape of the peaks corresponding to the first reflections. It can be noted that for higher crossover frequencies, lower frequencies are better represented. As the peak apex of each distinguishable peak is at the same time position for each crossover frequency, it is complicated to measure the influence of the crossover frequency in the time domain. It also visually looks like reducing the crossover frequency would improve the accuracy of the global shape of the RIR. These results would indicate that the crossover frequency could be reduced to its minimum but further research is needed.

Acoustic parameters present the advantage of being totally objective. Hybrid models were compared with real measurements in terms of reverberation times (T_{30}, EDT) and energy ratios (C_{80}). They were all calculated from the octave band containing the crossover frequency of the hybrid RIR to improve the relevance of the results. For the first case study, reverberation times were abnormally higher in the hybrid model than for real measurements, but the same trend emerged than in case study 2, indicating that the crossover frequency could be reduced to 360 Hz without compromising the quality of the hybrid model in terms of T_{30} and EDT.

In both case studies, errors between C_{80} values from hybrid and real responses have shown that the hybrid model was more accurate when the crossover frequency was lowered between 360 and 700 Hz. However, as for reverberation times, these frequency values could be a general one and not specific to the room since it is the case for both case studies. Therefore, further research should be done on other rooms to verify this point.

In the frequency domain, resonant peaks from the hybrid models are consistent with the ones from real measurements, and correspond to the observed room modes. However, reducing the crossover frequency results in the disappearance of these modes at frequencies higher than F_c.

Therefore, the minimum crossover frequency corresponds to the frequency threshold for which room modes become less prominent. Visually, this is roughly estimated at 320 Hz for case study 1 and 550 for case study 2. These results would need to be validated with objective criterion but these were too complicated to be implemented here due to too much disparity between the hybrid model and the real RIR.

These results have shown that, in order to optimize the efficiency of a hybrid model, its crossover frequency could be reduced to frequencies much lower than 1 kHz, which is the common crossover frequency for such a model. Results tend to show that the crossover frequency could be reduced to an approximate value between 360 and 700 Hz depending on the comparison tool, but there is no significant change when the crossover frequency reaches the wavelength corresponding to the smallest distance between two boundaries in the room. Results derived from acoustic parameters show that it could be reduced further to 360 Hz and further research needs to be done in this direction. Furthermore, listening tests are essential to validate this hypothesis and have not been done due to a lack of time.

CHAPTER 6

CONCLUSION

This study has been concerned with the optimisation of a hybrid acoustic model by the reduction of its crossover frequency. The hybrid model used in this thesis combines a finite difference time domain method for the low frequency range and a ray tracing / image source method for the mid-high frequency range. The modelling was based on two existing rooms, at three source / receiver positions, from which real measurements were made.

The geometric model was synthesized via ODEON Auditorium [2] using a calibration process to obtain the absorption coefficients. The numerical model was implemented in MATLAB, using a standard rectilinear FDTD scheme based on locally reacting surface conditions [15]. To conserve the frequency dependency of the boundaries, the FDTD RIR was obtained by filtering 8 generated responses, each with absorption coefficient specific to each octave band, and by summing them.

The hybrid model was built by combining the numerical and the geometric RIR after a calibration process. The crossover frequency of the response, which is referred frequently in this research, corresponds to the cut-off frequency of the filters used on the geometric and numerical model. In each situation, 60 hybrid RIRs were synthesized, each with its own crossover frequency ranging from 20 Hz to 1200 Hz.

Hybrid RIRs were compared with real measurements in the frequency domain, in the time domain, and in terms of acoustic parameters. Results have shown that, in each case, the crossover frequency could be reduced to much less than 1 kHz, which is the crossover frequency value often quoted in previous work. However, the frequency boundary between an accurate model and an acceptable one is still very vague, and requires further research.

In a time domain analysis of the RIRs produced, it has been shown that reducing the crossover frequency resulted in a modification of the shape of the sound rays, but not of their position in time. It also affects the global shape of the RIR which visually looks more similar to real measurements at lower crossover frequencies (300 Hz or less). However, these assumptions require numerical comparative measures, which were too complicated to implement here, due to a huge disparity between the signals.

Acoustic parameter analysis has shown that reducing the crossover frequency to a value of 360 Hz did not affect the quality of the hybrid model in terms of reverberation times and energy ratios. The model would even be more accurate in terms of C_{80} with a crossover frequency between 360 and 700 Hz. However, it looks like this value could be a general one and not specific to the room, which would require further research on other rooms to assert this assumption.

In the frequency domain, it has been shown that reducing the crossover frequency resulted in a disappearance of the room modes above this frequency, and a rough threshold was visually selected, corresponding approximatively to 320 Hz for case study 1 and 550 Hz for case study 2. These values are close to the wavelength equivalent to the smallest distance between two boundaries of the enclosure, but need to be validated with objective criterion which were too complicated to be implemented here due to too much disparity between the signals.

Results tend globally to show that the crossover frequency could be reduced to an approximate value between 360 and 700 Hz depending on the comparison tool, but there is no significant change when the crossover frequency reaches the wavelength corresponding to the smallest distance between two boundaries in the room.

6.1   Restatement of Hypothesis

This research was conducted to validate the following hypothesis :

"The crossover frequency of a hybrid acoustic model can be reduced to the wavelength equivalent to the smallest distance between two boundaries of the enclosure, without compromising the quality of the model."

In the case of the models that have been presented, it can be stated that, visually and with tools of measurement, the crossover frequency could be reduced to values significantly lower than 1 kHz (approximately between 360 and 700 Hz) without affecting the position of early rays, room

modes, and the values of some essential acoustic parameters. However, this threshold is very vague and there is no significant change when the crossover frequency reaches the wavelength corresponding to the smallest distance between two boundaries in the room. These results cannot be validated without a listening test that would confirm or not the hypothesis and the results above and that could not be achieved due to a lack of time.

Although a formal listening test has not been conducted, an auralisation of several sound samples has been processed with the synthesized RIRs and the real measurements. It has revealed no audible distinction between the hybrid RIRs with crossover frequencies ranging from 360 Hz to 1200 Hz. However, for crossover frequencies lower than 360 Hz, real differences could be heard from sound samples containing a majority of low frequencies. Therefore, it is possible that the hypothesis of this thesis and the results summarized previously could be confirmed with a formal listening test and experimentation on other enclosures.

6.2 Further Research

This section provides information about future work towards the further optimisation of acoustic models for auralisation. Each line of approach is based on the elements of analysis presented in this thesis.

Use of Numerical Comparison Tools

In Chapter 5, hybrid responses were compared with real measurements in the time domain and in the frequency domain. However, the analysis was exclusively visual, since objective criterion like *Mean Squared Error*, *Frequency Response Assurance Criterion*, or *Magnitude Index* [29] have been discounted due to too much disparity between the signals. Possible avenues to explore would be filtering the RIRs before the comparison, or using other comparison tools that would be relevant.

Listening Tests

As mentioned previously, a listening test could confirm or not the results discussed in this thesis. The main idea would be to compare real measurements with synthesized RIRs after an auralisation process with several sound samples. The sound samples would need to each have a different frequency range to establish the impact of the crossover frequency towards the frequency spectra of the RIRs.

This listening test could be in the form of an 'ABX' test, with a sound reference 'X' that would correspond to the real measurement, and two responses ('A' and 'B') corresponding to synthesized hybrid RIRs with distinct crossover frequencies. Another way of performing the test would be to use a 'MUSHRA' test (MUltiple Stimuli with Hidden Reference and Anchor). This test should help to define more precisely the minimal crossover frequency of the model.

Measurement of the Absorption Coefficients

The absorption coefficients used in the study were determined by an estimation via acoustic parameters. Measuring them directly would allow the model to be much more precise, which would reduce the gap between the real RIR and the synthesized ones and would help in the determination of a valid crossover frequency. This process could not be done in the research due to a lack of resources.

Extension of the Case Studies

As the experimentation has only been conducted on two rooms, it could be interesting to test it on larger enclosures. In theory, the crossover frequency could be reduced to even lower frequencies. However, it has to be noticed that running simulations on larger spaces would considerably increase the execution time and the memory requirements.

[1]   *Sketchup software website.*
      http://www.sketchup.com.
      [Online; accessed September 2017].

[2]   *Odeon room acoustics software website.*
      http://www.odeon.dk, 1984.
      [Online; accessed September 2017].

[3]   *International standards office 3382-1:2009. acoustics - measurement of room parameters - part 1 : Performance spaces*, ISO, 3382(1) (2009).

[4]   S. BILBAO, *Numerical Sound Synthesis*, Wiley, 2009.

[5]   ——, *Modeling of complex geometries and boundary conditions in finite difference/finite volume time domain room acoustics simulation*, IEE Transactions on Audio, Speech, and Language Processing, 21 (2013).

[6]   D. BOTTELDOOREN, *Acoustical finite-difference time-domain simulation in a quasi-cartesian grid*, The Journal of the Acoustical Society of America, 95 (1993).

[7]   ——, *Finite difference time-domain simulation of low-frequency room acoustics problems*, The Journal of the Acoustical Society of America, 98 (1995).

[8]   A. FARINA, *Simultaneous measurement of impulse response and distortion with a swept-sine technique*, Audio Engineering Society, (2000).

[9]   B. HAMILTON, S. BILBAO, AND C. WEBB, *Revisiting implicit finite difference schemes for 3d room acoustics simulation on gpu*, (2013).

[10] C. HARRIS, *Absorption of sound in air versus humidity and temperature*, The Journal of the Acoustical Society of America, 40 (1966).

[11] D. HOWARD AND J. ANGUS, *Acoustics and Psychoacoustics*, Focal Press, 1996.

[12] H. JEONG AND Y. WAI LAM, *Source implementation to eliminate low-frequency artifacts in finite difference time domain room simulation*, The Journal of the Acoustical Society of America, 131 (2012).

[13] L. KINSLER, *Fundamentals of Acoustics*, Wiley, 2000.

[14] K. KOWALCZYK, *Boundary and Medium Modelling using compact finite difference schemes in simulations of room acoustics for audio and architectural design applications*, PhD thesis, Queen's University, Belfast, 2008.

[15] K. KOWALCZYK AND M. VAN WALSTIJN, *Room acoustics simulation using 3d compact explicit fdtd schemes*, IEE Transactions on Audio, Speech, and Language Processing, 19 (2011).

[16] A. KROKSTAD, *Calculating the acoustical room response by the uuse of a ray tracing technique*, J. Sound Vib., 8 (1968).

[17] H. KUTTRUFF, *Room Acoustics*, Spon Press, 1973.

[18] H. D. LUKE, *The origins of the sampling theorem*, IEEE Communications Magazine, (1999).

[19] S. MULLER AND P. MASSARANI, *Transfer-function measurement with sweeps*, Audio Engineering Society, 49 (2001).

[20] D. MURPHY, A. KELLONIEMI, J. MULLEN, AND S. SHELLEY, *Acoustic modeling using the digital waveguide mesh*, IEEE Signal Processing Magazine, (2007).

[21] D. MURPHY, S. SHELLEY, M. BEESON, AND A. MOORE, *Hybrid room impulse response synthesis in digital waveguide mesh based room acoustics simulation*, Conference on Digital Audio Effects, 11 (2008).

[22] D. MURPHY, A. SOUTHERN, AND L. SAVIOJA, *Source excitation strategies for obtaining impulse responses in finite difference time domain room acoustics simulation*, Elsevier, Applied Acoustics, 82 (2014).

[23] D. T. MURPHY, *Digital Waveguide Mesh Topologies in Room Acoustics Modeling*, PhD thesis, University of York, 2000.

[24] S. OXNARD, *Efficient Hybrid Virtual Room Acoustic Modelling*, PhD thesis, University of York, 2016.

[25] B. POSTMA, *Serious Auralizations*, PhD thesis, LIMSI Orsay, 2017.

[26] V. PULLKI, *Virtual sound source positioning using vector based amplitude panning*, Audio Engineering Society, 45 (1997).

[27] L. SAVIOJA AND U. SVENSSON, *Overview of geometrical room acoustic modeling techniques*, The Journal of the Acoustical Society of America, 138 (2015).

[28] J. SHEAFFER, B. FAZENDA, D. MURPHY, AND J. ANGUS, *A simple multiband approach for solving frequency dependent problems in numerical time domain methods*, Forum Acusticum, Aalborg, Denmark, (2011).

[29] K. SHIN, *An alternative approach to measure similarity between two deterministic transient signals*, Journal of Sound and Vibration, 371 (2016).

[30] J. I. SMITH, *On the equivalence of the digital waveguide and finite difference time domain schemes*, (2004).

[31] A. SOUTHERN, S. SILTANEN, D. MURPHY, AND L. SAVIOJA, *Room impulse response synthesis and validation using a hybrid acoustic model*, IEE Transactions on Audio, Speech, and Language Processing, 21 (2013).

[32] A. SOUTHERN, S. SILTANEN, AND L. SAVIOJA, *Spatial room impulse responses with a hybrid modelling method*, Audio Engineering Society, AES 130th Convention (2011).

[33] G. STAN AND J. EMBRECHTS, *Comparison of different impulse response measurement techniques*, J. Audio Eng. Soc., 50 (2002).

[34] V. VALIMAKI, J. PARKER, L. SAVIOJA, J. SMITH, AND J. ABEL, *Fifty years of artificial reverberation*, IEE Transactions on Audio, Speech, and Language Processing, 20 (2012).

[35] S. VAN DUYNE AND I. SMITH, J.O., *The 2-d digital waveguide mesh*.

[36] M. VORLANDER, *International round robin on room acoustical computer simulations*, 15th International Congress on Acoustics Trondheim, Norway, (1995).

[37] ——, *Auralization*, RWTH, Springer, 2008.

[38] K. S. YEE, *Numerical solution of initial boundary value problems involving maxwell's equations in isotropic media*, (1966).