UNIVERSITY OF SOUTHAMPTON
FACULTY OF ENGINEERING AND THE ENVIRONMENT
INSTITUTE OF SOUND AND VIBRATION RESEARCH

INTERACTIVE AURALIZATION BASED ON HYBRID SIMULATION METHODS AND PLANE WAVE EXPANSION

by

Diego Mauricio Murillo Gómez

Supervisors
Dr. Filippo Fazi
Prof. Jeremy Astley

A thesis submitted in partial fulfillment for the
degree of Doctor of Philosophy

June 2016

Declaration of Authorship

I, Diego Mauricio Murillo Gómez, declare that this thesis titled, 'Interactive Auralization Based on Hybrid Simulation Methods and Plane Wave Expansion' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: _____

Date: June 2016

Dedicado a

***Mafe**, mi más puro y sincero amor.*

***Madre**, ejemplo de fuerza absoluta.*

***Padre**, apoyo incondicional.*

***Esposa**, mi amor eterno.*

***Hermano** y **Hermana**, unión inquebrantable.*

***Poesía en música**, mi brújula en la vida.*

Sería nada sin ustedes...

UNIVERSITY OF SOUTHAMPTON
Faculty of Engineering and the Environment
Institute of Sound and Vibration Research

Abstract

Doctor of Philosophy

INTERACTIVE AURALIZATION BASED ON HYBRID SIMULATION METHODS AND PLANE WAVE EXPANSION

by Diego Mauricio Murillo Gómez

The reconstruction and reproduction of sound fields have been extensively researched in the last decades leading to an intuitive approach to estimate and evaluate the acoustic properties of enclosures. Applications of auralization can be found in acoustic design, subjective tests, virtual reality and entertainment, among others. Different methodologies have been established to generate auralizations for room acoustics purposes, the most common of them, the use of geometrical acoustics and methods based on the numerical solution of the wave equation to synthesize the room impulse responses. The assumptions and limitations of each approach are well known, which in turn, restrict their application to specific frequency bands. If the aim is to reconstruct accurately the sound field in an extended range of frequencies, a combination of these methodologies has to be performed. Furthermore, recent advances in computational power have enabled the possibility to generate interactive atmospheres where the user is able to interact with the environment. This feature, although it expands the applications of the auralization technique, is nowadays mainly based on geometrical acoustics or interpolation methods.

The present research addresses the generation of interactive broadband auralizations of enclosures using a combination of the finite element method and geometrical acoustics. For this, modelling parameters for both simulation methods are discussed making emphasis on the assumptions made in each case. Then, the predicted room impulse responses are represented by means of a plane wave expansion, which in turn, enables interactive features such as translation and rotation of the acoustic fields. An

analytical expression is derived for the translation in the plane wave domain. Furthermore, the transformation of the plane wave representation in terms of spherical harmonics is also explored allowing the acoustic fields to be rotated. The effects of assuming a plane wave propagation within small enclosures and the consequences of using a finite number of plane waves to synthesize the sound fields are discussed.

Finally, an implementation of an interactive auralization system is considered for different reference cases. This methodology enables reconstruction of the aural impression of enclosures in real-time with higher accuracy at low frequencies compared to only geometrical acoustics techniques. The plane wave expansion provides a convenient sound field representation in which the listener can interact with the acoustics of the enclosure. Furthermore, the sound reconstruction can be performed by implementing several sound reproduction techniques extending the versatility of the proposed approach.

Acknowledgements

This journey has been one of the best experiences of my life. The growth from the academic perspective, but also from the personal point of view, is not possible to describe it in words. I would like to begin by thanking to the creative energy of the universe, whose good vibes allow me to finish this significant stage of my life. My most sincere and deepest thanks to my supervisor Dr. Filippo Fazi. His advice and support from the start to the end of my PhD were invaluable. His professional skills and love for science have been a source of inspiration that pushed me every day to be a better scientist. My immense gratitude to my other supervisor Prof. Jeremy Astley, who was also very supportive and provided me with exceptional advice when I need it. I feel very grateful to have had the opportunity of learning from these two excellent persons.

I would particularly like to extend warm thanks to all my friends from the VAAE team and ISVR whose discussions about audio and acoustics have been very enriching. A huge thanks to Colciencias and the University of San Buenaventura Medellín whose funding provided me the possibility of studying a PhD.

From a more personal perspective, I am immeasurably grateful to my wife Leidy Viviana and my daughter Mafe. Your support and company have been infinitely valuable to face all the challenges that involve to do a PhD. “Amor”, thanks for being my travel partner, my confident, my “parcera”. You have always been by my side in the good and bad times. Mafe, you are the most beautiful thing that has ever happened to me. I saw you confidently to overcome all the troubles of studying in a different language, which makes me feel so proud of you. You are my clever girl!

My eternal thanks to my Mother and Father. You are exceptional parents who taught me good values, which will stay with me for my entire life. I am very glad to show you that all your efforts and sacrifices have been worth it. Thanks to my brother David and sister Juanita, your love has always been a fuel to keep me going. I am much obliged with all my family, my uncles and aunts who have been a significant and encouraging guidance in my life.

A special thanks to all my friends in Southampton, in particular to Andres and Rosmery, whose company and unconditional support have been very important for Leidy, Mafe and me. Thanks for your genuine and irreplaceable friendship. I would also like to express thanks to my friend Mateo, our philosophical discussions have extended my perspective about life.

Finally, I would like to thank to the poetry in music (RAP), everything that I am today is because of it. My critical thinking, my standpoint about society and my will to help my country through education have been positively influence by it. From now, I am an example showing that is possible to achieve your dreams if you work hard and with passion. It does not matter the economic limitations; the real impediment is the lack of will. Thanks to my "parceros": Oscar (compa), Alex (jugla), Eddson (ning1), Michael (mic), Marvin (el gato), Jhon (meliko), Mauricio (mao), José (chepe), Sergio and my other mates. We made it!!!

Contents

List of Figures

List of Tables

Abbreviations

BEM	**B**oundary **E**lement **M**ethod
BMR	**B**alance **M**odel **R**eduction
BRIR	**B**inaural **R**oom **I**mpulse **R**esponse
C_{80}	**C**larity Index
CTC	**C**rosstalk **C**ancellation
DIVA	**D**igital **I**nteractive **V**irtual **A**coustics
EDT	**E**arly **D**ecay **T**ime
FDTD	**F**inite **D**ifference **T**ime **D**omain
FEM	**F**inite **E**lement **M**ethod
FIR	**F**inite **I**mpulse **R**esponse
GA	**G**eometrical **A**coustics
HOA	**H**igh **O**rder **A**mbisonics
HRTF	**H**ead **R**elated **T**ransfer **F**unction
HPIR	**H**ead **P**hone **I**mpulse **R**esponse
IDW	**I**nverse **D**istance **W**eighting
IIR	**I**nfinite **I**mpulse **R**esponse
ILD	**I**nteraural **L**evel **D**ifference
IPTF	**I**nter **P**ositional **T**ransfer **F**unction
IR	**I**mpulse **R**esponse
ISM	**I**mage **S**ource **M**odel
ITD	**I**nteraural **T**ime **D**ifference
PWE	**P**lane **W**ave **E**xpansion
RIR	**R**oom **I**mpulse **R**esponse
RT	**R**ay **T**racing

SDR	**S**ignal to **D**istortion **R**atio
SHE	**S**pherical **H**armonics **E**xpansion
T_{20}	**R**everberation Time
VBAP	**V**ector **B**ase **A**mplitude **P**anning

Table of Symbols

Mathematical symbols	
j	imaginary unit ($j = \sqrt{-1}$)
$\mathcal{F}$, $\mathcal{F}^{-1}$	direct and inverse Fourier transform
$(\cdot)^*$	complex conjugate of a complex number
$\lceil \cdot \rceil$	rounding to next integer operator
$\lvert \cdot \rvert$	absolute value
$\lVert \cdot \rVert$	norm of a matrix (largest singular value)
$\otimes$	convolution
$:=$	definition
$\nabla_{\mathbf{n}}$	normal derivative operator ($\nabla_{\mathbf{n}} = \hat{\mathbf{n}} \cdot \nabla$)
$\angle(\cdot)$	phase of a complex number $q = \lvert q \rvert \, e^{j \angle q}$
$\in$	...is in...
Sets, vectors and matrices	
Ω	unitary sphere
$\mathbf{x}$	position vector of an acoustic field point
$\hat{\mathbf{y}}$	unitary vector identifying the direction of wave propagation
$\hat{\mathbf{n}}$	unitary vector normal to a surface
$\mathbf{I}$	identity matrix
$\mathbf{H}$	plane wave propagation matrix
$\mathbf{Y}$	spherical harmonic and spherical Bessel propagation matrix
Physical entities	
$p(\mathbf{x})$	target acoustic pressure [Pa]
$\widetilde{p}(\mathbf{x})$	reconstructed acoustic pressure [Pa]
$\mathbf{u}(\mathbf{x})$	particle velocity [m s^{-1}]

$\mathbf{I}(\mathbf{x})$	time averaged acoustic intensity [kg s^{-3}]
W	acoustic power [watt]
Z	specific acoustic impedance [kg $m^{-2}s^{-1}$]
c	speed of sound propagation [m s^{-1}]
ρ_0	static density of the fluid [kg m^{-3}]
p_{ref}	reference pressure [20×10^{-6} Pa for air]
W_{ref}	reference power [10^{-12} watt]
k	wave number [rad m^{-1}]
t	time [s]
ω	angular frequency [rad s^{-1}]

Functions

$Y_n^m(\cdot)$	complex spherical harmonic of order n and mode m
$A_{nm}(\cdot)$	complex spherical harmonic coefficient n and mode m
$P_n^m(\cdot)$	associated Legendre function
$j_n(\cdot)$	n-th order spherical Bessel function
$h_n^{(2)}(\cdot)$	spherical Hankel function of second kind
$\Upsilon_n^m(\cdot)$	real spherical harmonic of order n and mode m

Mi objetivo en la vida es ser feliz y ayudar a los demás a serlo

Chapter 1

Introduction

Auralization is the technique of creating an audible perception from numerical information related to the acoustics of a specific environment [1]. It is a powerful tool because it provides the possibility to render the sound field according to the characteristics of the medium, which has relevant applications in the evaluation and understanding of the physical phenomenon under consideration. Applications of auralization can be found in a wide range of areas such as room acoustics, environmental acoustics, noise control, virtual reality, teaching, medicine, entertainment and telecommunications. Although it is feasible to create auralizations of enclosures using scale models, the implementation of simulation techniques such as methods for the numerical solution of the wave equation or geometrical acoustics has become the most common approach to estimate the room impulse responses. This is in part due to advances in electronics and computation technologies achieved in the last decades.

This thesis addresses the generation of interactive auralizations of enclosures whose impulse responses have been predicted numerically by means of the finite element method (FEM) and geometrical acoustics (GA). The main contribution of this research is the development of a methodology to reconstruct the acoustic field of rooms in real-time. The combination of a method based on the numerical solution of the wave equation and geometrical acoustics enhances the applications and realism of the auralization technique, in particular at low frequencies. However, the computational cost required for both approaches presents a significant challenge to synthesize interactively the sound fields. To tackle this problem, the plane wave expansion is proposed as a tool to represent the sound field to be auralized. Moreover, although the current research is conducted using the finite element method to predict the low

frequency content of the room impulse responses, different approaches such as the boundary element method or the finite difference time domain method can also be implemented based on this methodology.

The main motivation of this research is to improve the auralization technique by generating an interactive sound field reconstruction in which the listener is able to listen to the changes in the acoustic field based on their relative position with respect to the enclosure. The plane wave expansion provides a convenient description that allows for interactive features such as translation and rotation of the sound field. The translation is performed by the implementation of a mathematical operator in the plane wave domain rather than an interpolation process between Ambisonics impulse responses, which is an alternative strategy proposed by the scientific community [2]. Furthermore, the interoperability of the plane wave representation with the spherical harmonics or the straightforward implementation of vector base amplitude panning enables the rotation of the acoustic field. Finally, the use of FE and GA results to generate the plane wave expansion improves the accuracy of the proposed methodology enhancing the applications of the approach.

Geometrical acoustics is a well-established technique to simulate the acoustic behaviour of enclosures [3]. It is founded on the assumption that sound waves can be replaced by rays travelling in the enclosure with each ray carrying a portion of the energy produced by the source [4]. However, the postulation of waves propagating as rays is only sensible when the wavelength is small compared to the dimension of the surfaces of the room, otherwise wave effects such as diffraction are relevant [5]. Although the implementation of scattering coefficients has improved the effectiveness of geometrical acoustics at middle and high frequencies [6], diffraction remains the greatest drawback of this method. Recently, new approaches for edge diffraction calculation have been proposed [7, 8], nevertheless, assumptions such as infinite planes or rigid surfaces constrain their application.

Alternatively, the propagation of the sound in a compressible medium is calculated by solving the wave equation. An analytical solution is only possible in a few cases under specific conditions and for simple geometries, which leads to the implementation of numerical methods to approximate it. The finite element method [9] has been extensively used to analyse and solve problems in acoustics. The room impulse response is predicted by solving the Helmholtz equation numerically for a range of

frequencies and by applying the inverse Fourier transform at specific field points. Time domain formulation of the method has recently been proposed as well [10]. The main advantage compared to GA methods lies in the accuracy of the results at low frequencies because wave effects such as diffraction are inherently included in the solution. However, the high computational cost constrains their use at higher frequencies and for larger rooms. This is because the required resolution of the mesh where the solution is computed depends on the frequency.

Taking into consideration the advantages and disadvantages of the two simulation methods described above, it is sensible to combine them according to the frequency range of interest over which each is most effective. In this case, the finite element method is used at low frequencies where the modal behaviour and the diffraction effects are significant, and geometrical acoustics is implemented at medium and high frequencies where its assumptions are acceptable and yield reasonable results. Based on that, simulation parameters such as the geometrical detail of the room, the representation of the acoustic source and characterization of the boundary conditions are addressed for each method in Chapter 3. In particular, an analysis of the use of measurements of reverberation time to characterize the boundary conditions of the room is conducted. From this information, the diffuse absorption coefficients are estimated and the specific acoustic impedances of the absorbing surfaces are calculated. The implications of assuming that the specific acoustic impedance is entirely resistive are also discussed. This method leads to a simple and practical approach to synthesize room impulse responses when both simulation techniques are used. It is important to point out that the implementation of hybrid simulation methods for room acoustic simulations is not a novel topic in the scientific community [11, 12]. Nevertheless, a contribution of this research is how to use this data as an input to generate an interactive auralization based on the manipulation of the sound field due to a plane wave representation.

In Chapter 4, results from the finite element and geometrical acoustics simulations are processed to generate a plane wave expansion (PWE) of the acoustic fields. The plane wave representation has been adopted because it is a solution of the homogenous Helmholtz equation that allows for the translation and rotation of sound fields. Furthermore, different sound reproduction techniques such as binaural, Ambisonics, Wave Field Synthesis or VBAP can be straightforwardly implemented, thus making the plane wave expansion a convenient representation for auralization purposes. The

consequences of the assumption of plane wave propagation in small rooms and of the discretization of the plane wave expansion, which is required for the implementation of the method, are also analyzed.

Different strategies are proposed to generate a plane wave representation from each simulation method. In the case of FEM, an inverse method is implemented to estimate a set of plane waves whose complex amplitudes reconstruct a target acoustic field. This methodology has already been used by Støfringsdal and Svensson [13] for 2D cases. To the extent of the author's knowledge, this is the first implementation for 3D cases. The inverse problem is formulated by using acoustic pressure sampled at points from the finite element solution. The effects of Tikhonov regularization are investigated in terms of the accuracy of the synthesized acoustic fields and the complex amplitude of the plane waves. The relation between a plane wave expansion of L elements and a spherical harmonic expansion of order N is also considered, which is useful for the interoperability between these two domains.

The plane wave representation for the geometrical acoustic data is generated using a different approach. For each enclosure, the reflections that contribute to the omnidirectional room impulse response predicted at the origin where the plane wave expansion is performed are traced using a virtual directional microphone. The virtual directional microphone was provided by the developer of the commercial package used for the simulations (Catt-Acoustics). The plane wave expansion is generated by assuming that the reflection paths are produced by plane waves. A discussion about the suitability of this assumption is presented for different reference cases.

Nevertheless, because the plane wave expansions from the FE and GA results have been generated using different approaches, the number of elements in each PWE may be not the same. In this research, a higher number of plane waves has been adopted for the GA information to take advantage of the high-directional resolution provided by the virtual microphone used to trace the reflections. This involves a spatial resampling of the PWE from the GA data when a combination of the results from the two simulation methods is intended. Several techniques to perform the spatial resampling are discussed, which leads to the same number of elements for each plane wave expansion. Finally, a method to combine the plane wave expansions from the

two simulation methods is proposed.

Chapter 5 addresses the development of efficient techniques that allow for sound field manipulation. This is carried out based on a unified plane wave expansion that has been estimated from finite element and geometrical acoustic information. The interactivity of the auralization is achieved by allowing the listener to move around the enclosure. This involves the manipulation of the plane wave expansion in terms of translation and rotation. A theoretical formulation to perform the translation of acoustic fields directly in the plane wave domain is presented. Regarding rotation, the use of an interpolation method is discussed allowing the rotation to be conducted in the plane wave domain. Furthermore, a spherical harmonic transformation, which allows for the implementation of a rotation operator is evaluated as well.

Finally, Chapter 6 presents the development and implementation of an interactive auralization system composed of a graphical interface in which an avatar moves inside of a virtual enclosure (similar to a first-person video game). This is carried out using the commercial packages Max as an acoustic rendering and Unity to generate the virtual environments. The rendering of the acoustic field is performed in real-time based on the position and orientation of the avatar. The auralization is presented to the listener by means of a headphone-based binaural approach. However, different sound reproduction techniques can be easily implemented. Experiments on the ability of the proposed methodology to accurately reconstruct the acoustic field are presented for several reference cases to validate the theoretical results. Limitations of the current approach are discussed and techniques to improve them are also considered.

The proposed methodology allows for a more accurate reconstruction of the acoustic fields compared to the use of only geometrical acoustics techniques. Although the framework presented in this thesis is based on room acoustics applications, the potential offered to auralize interactively acoustic data from methods based on the numerical solution of the wave equation provides a tool to analyze different problems in acoustics. Such applications range from noise control, loudspeaker design to environmental acoustics.

The main contributions of this thesis to the existing knowledge in the field are:

- Demonstration of the suitability of the plane wave expansion to generate an interactive broadband auralization.
- Development of a complete framework for generate an interactive auralization based on a plane wave expansion.
- Demonstration of the suitability of the inverse methods to create a plane wave expansion from finite element data in 3D cases.
- The interactive auralization of acoustic fields that have been predicted by means of the finite element method and geometrical acoustics.
- Demonstration of the auralization system effectiveness as tool for the reproduction of the modal response of enclosures in real-time.
- Integration of several audio technologies to generate a real-time interactive auralization system.

A list of publications that cover part of the material presented in this thesis is reported below:

- Diego M. Murillo, Filippo M. Fazi and, Mincheol Shin, Evaluation of Ambisonics Decoding Methods with Experimental Measurements, EAA Joint Symposium on Auralization and Ambisonics, Berlin, 2014.
- Diego M. Murillo, Catriona Cooper and Filippo M. Fazi, Acoustic Survey of a Late Medieval Building Based on Geometrical Acoustics Methods, Forum Acusticum, Krakow, 2014.
- Diego M. Murillo, Filippo M Fazi and Jeremy Astley, Spherical Harmonics Representation of the Sound Field in a Room Based on Finite Element Simulations, 46 Congreso Iberoamericano de Acústica, Valencia, 2015.

The following awards have been received by the author:

- Young Scientist Conference Attendance Grant, 46vo Congreso Iberoamericano de Acústica, SAE, 2015.
- First place in the Audio Engineering Society UK Graduate Student Poster Competition, AES UK section, 2016.
- Young Scientist Conference Attendance Grant, 22nd International Congress on Acoustics, ICA-ASA, 2016.

Chapter 2

Theoretical Background and Literature Review

This chapter addresses the fundamental bases for the understanding and development of the research. It is composed of two main sections: the theoretical background and the literature review.

2.1 Theoretical background

In this section the mathematical notation used throughout the thesis is presented. Then, mathematical transforms that allow for the analysis of a signal in different domains are discussed. Finally, the physical models of sound waves propagating in air and their interaction with a bounded space are considered.

2.1.1 Mathematical notation

A description of the notation used through this thesis is given as follows:

- Vectors are represented by lower case boldface letters "$\mathbf{x}$".
- Matrices are represented by upper case boldface letters "$\mathbf{H}$".
- Scalars, functions or scalar fields in the frequency domain are represented by lower case italic letters "$p(x) = p(x, \omega)$".

- Scalars, functions, scalar fields in the time domain are represented by upper case italic letters "$P(x) = P(x,t)$".
- Where there is a potential of confusion or for symbols that are commonly used in the scientific literature such as (ρ) for density, the notation includes the time or frequency dependency.

2.1.1.1 Vectors and matrices

The components of a vector $\mathbf{x}$ are given by $[x_1, x_2, x_3]$ in Cartesian or $[r_x, \theta_x, \phi_x]$ in spherical coordinates. A component of a vector $\mathbf{x}$ is defined as ($x_1 := \mathbf{x} \cdot \hat{\mathbf{x}}_\mathbf{1}$) or ($r_x := \mathbf{x} \cdot \hat{\mathbf{r}}$), respectively, in which ($\cdot$) represents the scalar product between two vectors:

$$\mathbf{x} \cdot \mathbf{y} = \sum_{n=1}^{N} x_n y_n. \tag{2.1}$$

The direction and module of a vector $\mathbf{x} = [x_1, x_2, x_3]$ are defined as

$$x = |\mathbf{x}| := \sqrt{\sum_{n=1}^{3} |x_n|^2}, \quad \hat{\mathbf{x}} := \frac{\mathbf{x}}{|\mathbf{x}|}. \tag{2.2}$$

The relation between Cartesian and spherical coordinates for a vector $\mathbf{x}$ is given by

$$\mathbf{x} = [x_1, x_2, x_3] = [r_x \cos\phi_x \sin\theta_x, r_x \sin\phi_x \sin\theta_x, r_x \cos\theta_x], \tag{2.3}$$

where θ_x and ϕ_x are the elevation and azimuthal components of the vector $\mathbf{x}$, respectively.

2.1.1.2 Complex numbers and functions

The imaginary unit is represented by the symbol $j = \sqrt{-1}$. A complex scalar or function $q(\omega)$ can be described by its real and imaginary part or by its magnitude $|\cdot|$ and phase $\angle(\cdot)$, namely

$$q(\omega) = |q(\omega)|\, e^{j\angle q(\omega)}. \tag{2.4}$$

2.1.2 Fourier transform

The forward and inverse Fourier transform are defined as follows:

$$\mathcal{F}[F(t)] = f(\omega) = \int_{-\infty}^{\infty} F(t)e^{-j\omega t}dt \tag{2.5}$$

and

$$\mathcal{F}^{-1}[f(\omega)] = F(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} f(\omega)e^{j\omega t}d\omega, \tag{2.6}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote the direct and inverse Fourier transform, respectively. $f(\omega)$ and $F(t)$ are the function in the frequency and time domain, respectively. The interaction between these two domains helps with the analysis of the different properties of the function. Through the time domain it is possible to evaluate the variation in the amplitude of the signal, which leads the estimation of acoustic parameters such as reverberation time, the early decay time, clarity index, among others. In contrast, the analysis in the frequency domain allows for the inspection of the spectrum identifying the distribution of the energy across frequency. This is useful for example, when sound waves interact with the boundaries of a space with frequency dependent characteristics. Moreover, some mathematical operations are easier to compute in one domain than in the other. For example, the operation of convolution in the time domain is equivalent to the multiplication in the frequency domain [14]:

$$F(t) \otimes G(t) = \mathcal{F}^{-1}\left[f(\omega)g(\omega)\right], \tag{2.7}$$

where $\otimes$ denotes the convolution operation.

2.1.3 Wave equation

The space-time dependence of the acoustic pressure fluctuations is described by a partial differential equation called as the *wave equation*. The derivation of this expression is performed based on the mass conservation and the momentum conservation laws (equations (2.8) and (2.9))

$$\frac{\partial \rho_{tot}(\mathbf{x},t)}{\partial t} + \nabla \cdot (\rho_{tot}(\mathbf{x},t)\mathbf{U}(\mathbf{x})) = 0, \tag{2.8}$$

$$\frac{\partial \mathbf{U}(\mathbf{x})}{\partial t} + \mathbf{U}(\mathbf{x}) \cdot \nabla \mathbf{U}(\mathbf{x}) + \frac{1}{\rho_{tot}(\mathbf{x}, t)} \nabla P_{tot}(\mathbf{x}) = 0, \tag{2.9}$$

where P_{tot} is the total instantaneous pressure, $\mathbf{U}$ is the particle velocity and ρ_{tot} is the total instantaneous density. To linearize the previous equations, it is expected that changes in the pressure P and density ρ are small compared to the ambient values p_0 and ρ_0. In addition, the particle velocity $\mathbf{U}$ should be much smaller than the speed of sound c. Under these conditions and assuming that the propagation of the wave occurs in an ideal gas (which air is a good approximation of), equations (2.8) and (2.9) can be simplified as:

$$\frac{\partial \rho(\mathbf{x}, t)}{\partial t} + \rho_0 \nabla \cdot \mathbf{U}(\mathbf{x}) = 0 \tag{2.10}$$

and

$$\rho_0 \frac{\partial \mathbf{U}(\mathbf{x})}{\partial t} + \nabla P(\mathbf{x}) = 0. \tag{2.11}$$

Relating the above two equations with the equation of state (2.12) it is possible to establish the wave equation for a homogeneous medium as equation (2.13).

$$P(\mathbf{x}) = c^2 \rho(\mathbf{x}, t). \tag{2.12}$$

$$\nabla^2 P(\mathbf{x}) - \frac{1}{c^2} \frac{\partial^2 P(\mathbf{x})}{\partial t^2} = 0, \tag{2.13}$$

in which ∇^2 is the Laplacian operator and c is the speed of the sound. Furthermore, The wave equation can be expressed in the frequency domain performing the Fourier transformation defined in equation (2.5):

$$p(\mathbf{x}, \omega) = \int_{-\infty}^{\infty} P(\mathbf{x}, t) e^{-j\omega t} dt, \tag{2.14}$$

$$P(\mathbf{x}, t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} p(\mathbf{x}, \omega) e^{j\omega t} d\omega. \tag{2.15}$$

The derivative of equation (2.15)

$$\frac{\partial P(\mathbf{x})}{\partial t} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \omega j p(\mathbf{x}) e^{j\omega t} d\omega, \tag{2.16}$$

yields

$$\mathcal{F}\left(\frac{\partial P(\mathbf{x})}{\partial t}\right) = j\omega p(\mathbf{x}). \tag{2.17}$$

The relation expressed in equation (2.17) can be used to compute the Fourier Transform of the wave equation, leading to the homogeneous Helmholtz equation

$$\nabla^2 p(\mathbf{x}) + k^2 p(\mathbf{x}) = 0, \tag{2.18}$$

where $k = \frac{\omega}{c}$ is the wavenumber, ω is the angular frequency and c is the speed of the sound.

2.1.4 Sound waves

2.1.4.1 Plane waves

A relevant solution of the Helmholtz equation is given by a harmonic function expressed as $|q|\, e^{(-jk_0(\mathbf{x}\cdot\hat{\mathbf{y}})+\angle_q)}$ where the sign of the argument of the complex exponential identifies whether it is an incoming or outcoming propagation. The vector $\mathbf{x}$ defines a field point position and $\hat{\mathbf{y}}$ identifies the direction of the wave propagation. $|q|$ an $\angle_q$ are the magnitude and phase of the wave, respectively. This solution represents an ideal sound wave with constant pressure in any plane perpendicular to the propagation vector (see figure 2.1). Even though this type of wave is not found in reality, its study is significant because it reduces the complexity of many acoustic problems. An approximation of a plane wave can be assumed at a large distance from the acoustic source, where the curvature of the wave can be ignored.

The Fourier transform of the momentum equation (2.11) establishes the following relation between the complex sound pressure and the complex particle velocity:

$$j\omega\rho_0\mathbf{u}(\mathbf{x}) + \nabla p(\mathbf{x}) = 0. \tag{2.19}$$

An important quantity in acoustics is called the specific acoustic impedance $Z(\mathbf{x}, \omega)$ and it is defined as the ratio between the complex pressure and the complex particle velocity. This complex value determines the relative amplitude and phase between

the pressure and particle velocity. For a plane wave $Z(\omega) = \rho_0 c$, which is also called the characteristic impedance of the medium.

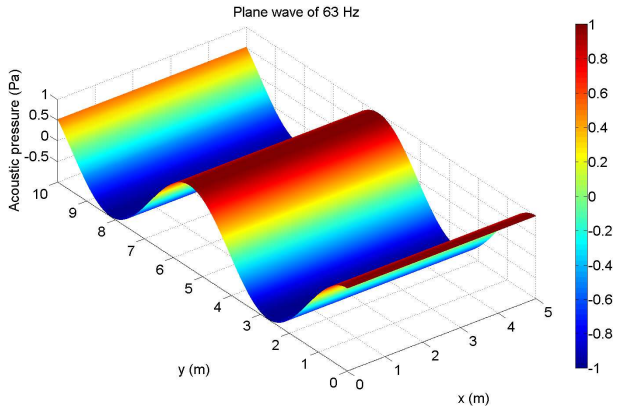


FIGURE 2.1: Instantaneous pressure in a plane wave at 63 Hz. Colours red and blue represent points of maximum and minimum acoustic pressure, respectively.

2.1.4.2 Spherical waves

Another elementary type of wave is a spherical wave (see figure 2.2). In this case, the sound pressure is a function of the radius r in a spherical symmetrical propagation, thus making it convenient to express the wave equation in polar coordinates [14]

$$\frac{\partial^2(rP(r))}{\partial r^2} - \frac{1}{c^2}\frac{\partial^2(rP(r))}{\partial t^2} = 0, \tag{2.20}$$

Assuming that the acoustic source is located at the origin of the coordinate system, the solution for a single harmonic signal of frequency ω_0 is given by[1]:

$$P(r) = \mathrm{Re}\left\{\frac{Qe^{j(\omega_0 t - kr)}}{r}\right\}. \tag{2.21}$$

Equivalently, the fundamental solution for the Helmholtz equation is

$$p(r) = \frac{q}{r}e^{-jkr}, \tag{2.22}$$

where $q = |q|e^{j\angle q}$ is the complex amplitude of the wave. In the spherically symetric case, the Fourier transform of the momentum equation (2.11) becomes:

[1] In an unbounded medium only outward waves are feasible

$$j\omega\rho_0 u_r(r) + \frac{\partial p(r)}{\partial r} = 0, \tag{2.23}$$

thus

$$u_r(r) = \frac{q}{j\omega\rho_0}\left[\frac{jk}{r} + \frac{1}{r^2}\right] e^{-jkr}, \tag{2.24}$$

hence the specific acoustic impedance is

$$Z(r,\omega) = \rho_0 c\left[\frac{jkr}{1+jkr}\right]. \tag{2.25}$$

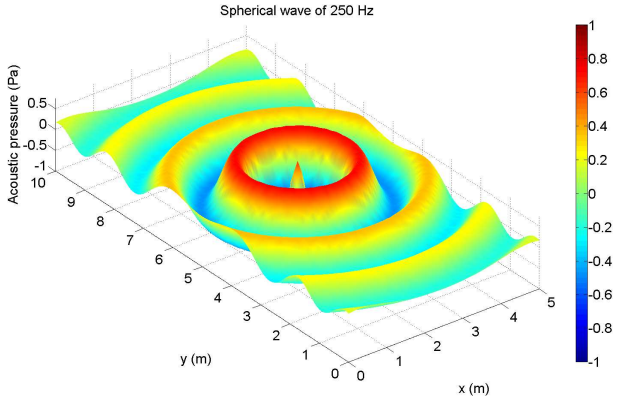


FIGURE 2.2: Instantaneous pressure in a spherical wave of 250 Hz. Colours red and blue represent points of maximum and minimum acoustic pressure, respectively.

2.1.5 Acoustic quantities

It is appropriate to average the instantaneous sound pressure $P(\mathbf{x})$ when it has a random behaviour with respect to time (e.g. noise). The mean square value of the pressure $\overline{P^2}(\mathbf{x})$ and the root mean square pressure $P_{rms}(\mathbf{x})$ are used to obtain a representative value in the considered time interval (T).

$$\overline{P^2}(\mathbf{x}) = \lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2} P^2(\mathbf{x})dt, \tag{2.26}$$

$$P_{rms}(\mathbf{x}) = \sqrt{\overline{P^2}(\mathbf{x})}. \tag{2.27}$$

However, because the range of variations in the acoustic pressure perceived by the humans may be very wide (from 20×10^{-6} Pa to 64 Pa), it is reasonable to express the pressure in a logarithmic scale. Based on that, the sound pressure level is defined as

$$L_P = 20 \log_{10} \left(\frac{P_{rms}}{p_{ref}} \right), \tag{2.28}$$

where p_{ref} is the reference pressure which corresponds to the human hearing threshold at 1 kHz (20×10^{-6} Pa for air). Although the sound pressure is an important parameter for the characterization of sound waves, other quantities can be used to analyze different properties of the waves and the sources that generate them. Among these quantities, the most important are the acoustic intensity and the sound power.

2.1.5.1 Acoustic intensity

The instantaneous acoustic intensity is the rate of energy flow through unit surface area. It can be calculated by taking the product of the sound pressure and particle velocity. Therefore, the acoustic intensity averaged over a integration time (T) is given by

$$\mathbf{I}(\mathbf{x}, t) = \frac{1}{T} \int_0^T P(\mathbf{x}) \mathbf{U}(\mathbf{x}) dt. \tag{2.29}$$

For one-dimensional plane waves equation (2.29) yields

$$\mathbf{I}(x_n, t) = \frac{|P(x_n)|^2}{2\rho_0 c}. \tag{2.30}$$

The Fourier transform of equation (2.29) leads to:

$$\mathbf{I}(\mathbf{x}, \omega) = \frac{1}{2} \operatorname{Re} \left\{ p(\mathbf{x})^* \mathbf{u}(\mathbf{x}) \right\}, \tag{2.31}$$

in which $(\cdot)^*$ indicates complex conjugate.

2.1.5.2 Sound power

The sound power is the amount of energy radiated by an acoustic source per unit time. This quantity can be characterized by integrating the intensity going out

through a hypothetical surface that encloses the source. This relation is expressed by the equation

$$w = \int_S \mathbf{I}(\omega) \cdot \widehat{\mathbf{n}} dS, \tag{2.32}$$

where $\widehat{\mathbf{n}}$ is the unitary vector normal to the surface pointing towards the exterior. In far field ($kr >> 1$), the sound power can be calculated from the sound pressure for a spherically symmetric propagation as

$$W = \frac{|P(r)|^2}{\rho_0 c}(4\pi r^2). \tag{2.33}$$

Similarly to the sound pressure, it is convenient to express the sound power in a logarithmic scale. Therefore, the sound power level is defined as:

$$PWL = 10 \log_{10} \left(\frac{W}{W_{ref}} \right), \tag{2.34}$$

in which W_{ref} is the reference power (10^{-12} Watts). Finally, the following equation specifies the relation between the sound pressure level and the sound power level for spherical propagation [15]:

$$L_P(r) = PWL - 10 \log_{10}(4\pi r^2) + 10 \log_{10}(\frac{\rho_0 c W_{ref}}{p_{ref}^2}). \tag{2.35}$$

2.1.6 Sound propagation in a bounded space

If a sound wave is travelling within a bounded space, it is necessary to take into consideration the influence of its boundaries on the propagation. When a plane wave impinges on a wall, a fraction of its energy is absorbed, some energy is transmitted through the wall and the rest is reflected into the medium. This reflected wave has a different amplitude and phase with respect to the incident wave according to the specific properties of the wall. These changes in phase and amplitude can be expressed by the complex reflection factor [4]

$$R(\omega) = |R(\omega)| \, e^{j \angle_{R(\omega)}}, \tag{2.36}$$

whose magnitude $|R(\omega)|$ and phase $\angle_{R(\omega)}$ depend on the frequency and the angle of arrival of the incident wave. In turn, the loss of energy in the reflected wave due to the boundary can be determined by the absorption coefficient α, which is related to the complex reflection factor by

$$\alpha(\omega) = 1 - |R(\omega)|^2 . \tag{2.37}$$

2.1.6.1 Specific acoustic impedance

Another relevant quantity commonly used to describe the acoustic properties of a boundary is the normal specific acoustic impedance $Z_n(\omega)$ defined as the ratio between the sound pressure p at the surface of the wall and the particle velocity normal to the wall u_n. The specific acoustic impedance normalized by the characteristic impedance of the medium ($\rho_0 c$) is called non-dimensional specific acoustic impedance z_n.

$$Z_n(\omega) = \left(\frac{p}{u_n} \right), \tag{2.38}$$

$$z_n(\omega) = \frac{Z_n(\omega)}{\rho_0 c}. \tag{2.39}$$

2.1.6.2 Relation between impedance and the absorption coefficient

The relation between $Z_n(\omega)$, $R(\omega)$ and $\alpha(\omega)$ can be derived by performing the analysis of an incident plane wave impinging on a wall with incidence angle θ. This situation is ilustrated in Figure 2.3.

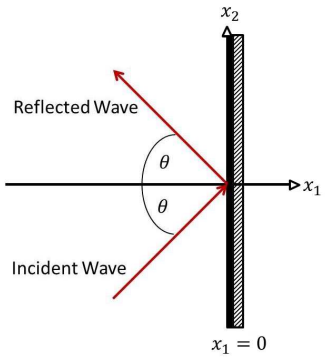


FIGURE 2.3: Diagram of a plane wave reflection

The sound pressure $p^{(i)}$ and the normal component of particle velocity u^i_{x1} of the incident wave are expressed as:

$$p^i(\mathbf{x}) = qe^{-jk(x_1\cos\theta + x_2\sin\theta)}, \tag{2.40}$$

and

$$u^i_{x1} = \frac{q}{\rho_0 c}\cos\theta e^{-jk(x_1\cos\theta + x_2\sin\theta)}. \tag{2.41}$$

In the case of the reflected wave, the sign of the component in x_1 is negative due to the direction of propagation. In addition, the reflection factor establishes the changes in amplitude and phase with respect to the incoming wave.

$$p^r(\mathbf{x}) = R(\omega)qe^{-jk(-x_1\cos\theta + x_2\sin\theta)}, \tag{2.42}$$

$$u^r_{x1} = -\frac{R(\omega)q}{\rho_0 c}\cos\theta e^{-jk(-x_1\cos\theta + x_2\sin\theta)}. \tag{2.43}$$

The total sound pressure $p^t_{\mathbf{x}}$ at the boundary ($x_1 = 0$) is found by the addition of the incident and reflected pressure. The same principle applies to the total particle velocity u^t_{x1}.

$$p^t(0) = qe^{-jk(x_2\sin\theta)}(1 + R(\omega)), \tag{2.44}$$

$$u^t_{x1} = \frac{q}{\rho_0 c}\cos\theta e^{-jk(x_2\sin\theta)}(1 - R(\omega)). \tag{2.45}$$

Therefore, the specific acoustic impedance expressed in terms of $R(\omega)$ is

$$Z_n(\omega) = \frac{\rho_0 c}{\cos\theta} \frac{1 + R(\omega)}{1 - R(\omega)}. \tag{2.46}$$

Solving for $R(\omega)$ yields

$$R(\omega) = \frac{Z_n(\omega)\cos\theta - \rho_0 c}{Z_n(\omega)\cos\theta - \rho_0 c} = \frac{z_n(\omega)\cos\theta - 1}{z_n(\omega)\cos\theta + 1}. \tag{2.47}$$

Finally, the absorption coefficient is calculated based on equation (2.37) as

$$\alpha(\omega) = \frac{4\,\mathrm{Re}\left\{z_n(\omega)\right\}\cos\theta}{\left|z_n(\omega)\right|^2\cos^2\theta + 2\,\mathrm{Re}\left\{z_n(\omega)\right\}\cos\theta + 1}. \tag{2.48}$$

2.1.7 Sound field representation

Two sound field representations that satisfy the homogeneous Helmholtz equation are reviewed in this section. Firstly, an integral representation based on the plane wave expansion is addressed. Then, a series representation through the decomposition of the sound field using spherical harmonics is considered.

2.1.7.1 Plane wave expansion

The complex sound pressure due to a plane wave arriving from the direction $\widehat{\mathbf{y}}$ at the point $\mathbf{x}$ is given by

$$p(\mathbf{x}) = q e^{jk\mathbf{x}\cdot\widehat{\mathbf{y}}}, \tag{2.49}$$

where $q = |q|\, e^{j\angle(q)}$ is the complex amplitude of the plane wave. For an infinite number of plane waves arriving from all possible directions, equation (2.49) is transformed into

$$p(\mathbf{x}) = \int_{\widehat{\mathbf{y}}\in\Omega} e^{jk\mathbf{x}\cdot\widehat{\mathbf{y}}} q(\widehat{\mathbf{y}}) d\Omega(\widehat{\mathbf{y}}), \tag{2.50}$$

in which $q(\widehat{y})$ is the amplitude density function and Ω is the unitary sphere. The representation of the sound field using a plane wave expansion (PWE) offers the main advantage of being adaptable to several audio reproduction techniques [13]. Binaural reproduction can be also performed by the convolution of the plane waves with the

HRTFs according to the direction of arrival. The relation and transformation between the PWE, Ambisonics and Wave Field Synthesis are shown in [16]. The disadvantage of this representation is due to the assumption of plane waves itself, making the approach suitable for sound fields generated by sources located a large distance from the listener ($kr >> 1$). In addition, the implementation of an infinite number of plane waves is not feasible in the reproduction stage, whose discretization generates artifacts in the sound field reconstruction.

2.1.7.2 Spherical harmonic expansion

Analogous to the Fourier transform, spherical harmonics are orthonormal functions whose weighted superposition allows a represention of a function f defined on a sphere. Any function which is integrable on the unit sphere can be expanded using spherical harmonics by [14]

$$f(\theta, \phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} A_{nm} Y_n^m(\theta, \phi). \tag{2.51}$$

The coefficients A_{nm} can be calculated as follows:

$$A_{nm} = \int_{\Omega} Y_n^m(\theta, \phi)^* f(\theta, \phi) d\Omega, \tag{2.52}$$

where $Y_n^m(\theta, \phi)$ are the spherical harmonics and $(\cdot)^*$ denotes the complex conjugate. The spherical harmonics are defined as

$$Y_n^m(\theta, \phi) = \sqrt{\frac{(2n+1)}{4\pi} \frac{(n-m)!}{(n+m)!}} P_n^m(\cos\theta) e^{jm\phi}, \tag{2.53}$$

in which P_n^m is the Legendre associated function.

The spherical harmonics are part of the solution of the wave equation (constituting the angular component) when it is expressed in spherical coordinates. The wave equation in spherical coordinates is given by

$$\frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial P}{\partial r} \right) + \frac{1}{r^2 \sin\theta} \frac{\partial}{\partial \theta} \left(\sin\theta \frac{\partial P}{\partial \theta} \right) + \frac{1}{r^2 \sin^2\theta} \frac{\partial^2 P}{\partial \phi^2} - \frac{1}{c^2} \frac{\partial^2 P}{\partial t^2} = 0, \tag{2.54}$$

whose solution, obtained by separation of variables, leads to the following expression for the interior case (where all sources lie outside the region of interest) [14]:

$$p(r,\theta,\phi,\omega) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} A_{nm}(\omega) j_n(kr_x) Y_n^m(\theta,\phi), \tag{2.55}$$

where j_n is the spherical Bessel function of the first kind of order n and $Y_n^m(\theta,\phi)$ are the spherical harmonics defined by equation (2.53). The relation between a plane wave and the spherical harmonics is determined by the Jacobi-Anger expansion [17]

$$e^{jk\mathbf{x}\cdot\hat{\mathbf{y}}} = 4\pi \sum_{n=0}^{\infty} j^n j_n(kr_x) \sum_{m=-n}^{n} Y_n^m(\theta_x,\phi_x) Y_n^m(\theta_y,\phi_y)^*, \tag{2.56}$$

The relation for a spherical wave is given by [14]:

$$\frac{e^{jk|\mathbf{x}-\mathbf{x_{src}}|}}{4\pi\,|\mathbf{x}-\mathbf{x_{src}}|} = -jk \sum_{n=0}^{\infty}\sum_{m=-n}^{n} j_n(kr_x) h_n^{(2)}(kr_{src}) Y_n^m(\theta_x,\phi_x) Y_n^m(\theta_{src},\phi_{src})^*, \tag{2.57}$$

where $\mathbf{x_{src}}$ is the position of the acoustic source and $h_n^{(2)}(kr_{src})$ is a spherical Hankel function of second kind.

2.1.7.2.1 Real-valued spherical harmonics The real-valued spherical harmonics can be estimated from their complex pairs using the relation established in equation (2.58). This set of functions are useful to generate the encoding and decoding stages of Ambisonic signals, which will be addressed in further Chapters.

$$\Upsilon_n^m = \begin{cases} \frac{i}{\sqrt{2}}\left(Y_n^{-|m|} - (-1)^m Y_n^{|m|}\right) & \text{if } m < 0 \\ Y_n^0 & \text{if } m = 0 \\ \frac{1}{\sqrt{2}}\left(Y_n^{-|m|} + (-1)^m Y_n^{|m|}\right) & \text{if } m > 0 \end{cases} \tag{2.58}$$

2.2 Literature review

In this section, a review of the scientific literature related to the current investigation is carried out. This is done by a brief introduction about the history of auralization. Subsequently, the main steps to generate an auralization are addressed taking into

consideration several approaches proposed by the scientific community. Then, techniques to create interactive auralizations are reviewed. Finally, an analysis of the topic is conducted to point out the contribution of the present research to the field.

2.2.1 History of auralization

One of the first attempts at rendering audio to obtain an audible impression of an enclosure was made by Spandöck in 1934 [18]. The experiment was based on building a scale model (1:5) of a hall where an anechoic sound was reproduced at a faster rate of 5 times. Simultaneously with the playback, the sound in the scale model was recorded using a microphone and a wax cylinder. Finally, the recorded signal was reproduced, making a correction in the playback speed, through headphones. Later, in the 60's, this methodology was used with decreased size of the scale models (1:10) [1]. In the same decade, Schroeder proposed a different method for the evaluation of concert halls before their construction. For this, an anechoic signal was treated by adding echoes and reverberation according to the simulation of the sound propagation in the hall. The reproduction of the results was performed through loudspeakers [19].

However, it was in the late 80s and early 90s when auralization began to be thoroughly investigated [1]. The main objective was to recreate the listening experience taking into account the properties of sound propagation and of the receiver. Because the auralization process requires many stages, different concepts about its meaning were formulated. The main difference between the various researches is related to the steps involved in the reconstruction of the sound field, this means, whether to take into account the estimation of the room impulse response as part of the auralization chain or not [1, 20]. This was only a matter of definition because the elements required for creating an auralization are the same in both cases. The Room Impulse Response (RIR) has to be estimated or measured according to the selected audio reproduction system (binaural or multichannel), then it is convolved with an anechoic audio signal and finally it is reproduced [21]. Nowadays, auralization may be understood as a unifying process of three stages to create audible information: the sound generation, the sound transmission, and the sound reproduction [3].

2.2.2 Sound generation

The sound generation includes the study of the characteristics of the source in terms of directivity, acoustic power, and how it must be recorded in an anechoic environment. In order to create a realistic auralization, a virtual "violin" source, for instance, should maintain the natural output level and the directional radiation of the real instrument. This means a narrow directivity lobe at the high frequencies and a wider lobe at low frequencies. Dalenbäck et al. [22] studied the perceptibility of changes in the source directivity, geometric shape and sound absorption variations to evaluate the relevance of the directivity in the auralizations. The results showed that it is clearly possible to discern these differences and therefore these affect the auralization. Similar results were found by Wang et al. [23] who focused only on the change of source directivity. The findings were conclusive only with extremely directional patterns due to the limited directivity data considered in the experiments. In addition, Hoare et al. [24] analyzed the effect of the source directivity in out-door auralizations. Based on the results of their work, they could establish that the radiation pattern has a relevant effect in the perception of auralized sound.

The directivity radiation of a sound source can be determined based on the standard ISO 10140-5 [25] or following the methodologies presented by Martin [26], Jacques et al. [27] or Fernandez et al. [28]. The general procedure consists in surrounding the source by a virtual grid, at the nodes of which the sound pressure or the particle velocity is measured in free field conditions. The registered values are plotted on a polar diagram representing the relative output of the source according to the azimuth and elevation angles. Regarding the acoustic power, this can be estimated using the standards ISO 3741 [29] or ISO 9614 [30]. This information can be easily included in the most popular room acoustic software packages [31–33].

Another aspect is the availability of sound material to be convolved with the RIR because it has to be recorded an anechoic environment. Due to the complexity of this procedure, not much data is available. Hansen and Munch [34] recorded different instruments under anechoic conditions for the Archimedes project. The instruments used were guitar, cello, brass, percussion and voice. One relevant point is that the recordings were performed with a monophonic capturing technique. Using only one microphone to record the instrument, the anechoic recorded signal contains therefore the frequency response of the source for a specific angle at a given time.

However, it is important to emphasize that the sound emission of an instrument is not homogeneous and depends on its own components and characteristics. In Figure 2.4 the frequency response is presented of a classic guitar [35]. At low frequencies, the hole is the element that produces most energy while at high frequencies is the top plate; therefore, the directivity of this instrument is not fixed in time and will depend on the musical piece being played. Recent repertoires of anechoic audio material can be found in references [36, 37].

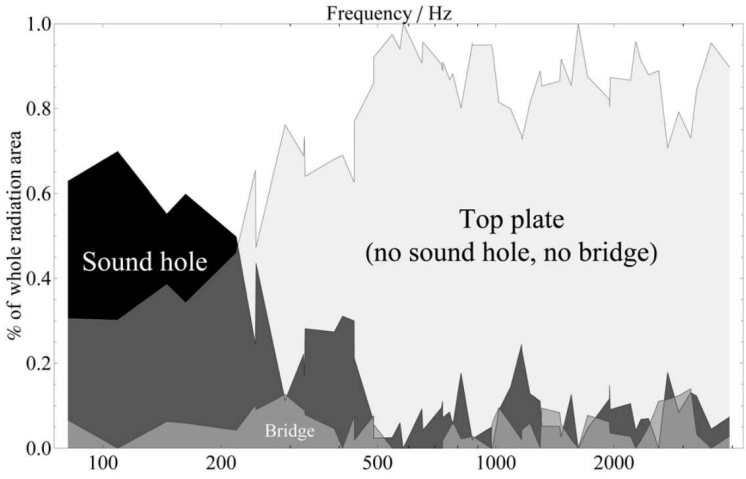


FIGURE 2.4: Frequency response of a classical guitar [35].

Due to the fact that the directivity of an instrument is not constant, Otondo and Rindel [38] proposed a new technique to determine the effects of the directivity changes in auralization. The method is based on three steps: the multichannel recording of an instrument, the simulations of RIRs using an omnidirectional source, which has been tessellated according to the number of microphones used in the recording and the convolution of each RIR with the equivalent microphone's signal. The results showed that this methodology provides a more realistic and natural auralization. A more extensive study was performed by Vigeant et al. [23] including more instruments and more channels to construct the auralizations. The outcomes were similar to the previous research showing the advantages of using this technique.

In an interactive system, it is necessary to include the directivity of a source because users should be able to listen the changes in frequency response when they are moving around the source. However, the technique proposed by Otondo and Ringel requires that the number of calculated room impulse responses must correspond to the number of microphone recordings. This involves a great amount of calculations,

whose computational cost hinders the implementation in real-time. A more feasible approach may be the use of an acoustic source whose directivity pattern has been included in the simulation [22]. Under this perspective, if the source is not omni-directional, the lateral reflections will contain less energy at high frequencies and therefore will affect the frequency response of the RIR.

2.2.3 Sound propagation

If a spatial RIR is measured using a microphone array, the properties of the source and the room are indeed in the measurement. In this case the target is to analyze the multiple signals to determine the spatial characteristics of the enclosure. Otherwise, if the RIR is synthesized, it is necessary to estimate the propagation of the sound from the source to the receiver taking into account physical phenomena like attenuation, absorption, reflection, diffraction and diffusion. Two methodologies are commonly used to predict RIRs, methods based on the numerical solution of the wave equation or based on geometrical acoustics [1, 39].

2.2.3.1 Methods based on the numerical solution of the wave equation

The propagation of the sound in a compressible medium like air is described by the wave equation. However, an analytical solution is only possible in a few cases under specific conditions, which requires the implementation of numerical methods to approximate it. The finite element method (FEM), the boundary element method (BEM) and the finite difference time domain method (FDTD) are some of the techniques that have been widely used for this purpose. The main advantage of these methods with respect to GA lies in the accuracy of the results, but the high computational cost restricts their use to low frequencies or relatively small-sized rooms.

The Finite Element Method

FEM has been extensively used since the late 60s to analyze and solve problems in acoustics [9]. The main applications in room acoustics are the calculation of the modes of enclosures, the estimation of the room frequency response and the prediction of sound transmission through porous elements, among others. The RIR is

calculated by solving the Helmholtz equation numerically for a range of frequencies and by applying the inverse Fourier transform at specific field points [40]. This requires the distribution of the problem in elementary fragments based on a mesh, the characterization of the boundary conditions that govern the system, and the solution of a system of equations [41]. The sound pressure between nodes can be obtained by an interpolation process.

Regarding auralization, FEM has been used to calculate the low frequency component of the RIR and subsequently combined with GA methods to obtain a broadband frequency response. Granier et al. [42] implemented this methodology to reconstruct the sound field of a car audio installations. Although the results were not highly satisfactory (due the lack of data for the geometrical acoustic (GA) model, the headphones equalization and the crossover between the low and high frequency components of the impulse response), FEM showed a good approximation to the measured data at low frequencies. Aretz et al. [11] obtained better results using the same approach in a regular room. The RIR at frequencies below the Schoeder frequency was calculated with FEM whilst at higher frequencies GA methods were used. In contrast with the results reported by Granier, the subjective test shown good agreement between the measured and synthesized data. Tafur et al. [12] created broadband auralizations using the same technique introduced above to calculate the RIR. The outcomes showed that the synthesized RIRs are closer to the measured one when FE simulations are included.

As mentioned above, FEM is mainly useful to calculate the low frequency component of the RIR. The reason of this limitation is related to the computational cost required to calculate the solutions at high frequencies. The discretization of the domain involves a mesh, the resolution of which depends on the wavelength; in practice it is recommended to use between 6 and 10 nodes per wavelength [40, 41, 43] to ensure convergence to the correct solution. Moreover, in a 3D problem the number of nodes grows with frequency by a power of 3 thus increasing the computation demand for larger rooms.

The Boundary Element Method

This method is based on the calculation of the pressure and the particle velocity at the boundaries of the domain. Then, these quantities are used to predict the acoustic field inside or outside the domain by applying the Green's theorem and the Green's function definition to the Helmholtz equation (Kirchhoff-Helmholtz Integral) [44]. The approach is based on the discretization of the boundary surface according to a mesh and on the calculation of the sound pressure and particle velocity at its nodes. Like FEM, the resolution of the mesh depends on frequency, thus limiting the applicability of BEM to low frequencies. Although the use of BEM entails a smaller number of equations compared to FEM, the time consumed to obtain the solutions can be higher because the matrices of equations are full and frequency dependent. [15].

There is an extensive literature about the application of the boundary element method to solve problem in acoustics. In the following, some researches regarding room acoustics applications are considered. BEM has been implemented to solve problems in acoustics associated to transmission loss, sound radiation and the sound field inside cavities or enclosures [45]. Seybert et al. [46] coupled the formulation of the interior and exterior integral equations in order to reduce the numerical errors when these two cases are contemplated together. In addition, they implemented this approach in three cases: sound radiation from a circular duct, sound propagation from a source within a semi-closed enclosure and acoustical response of a slotted cavity. The results showed the capability of BEM to calculate sound transmission for room acoustics applications.

Kopusz and Lalor [44] compared BEM and FEM by predicting the sound field inside small rectangular cavities. The results indicated a very good similarity of the two approaches. Moreover, it was shown that when it is used for coupled cavities, the type of partition between them has a large effect on the resulting acoustic field. A more recent study was carried out by Haitao et al. [47] who used the Element-Free Galerkin Method to calculate the sound field in rectangular and cylindrical cavities. BEM was implemented as a reference to validate the experiments. The simulations indicate a good similarity between the results using both methods.

Time domain formulation of the BEM has been also proposed for room acoustics applications. Hargreaves [48] investigated the applicability of time domain BEM to

model acoustic surface treatments. The results indicate that although there was a significant improvement in the algorithm used to calculate the solutions (optimization of the integration accuracy and dealing with non-rigid boundary conditions, among others), the method is in an early stage for more general applications.

A different approach has been recently proposed using a combination of the fast multipole method and BEM (FMBEM)[49]. The general concept is based on an iterative algorithm that is applied to the linear system obtained from the boundary integral equation. The matrix-vector multiplications, which are the largest computational load in the iterative process, are accelerated by means of the fast multipole method. This technique leads to a faster computation of the solution when it is compared to a classical BEM formulation. Moreover, the implementation of fast multipole methods allows large-scale engineering problems based on the boundary integral equation to be solved, which expands the applicability of BEM.

The Finite Difference Time Domain

In contrast to FEM and BEM, which are usually frequency domain methods, FDTD obtains the solution of the wave equation in the time domain directly. This is achieved by replacing the spatial and time derivatives with finite difference approximations [50]. The basic approach is based on the discretization of the medium with a temporal-spatial grid where the sound pressure and particle velocity are calculated at specific nodes.

The homogeneous wave equation in one-dimension is expressed in cartesian coordinates as:

$$\frac{\partial^2 P(\mathbf{x})}{\partial t^2} = c^2 \frac{\partial^2 P(\mathbf{x})}{\partial x^2}, \tag{2.59}$$

in which P is the acoustic pressure which depends on the space and time ($P(x_i, t_m)$ or P_i^m) and c is the speed of sound. The indices refer to the step in which the quantity is to be calculated. The discretization of equation (2.59) in temporal and spatial domain yields the following approximation [51]

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

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

$$P_i^{m+1} = \psi^2 (P_{i+1}^m - 2P_i^m + P_{i-1}^m) + 2P_i^m - P_i^{m-1},$$

where $\psi = c\frac{(\Delta t)}{(\Delta x)}$ is called the Courant number. Finally, the last equation may be re-arranged as

$$P_i^{m+1} = 2(1 - \psi^2)P_i^m + \psi^2 (P_{i+1}^m + P_{i-1}^m) - P_i^{m-1}. \tag{2.60}$$

Similar expressions can be obtained for the boundary conditions [51]. The consequence of the discretization of the medium by a space-temporal grid is responsible for the emergence of the dispersion error which is an important limitation in the use of FDTD [51]. However, several approaches to reduce it have been proposed. Savioja and Välimäki [52] compared three mesh structures: a rectangular mesh, an interpolated (rectangular) mesh and a triangular mesh. It was found that the interpolated and triangular mesh reduce dependency of error on direction of the sound propagation, thus decreasing the dispersion error. Subsequently, the last two structures were warped with a finite impulse response (FIR) filter to compensate the frequency dispersion improving the accuracy of the calculations. Finally, the results indicated that the warped triangular mesh provides the highest computational accuracy. However, the interpolated mesh provided a good improvement of the results compared with the conventional rectangular mesh, which is easier to implement. Also, Kowalczyk and Walstijn [53] analyzed the influence of using different types of grids (leapfrog, octahedral, cubic close-packed and interpolated) to discretize the medium. The results suggest that the interpolation methods have the best performance for FDTD simulations compared with the conventional meshes. In addition, as with FEM and BEM, the required resolution between nodes depends on the wavelength of the sound thus constraining its applicability at high frequencies.

The use of FDTD in auralization has been focused on the estimation of the low frequency component of RIR. Botteldooren [50] carried out simulations of the sound

pressure level distribution in a hall and the impulse response between one source-receiver position. The results showed the usability of the FDTD methods for modelling the sound propagation in the low frequency range. A more recent work was carried out by Escolano et al. [54] who implemented FDTD to calculate the RIR for wave field synthesis applications. Good similarities between the calculations and the measurements were found. Also, the combination of FDTD and GA methods to recreate the RIR has been applied. Southern et al. [55] used FDTD and beam tracing to obtain results over a broad frequency band. The main issues with this approach were the gain difference and the crossover between the two impulse responses to create a unified transfer function. The findings suggest that this hybrid method provides a good representation of the sound field at low and high frequencies.

2.2.3.2 Methods based on geometrical acoustics

The geometrical approach is founded on the assumption that sound waves can be replaced by rays or particles travelling in the enclosure with a portion of the energy produced by the source [4]. This method has been widely used from the nineties for room acoustics as it provides an acceptable accuracy in the range of middle and high frequencies. Likewise, the majority of commercial software for room acoustics design is based on these techniques and on its variations [31–33].

However, the hypothesis of sound waves propagating as sound rays presents some limitations. The interference effect usually is not taken into account; if several rays impinge on the same point in the space, their energy is added without considering their phase relations. In addition, when the wavelength of the sound is comparable with the dimensions of the surfaces on which the wave impinges, the diffraction phenomenon is relevant. Also, if the roughness of the surfaces is large compared to the wavelength of the sound, the diffusion changes the direction of the reflections. These two last phenomena are not predicted by standard GA methods and limit the usability of geometrical acoustics to predict a broad range of frequencies of the RIR. Nevertheless, some alternatives have been proposed to recreate diffusion, which improve the estimation of the high frequency content of the RIR [6].

In the basic approach, rays can be distributed in the room in a deterministic or stochastic way. The contribution of each ray over a period of time T generates the RIR. When a ray travels around the room its energy is attenuated by the distance of

travel, by the absorption of air and by the absorption coefficient of surfaces (if it impinges a wall). The last term is related to the complex reflection factor $R = |R|e^{j\angle_R}$ (section 2.1.6), which represents the variations in amplitude and phase occurring when a wave is reflected by an infinite boundary. R is a property of the wall and depends on the frequency and incidence angle of the sound wave. The absorption coefficient is defined as $\alpha = 1 - |R|^2$ and characterizes the fraction of incident energy lost during the reflection [4]. Although the absorption coefficient is dependent on both the frequency and the angle of incidence of the waves, the difficulty of obtaining information about the angular dependence leads to a simplification in its implementation assuming random incidence. The diffuse incidence absorption coefficient can be characterized using the ISO 354 [56]. This postulation is acceptable when a significant number of rays hits the wall with different incidence angles. The most common methods in geometrical acoustics are the Image Source Model (ISM) and Ray Tracing (RT), the first being deterministic and the second stochastic.

The image source method

The ISM is based on the assumption that a specular reflection may be represented as an additional source that is a mirrored version of the original source with respect to the wall that generates the reflection. This concept is explained graphically in Figure 2.5. In this case, the two red-shadow sources are related to the first-order reflections from the floor and ceiling respectively. Also, it is possible to mirror these virtual sources to emulate a higher reflection order (blue-shadow).

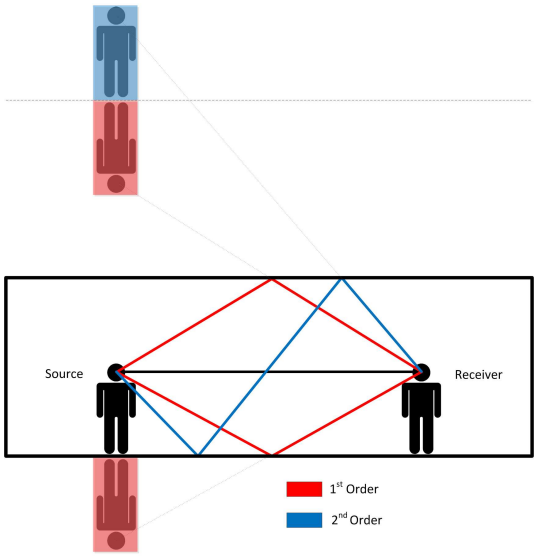


FIGURE 2.5: Image source method

In theory, it is possible to compute deterministically all the reflection paths correctly using an infinite number of virtual sources in a cuboid rigid-wall room. However, this method is only an approximation because of the effect of the finite impedance of the walls and of the finite number of sources modelling the space. In addition, the number of image sources grows exponentially with the order of reflection, thus considerably increasing the computational cost for higher orders [23, 57].

The visibility of the virtual sources is another important issue. In a cuboid room, all image sources reach the receiver point, but if the geometry of the room is more complex, it is necessary to evaluate which reflection paths arrive at the receiver and discard the others. This problem is shown in Figure 2.6 where the virtual source generated by the floor is visible to receiver 1 but not to receiver 2. The calculation of the visibility is required for each source-receiver path making this process the most time consuming part of the ISM algorithm. Because of this, some techniques have been proposed to improve the visibility algorithm of the image source method and accelerate its computation [58].

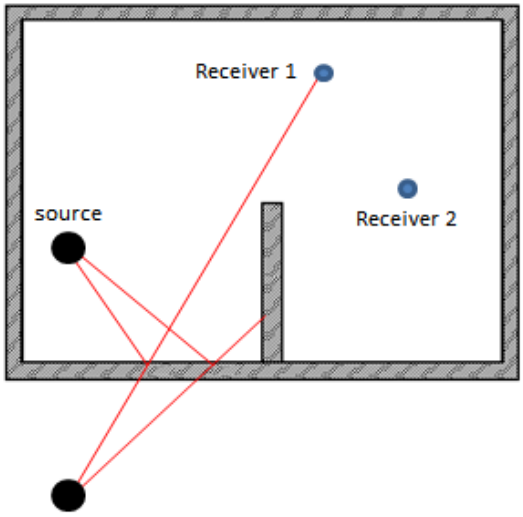


FIGURE 2.6: Visibility of image sources

Furthermore, the restriction of ISM to calculate only specular reflections makes this method not suitable to represent the last part of the RIR where the scattered energy predominates [59]. Moreover, when the roughness of the walls is larger than the wavelength, the specular assumption is no longer accurate due to the diffusion effect. Regarding diffraction, some methods have been proposed to include this phenomenon in the model. Pulkki et al. [8] used virtual sources to represent the diffraction effect due to edges. The basic procedure consists in the determination of polygons (surfaces) which are connected together. Then, the diffractive edge is simulated implementing additional image sources.

ISM has been used in relation to auralization since the seventies when Allen and Berkley [60] implemented an algorithm to compute the reverberation time in a rectangular room. Based on this approach, they were able to estimate the impulse response of the room and auralize the result. However, the model presented the limitation of simulating only rectangular rooms. Other issues were the assumption of angle and frequency independent absorption coefficients. In the eighties, Borish [61] extended the algorithm of ISM to arbitrary polyhedra making it suitable for any room geometry. Some techniques to deal with the visibility of image sources and the obstruction due to e.g. balconies were proposed too. From this time, in general, the research on ISM was focused on improving the algorithm to reduce its computational cost [58, 62]. Furthermore, ISM has been complemented with ray tracing in order to compute more accurately the last part of the impulse response [63, 64]. Another

interesting application was proposed by Savioja et al [39] who used ISM to represent the early part of the RIR. Under this hypothesis, they created an interactive system called DIVA (p. 43), which produces auralizations in real-time. Regarding the RIR, only the direct sound and the early reflections are calculated using an ISM. In contrast, the reverberation tail is recreated using a different approach based on the assumption of diffused and incoherent reflections by the implementing a feedback delay network.

Ray tracing

This approach assumes that the sound produced by a source can be modelled as rays or particles radiated from the source's position at the same time and in different directions. Under this concept, the source is tessellated in a finite number of rays propagating at the speed of sound and each of them carries a fraction of the total energy. In an omnidirectional source the rays are randomly dispersed on the sphere following a uniform distribution [65]. Also, if a directivity function is required, this can be achieved by distributing the particles in a particular region or by weighting the energy of the rays according to their direction of propagation [3]. Then, each ray is propagated into the room until a defined time has elapsed or the energy associated with the ray reaches a minimum threshold. The ray loses its energy because of air absorption, geometrical divergence and by the absorptive properties of the boundaries.

In classical Ray Tracing (RT) the rays are propagated following the GA rules, this means producing specular reflections when the rays hit a boundary. However, the relevance of including diffusion in the estimation of the RIR has been proved [6, 66], which led to the proposal of techniques to recreate this phenomenon. The most general approach is the inclusion of the random incidence scattering coefficient which can be understood as the ratio between the energy scattered in non-specular direction and the energy reflected specularly. The procedure to measure this coefficient is stated in the ISO 17497 [67]. The numerical implementation is commonly made in the following way: when a ray hits a wall, a random number it is generated (between 0-1) and compared with the scattering coefficient. If this number is higher a new direction is established using two more random numbers based on a directional distribution, e.g. Lambert's law, otherwise the reflection remains specular [3].

Because of the stochastic nature of the ray tracing algorithm, a detector is required to count the rays at the listener position. The rays passing through or near to the receiver's position are therefore counted by a detector which may be a plane or a volumetric form. The information regarding the time of arrival, the energy of the ray and its direction is usually stored. The size of the detector has a relevant influence on the results because if it is too small the rays cannot hit it (discarding reflection paths). In contrast, if its size is too big it is possible to create invalid paths as shown in Figure 2.7 where the detection zones capture information about reflections which do not reach the receivers. Also, big detectors may generate overshadows to other receivers [61]. Different criteria have been proposed to treat the size of the detector, one approach being a relation which depends on the number of rays. According to Lehnert [68] the radius of an omnidirectional detector (sphere) can be calculated by

$$r_i = ct_i\sqrt{\frac{2\pi}{N}}, \tag{2.61}$$

where i indicates the index for a given ray, r_i is the radius of the detector for the ray i, c is the speed of sound, t_i is the time that the ray $_i$ takes to arrive at the receiver and N is the number of rays. From equation (2.61) it is evident that the size of the detector depends on each ray and will be different for each of them. However, the dependency of the number of rays related to the volume of the room has not been taken into consideration. Based on this, Xiangyang et al. [69] proposed the two relations below to establish the size of the detector.

$$r = \left(\frac{15V}{2\pi N}\right)^{1/3}, \tag{2.62}$$

$$r = \eta d_{sr}\sqrt{\frac{4}{N}}, \tag{2.63}$$

where V is the volume of the room, N is the number of rays, d_{sr} is the distance from the source to the receiver and $\eta = \log_{10}(V)$ is an empirical factor suggested by the authors. The former relation takes into account the volume of the room (equation (2.62)) and the latter the volume of the room and the distance between the source and the receiver (equation (2.63)).

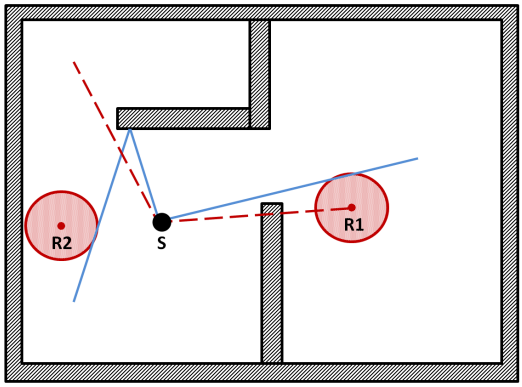


FIGURE 2.7: Invalid reflection paths captured by the detectors

As RT is a stochastic process, the number of rays has a significant effect on the prediction of the RIR [3]. Lehnert [68] also showed that the optimal amount of rays depends on the size of the room. If the number of rays is low the RIR will not include enough reflections, thus creating an inaccurate prediction. By increasing the size of the detector it is possible to reduce the number of rays but this solution raises the risk of invalid path detections. This is a crucial element in the RT algorithm because the probability of detection improves by increasing the amount of rays but at the expense of a higher computational cost. Different improvements have been proposed such as the use of similar algorithms like beam tracing or pyramid tracing.

Notwithstanding its limitations, RT has been widely used for decades to predict the RIR. One of the earliest implementations was made by Krokstad et al. in the sixties [70] who estimated the reverberation time in a rectangular and a fan-shaped enclosure. In the following decades, RT continued to be used to calculate different acoustic parameters in rooms [71, 72]. Also, some improvements in the algorithm were made such as including frequency-dependent absorption coefficients and transparent detectors (tackling the problem of shadows between receivers). In the eighties and nineties, the implementation of deterministic ray tracing methods [73] and the inclusion of scattering coefficients [6] were the most important innovations to RT.

Deterministic ray tracing

This method is a variant of the classical ray tracing, the main difference being that instead of rays, the sound propagation is represented by divergent geometric entities

such as cones or pyramids from the source. In this way, it is possible to use a specular ray tracing process to check the visibility of the image sources if a combination of the algorithms is intended. Another difference is the possibility to use a point receiver instead of a volume detector improving the problem of invalid detections. Nevertheless, the divergent geometric element used to represent the sound propagation also has an effect on the algorithm. When using cone tracing, a correction is required due the overlap between beams when the source is tessellated. This is solved using triangular or pyramid elements [3]. However, other problems must be solved like the continuity of the beam when this hits two or more surfaces simultaneously (creating invalid detections) and the direction of the beam after a reflection (discarding valid detections) [73]. Currently, some commercial packages for room acoustics are available that implement deterministic ray tracing methods for the computation of RIRs [31, 74].

Combination of ISM and RT

ISM can predict all the specular components but the computational cost required for the higher reflection orders restricts its applicability to the early part of the RIR. On the other hand, ray tracing algorithms can model the later part where it is desired to estimate a diffused behaviour with a significant number of reflections. Given the arguments above, it is reasonable to use these two techniques to calculate different parts of the RIR.

The first approach of this combined technique was proposed by Vorländer [63] who demonstrated the feasibility of using RT to validate the visibility of image sources. With this new methodology Vorländer could improve the accuracy on the estimation of the RIR and decrease the computational cost required by the ISM algorithm. Lewers [73] applied the combination of a deterministic ray tracing (pyramid) and a radiosity to estimate the RIR. The triangular (pyramid) tracing was used to find the paths between image sources and receiver, then a radiosity method was implemented to recreate the diffused reverberant tail. A more recent application was proposed by Lehmann and Johansson [57] who developed an algorithm to simulate the last part of the room impulse response based on the ISM. In their work, they calculate only the early part of the RIR using an ISM. The late part is emulated with decaying random noise whose envelope is calculated from the energy decay curve of the ISM.

The results suggest a good correlation with the classical ISM enabling a significant reduction in computation time.

A comprehensive overview of geometrical acoustics methods has been recently published by Savioja and Svensson [75]. The review includes the state of the art of this family of techniques, including the modelling of edge diffraction.

2.2.3.3 Hybrid methods

Taking into consideration the advantages and disadvantages of the two simulation methods described above, it is sensible to combine them according to the frequency range of interest over which each is most effective. At low frequencies, the response of the enclosures is usually dominated by isolated modes and the wavelength of the sound is comparable to the size of the surfaces, which increases the relevance of diffraction. These effects cannot be predicted by GA yielding to the implementation of techniques such as FEM, BEM or FDTD. At high frequencies, the computational cost of the methods based on the numerical solution of the wave equation makes their application impractical; however, GA assumptions are valid providing reliable results. Hybrid approaches using FEM and GA [11, 12] or FDTD and GA [50, 54, 55] can be found in the scientific literature. However, the consolidation of the method, the choice of modelling parameters and the unification of the results are still an open topic for research in more general applications.

2.2.4 Sound reproduction

The sound reproduction stage corresponds to the presentation of the audio material to the listener. The main objective is to recreate the three dimensional hearing experience based on the acoustic properties of the simulated or measured room. Two methodologies are frequently implemented for this purpose, multichannel and binaural techniques. The multichannel methods aim to reconstruct the sound field in a controlled area surrounding the listener, whilst in the binaural approach the goal is to reconstruct the acoustic pressure at the listener's ears. In the following, a review of two techniques commonly used to reproduce auralizations is considered: Ambisonics (multichannel) and binaural synthesis.

2.2.4.1 Ambisonics

Ambisonics is a multichannel sound reproduction technique based on the expansion of the sound field using spherical harmonics. Its development began in the early seventies when Gerzon [76] and other researchers applied the theory of this orthonormal basis to capture and reproduce sound fields. These early implementations were based on the first order of the series (see equation (2.55)), which in the recording step corresponds to a four coincident microphones signals, one being the omnidirectional signal (W) and other three orthogonal figures of eight representing each coordinate axis (X, Y, Z). These four signals are usually referred to in the scientific literature as the B-format. As in any series expansion, an exact reconstruction is only achieved when an infinite number of coefficients is considered. Nevertheless, this is not possible in a practical implementation which generates artifacts in the sound reproduction.

The errors in the sound field depend mainly on the frequency of the signal and the spatial location of the listener. More precisely, the accuracy on the reconstruction of a sound field over a specific region using Ambisonics is influenced by the number of spherical harmonic coefficients considered, which in turn is limited by the number of loudspeakers available. Work conducted by Ward and Abhayapala [77] has suggested the following relation between the number of loudspeakers L, the wavelength λ and the radius r of a sphere where the reconstruction is accurate

$$L > ([e\pi\frac{r}{\lambda}] + 1)^2, \tag{2.64}$$

in which e is Euler's number. Equation (2.64) is commonly simplified as [78]

$$\sqrt{L} - 1 >= kr, \tag{2.65}$$

where L is the number of loudspeakers, k is the wavenumber and r is the radius of a sphere in which the reconstruction is accurate. In general, the use of a finite number of coefficients produces a trade-off between the highest frequency and the region where the accuracy of the representation is acceptable. To improve the sound reproduction based on Ambisonics, different methods have been proposed to overcome its limitations. Gerzon [79] for example, developed a meta-theory based on psychoacoustic concepts which have been used for the optimization of Ambisonics decoders [80, 81]. The use of High Order Ambisonics (HOA) has increased the spatial resolution and the accuracy at high frequencies. Nowadays, HOA is widely

used to capture and synthesize sound fields [78, 82, 83]. Circular or spherical microphone arrays based on HOA expansion allow spatial information to be obtained on the environment in comparison with a monaural impulse response [84]. In addition, beamforming allows for a more robust analysis of the spatial characteristics of the enclosure. Because spherical microphone arrays are based on spherical harmonics, the recorded information is compatible with Ambisonics enabling a straightforward auralization of the environment.

Likewise, the compatibility of Ambisonics with other sound reproduction techniques has been addressed by the scientific community. Noisternig et al. [85] proposed a method to generate binaural signals from Ambisonics. The concept is based on the use of a virtual loudspeaker array to decode the Ambisonic components and perform the convolution between the signal of each loudspeaker with the Head Related Transfer Function (HRTF) corresponding to the direction of each loudspeaker. Finally, the convolved signals are superimposed to create the binaural output. Some advantages of this approach are the independence between the number of sources to encode and the number of HRTF filters and the simplicity of user's head rotation which is performed by rotation matrices in the Ambisonics domain. Nishumura and Sonoda [86] developed a technique to generate individualized B-format signals for binaural applications. Instead of taking the zero and first orders directly from the HOA representation (conventional method), all the coefficients are used to compute the binaural signals. Then, an individualized B-format is created from the binaural signals through the pseudo-inverse of the HRTF matrix. A comparison with the conventional method showed that the interaural phase differences are closer to the original HRTF but the signal-to-noise ratio is lower. This shortcoming may compromise the performance of the system in the reproduction stage.

Enzner and collaborators [87] focused on the rotation of the listener to compare two different methods for binaural rendering based on Ambisonic data. On the basis of decoding the Ambisonic signals in a set of virtual speakers, the user's rotation was computed through two different methods. The first corresponds to a sound field manipulation performed by the modification of the sound field coefficients using a rotation operator in the spherical harmonic domain. In the second approach, a plane wave expansion is performed and each signal is convolved with a high resolution HRTF according to the orientation of the listener. The outcomes indicate that the two methods have similar performance in terms of localization, but other factors

such as the relevant amount of memory required in the second case can constrain its implementation.

2.2.4.2 Binaural technology

The binaural technology is based on the concept that recreating the acoustic signals at the listener's eardrum will generate the same auditory perception as the real sound event. This is possible because these input signals contain the information about the source, the environment and the listener himself [88]. When a source generates sound within an enclosure, the listener perceives the direct sound and the reflections caused by the room according to their relative positions. In addition, the signals arriving to him/her are modified by his/her torso, shoulders, head and ears. The variation in the incoming signals produced by the listener according to the incidence angles of the waves can be described by a HRTF, which is the transfer function that describes the transmission of the sound coming from a source located at a certain distance and direction to the listener's eardrums (also it is suitable to use the entrance of the ear canals). If the source is far enough away it is possible to assume a far field condition and neglect the distance dependence [88]. The difference between the signals at both ears due to the HRTF generates the cues used by the auditory system to localize sound sources.

There are several methods to estimate a HRTF including measurements and modelling. The measurements are usually carried out in an anechoic environment discretizing the elevation and azimuth axis in several angles surrounding the listener who may be a person or an artificial head. If a human is tested, the HRTF is measured with small microphones placed at the entrance to or inside the ear canal. This method provides the best results for this particular person but not for general purposes because the HRTFs diverge among individuals. The implementation of an artificial head offers the possibility to standardize the information, which can be favourable for practical applications such as auralizations with large number of users and comparisons between research results. The drawback of using a generic artificial head lies in the mismatch compared with the user's HRTF, leading to problems like front-back ambiguity or apparent elevation of the source. Investigations made by Møller et al. [89] showed that the use of non-individualized HRTFs increases the error in the source localization and its perceived distance, mainly in the median plane.

Because the HRTFs are measured for a finite number of angles, it is usually necessary to use interpolation methods to estimate the information between measured points. In this case, the spatial resolution (distance between measured points) is relevant because it affects the quality of the prediction. Langendijk and Bronkhorst [90] investigated the spatial resolution required for using a linear interpolation without producing audible distortions. The test was based on the reproduction of broadband noise (flat and scrambled-spectrum) using interpolated data obtained from three sets of HRTFs with a spatial resolution of 5.6°, 11.3° and 22.5°, respectively. The results indicated that a resolution of 5.6° is required to provide an adequate representation of virtual sources (although in the case of scrambled-spectrum signal it is possible to use a resolution between $10° - 15°$ resulting in similar performance). Experiments carried out by Breebaart and collaborators [91] suggest that a spatial resolution of 10° is enough for high accuracy in binaural processing. A threshold of 5° in the spatial resolution was reported by Hartung et al. [92] when it is desired to generate a smooth rendering of sound source movements or head rotation for interactive applications. The same resolution has been also recommended by other authors [93].

A related study was done by Zhong et al. [94] who evaluated the spatial symmetry of the HRTFs in the azimuth and the elevation axis. The reason for this research is supported by the postulation that if the HRTFs are symmetric it is possible to reduce the amount of data. The left-right, back-front and median plane were evaluated in 52 human subjects using the asymmetry coefficient (representing the level of asymmetry) calculated from a cross-correlation process. The findings indicated that it is feasible to assume symmetry in the HRTF models up to 5 kHz. From this threshold this assumption has to be avoided because the asymmetry is relevant and changes the characteristics of the HRTFs. Finally, some research has been focused on the evaluation of different interpolation methods to increase the spatial resolution of the HRTFs. The review of these techniques will be addressed in the section of Direct RIR rendering (2.2.5.2).

A binaural signal is frequently reproduced via headphones or loudspeakers with crosstalk cancellation. In the first case, the main advantages lie in the practicability of the implementation and the portability of the system itself. However, these benefits are at the expense of reducing the immersive experience due to the use of transducers directly on the head of the listener. Another limitation is that headphones do not account for movements of listener's head, unless head tracking is used.

Furthermore, the non-linearity of the frequency response of the headphones leads to spectral colouration of the auralized signal that has to be compensated by means of inverse filters [88]. Schärer and Lindau [95] evaluated several methods of equalization for binaural signals. The experiments revealed important differences in the transfer function of the same pair of headphones using the same subject. These differences were attributed to the coupling of the transducers with the listener. Moreover, it was found that individual headphone equalization allows for a more realistic auralization compared to a generic equalization; this finding was also supported by later experiments [96]. Nevertheless, a generic equalization is preferable than reproducing the auralized signal without any compensation [95–97]. An interesting feature of the binaural technology using headphones is its application in mobile devices. Nowadays, the hardware resources of smartphones allow rendering of binaural environments taking into account some properties such as listener's orientation [98]. Although, these implementations are at an early stage, the continued development in the mobile device industry and the potential of binaural rendering opens a wide area for research [99].

In contrast to headphones, which deliver a different signal to each ear, the binaural reproduction by loudspeakers is corrupted by the crosstalk between the signals at the two ears of the listener. This means that the signal of the loudspeaker used to stimulate the left ear is also perceived by the right ear and vice versa. The crosstalk between signals affects the binaural rendering and therefore must be removed. Inverse filters that compensate for the crosstalk are implemented for this purpose [88]. Several methods have been proposed to generate the inverse filters and regularize the inverse matrix when it is ill-conditioned. Compared to headphones, Crosstalk Cancellation (CTC) allows listening tests to be conducted without the inherent bias that occurs when a real source is collated with a synthesized source using headphones (the listener has to remove the headphones to hear the real source). However, some disadvantages are the narrow sweet spot and the loss of signal amplitude due to the filtering process. In addition, as is common in all binaural reproduction systems, the use of individualized HRTFs plays an important role in the realism of the rendering [100]. Recent implementations of CTC are focused on improving the performance of the sound field rendering [101], the expansion of the listening area and real-time processing for virtual reality environments [102, 103].

2.2.5 Interactive auralization

Virtual reality systems that allow the user to have a multisensory experience are being widely investigated because their potential application in areas such as entertainment, education, subjective evaluations, etc. Regarding auralization, this technique enables the user to hear the sound field of a specific space but also allows him/her to interact with the environment. This means to provide the ability of movement within the virtual environment, reconstructing the acoustic field according to the position. The development of a real-time system requires a heavy computational load because the RIR has to be calculated in real-time based on the source/listener locations and then convolved with audio material recorded in an anechoic environment [39]. In addition, if a binaural reproduction is performed by headphones or using loudspeakers, the binaural RIR should take into account the user's rotation head and the changes of the HRTF with respect to the angles of incidence of the sound waves [20].

In general, a real-time auralization system should consider the following issues [104]: rotation and translation of the receiver, rotation and translation of the source, variations in the directivity of the source, changes in the room geometry and its acoustic properties. Because of the complexity of this process, some researchers have made simplifications that allow the generation of real-time systems satisfying certain conditions previously mentioned. One approach is to divide the impulse response in different parts and render each of them with a different method. Lehnert and Blauert [20] suggested to compute only the early reflections using the ISM method. The late part of the RIR is not calculated but instead, a statistical approach under the diffuse field assumption is implemented to recreate the reverberation tail. The technique of separating the RIR into two parts has been commonly used for real-time sound field computation where only the early part is calculated using different GA approaches.

Currently, it is possible to use algorithms to model in real-time specular reflections, diffused reflections and edge diffraction generating interactive auralizations. However, these algorithms are based on certain assumptions (e.g. GA methods or uniform theory of diffraction) and therefore they are not accurate in all situations. In addition, the real-time implementation demands simplifications of the number and order of reflections used to represent the early part of the RIR, thus decreasing the accuracy of the auralization. The two main methods commonly used to generate RIRs in real-time are parametric RIR rendering and direct RIR rendering [39].

2.2.5.1 Parametric RIR rendering

Using this approach the RIR is calculated in real-time according to the source, the receiver and the enclosure. However, due to the difficulty of this procedure, some simplifications and assumptions are required to enable the computation in real-time. Usually, only the early part of the RIR is calculated. The late part does not use convolution but instead, another reverberation synthesis method such as feedback delay network is implemented. Other perceptual-based approaches have also been implemented for interactivity purposes [105, 106].

One of the first attempts to create an interactive auralization was proposed by Furlong et al. [105] who suggested the use of a simplified RIR based on the research of Yoichi Ando about the acoustical response listener preference. Instead of focusing on a detailed RIR, they tried only to match the sound pressure level, the temporal distribution and energy of the early reflections, the reverberation time, and the interaural cross-correlation. The reason for this assumption was that, according to Ando's research, these are the most significant objective components for subjective preference. Although the paper describes the method to calculate these parameters, no results were presented related to the effectiveness of this simplification.

Another perceptual approach was developed by the Institut de Recherche et Coordination Acoustique/Musique (IRCAM) to change the sound field interactively [106]. The Spatialisateur is a spatial sound-processing software which allows the energy and spectrum of the direct sound and of early reflections to be manipulated. The ratio between the different parts of the RIR can be changed as well. Although the software provides an intuitive interface to recreate the sound field according to the user's desires, a direct calculation of the RIR is not performed.

A different method was applied by Savioja et al. [39] in the late 90's with the development of DIVA (Digital Interactive Virtual acoustics). DIVA is formed of different tools whose combined use allows one to generate the visual and auditory impressions of an interactive virtual environment. The auralization is carried out using the classical ISM for the calculation of the first order reflections, or second order depending on the complexity of the enclosure, and a simplified feedback delay network algorithm to recreate the reverberation in the room. The parameters used for modelling the reverberation are extracted from a FDTD/RT hybrid calculation previously computed.

Several strategies were adopted in order to improve the performance of the system concerning to real-time processing. The visibility of the ISM was done following two techniques. Firstly, only the surfaces that are visible to the source are taken into consideration for the computation of the image sources. Secondly, using ray tracing (previously computed) they statistically verify the visibilities of all surface pairs, thus reducing the number of possible image sources. Using these strategies they could generate an update rate of 20 Hz, which is enough for real-time purposes. The main restrictions of the system are the limitation in the number of image sources used to reconstruct the early part of the RIR, the calculation of only specular reflections, and the absence of diffraction.

Funkhouser and co-workers [107] used the beam tracing algorithm in 2004 to generate auralizations in real-time. The greatest improvement in relation to the work of Savioja et al. was the inclusion of diffraction. The creation of an interactive auralization is carried out in four phases. The first two steps are performed off-line and correspond to the spatial subdivision of the environment and the generation of a beam tracing tree. The spatial subdivision is made by dividing the space into convex polyhedral cells and by storing their relations using winged-pair data structures. This method allows for a faster calculation of the propagation paths during the beam tracing stage. After the spatial subdivision is concluded, a beam tracing technique is executed to find the propagation paths of a stationary source in the space. The beams are classified into transmission, specular, or diffraction according to the type of cells and stored in a beam tree for further use. The next two steps are the path generation and the auralization, which are completed in real-time. The path generation depends on the receiver location and use the beam trees to identify the propagation sequences in its position. Finally, the auralization is created using all the contributions made by the beams and a statistical approximation of the late reverberation. This methodology was tested for different room models and showed an improvement compared with the use of ISM in terms of computational cost. However, the system is only able to produce 8 reflections for a frequency update of 6 Hz, which is low for interactive rates [108]. With 4 reflections the system generates paths in an interactive rate. The main limitations are the use of only planar polygons in the model, the limited number of reflections for interactive applications and the absence of diffusion.

Noisternig et al. [109] also implemented a beam tracing method in 2008 to create real-time auralizations. Although the algorithm is not able to model the phenomena

of diffraction or diffusion, it was developed as open source and is available to everyone. The reproduction of the auralizations is made using HOA. The system is the integration of 4 independent units: 3D geometric scene, acoustic modelling, spatial audio encoding, and spatial audio decoding, running together to generate an interactive environment. As in the approach used by Funkhouser et al., the beam tracing is applied to create a beam tree which is useful to determine the reflection paths. The validation of the algorithm was performed by comparing the image sources and some room acoustic parameters produced by the software and a reference (Catt-Acoustics). The outcomes showed good agreement between results but the necessity was also evident of modelling diffusion. Regarding the performance of real-time applications, the system was able to recreate a 3rd order reflection path in a room (235 polygons) with an update frequency of 42 Hz for a predefined path around the source. For free movements in the same enclosure the update frequency was 71 Hz for 1st order of reflections and 3.5 Hz for 2nd order.

A different GA approach was applied by Chandak et al. [110] in 2008 who used an adaptive 4-side frustum (pyramid tracing) to compute the propagation paths in several environments. The model is able to generate paths due to specular reflections taking into account also the diffraction effect. The basic idea is to divide the pyramid using a quad-tree structure in an adaptive mode when there is an intersection with a polygon. This subdivision is useful for example when a pyramid partially hits a polygon, in which case the specular pyramid may not contain the correct reflection volume, generating an error which can be reduced using the subdivision technique. The experiments showed that it was required to use a subdivision of order between 4 or 5 to obtain results similar to the reference (Catt-Acoustics). Also, according to the complexity of the environment (e.g. 174 polygons), the algorithm can generate, using a 2nd order of reflection and 4th order of subdivision, a frequency update of 15 Hz which is applicable for real-time purposes. The limitations of the approach are related to the lack of diffusion in the model, the application of the uniform theory of diffraction, which is only valid for long edges (compared to the wavelength), and the update of the information without interpolation, which may create artifacts in the auralization.

An extension of the work done by Chandak and colleagues was made by Taylor et al. [111] in 2009 who developed the RESound system using different GA methods to generate an interactive auralization. The basic approach is to divide the RIR into two

parts: the first part (early part) takes into consideration specular reflections, diffusion and edge diffraction. To achieve this, the system uses two different techniques to propagate the reflections into the room. Specular components and the edge diffraction are computed using an adaptive frustum tracing and the diffusion is simulated with a discrete RT. The late part of the impulse response is recreated based on Eyring's reverberation equation [112]. The performance of the system depends on the number of subdivisions in the frustum and the order of reflections. The results showed that to achieve an equivalent accuracy when finding specular paths compared with the classical ISM, it is necessary to use a subdivision of 5. Similar results were found for the computation of edge diffraction. Under this condition, a calculation with 3rd order reflections takes 359 ms in the simplest enclosure (a room). For the RT (in the same enclosure), the time required for the computation is 274 ms. These values are too long for interactive purposes, so it is necessary to reduce the order of reflection and the subdivision of frustum if a real-time implementation is desired. The main limitations of REsound are related to the assumptions made in the algorithm. The GA applies only in a specific range of frequencies when the wavelength of the sound is smaller than the polygons of the room but larger than the roughness of the walls' materials. Also, the uniform theory of diffraction is only valid when the edges are larger than the wavelength of the sound. Finally, the computational cost required to generate a high order frustum subdivision and high order reflections leads to a reduced accuracy of the simulations in real-time applications.

2.2.5.2 Direct RIR rendering

This approach is based on the pre-computation of the RIRs for different listener positions and the interpolation between them in real-time according to the user movements. The implementation of this method is commonly done with binaural simulations to compensate for the head's rotation. Also, it has been used to reconstruct the sound field due the listener movement within the enclosure. The main limitation of this methodology is related to the fact that the RIRs are computed off-line. This means that the properties of the room, the source and the receivers cannot be changed unless another simulation is performed.

One early implementation of this technique was carried out by Reilly and McGrath [113] who performed a real-time auralization enabling the movement of the user's

head. The procedure consisted in the pre-computation of the Binaural Room Impulse Response (BRIR) at 128 points corresponding to 32 azimuth angles with 4 different elevation angles. In the real-time step, the tracking of the head's orientation was made using a transmitter-receiver system based on polarized magnetic fields. The auralization was executed using the BRIR related to the closest azimuth and elevation index. To avoid artifacts in the signal due to the switching between the RIRs a cross-fading algorithm was implemented to smooth this transition. Although no detailed information about the HRTF was reported, the results indicated the feasibility of using this technique to generate auralizations in real-time.

Because the interpolation is the core of the direct RIR rendering, the different aspects related to its implementation and its role with moving sources are reviewed in more detail below. Although generally the literature is focused on head's rotation, this can be useful for the generation of interactive auralizations if the user's turning is understood as a rotation of the enclosure as well. Basically, the interpolation is a method to predict unknown values from a set of discrete known values. The term interpolation means that the predicted data are within the range of the known values.

A comparison between interpolation techniques for the prediction of HRTFs was made by Hartung et al [92]. The research was focused on evaluating the performance of the Inverse Distance Weighting (IDW) and the spherical spline interpolation methods in the time and frequency domains. The study consisted in the measurement of the HRTFs using an artificial dummy head. The resolution of the measurements corresponded to 5° and 7° in azimuth and 10° in elevation (between $-70°$ and 90°). After that, they extracted the data corresponding to steps of 15° in azimuth and used it to predict a new HRTFs with 5° of resolution. The interpolations were carried out in the time and the frequency domain comparing the monaural level, the Interaural Time Difference (ITD) and the Interaural Level Difference (ILD) of the interpolated data to the measured data. Also subjective tests were carried out to determine if changes were perceived between the interpolations and the measurements. The results indicate that the interpolation techniques are able to recreate the ITDs independently of the domain in which the interpolation is performed. However, the frequency domain provided the best performance in terms of the monaural level and the ILDs. In all the experiments the spherical spline method was more accurate than the IDW. The listening tests revealed that although interpolation methods allow for an accurate representation of the HRTFs (in terms of magnitude, ITD and ILD),

users were able to identify the measurement from the interpolation. A relevant point in this research was the time required for calculating the interpolations, which was extremely high for real-time purposes.

In order to improve the efficiency of the interpolation process for real-time applications, Freeland and co-workers [114] proposed a method based on the Inter-Positional Transfer Functions (ITPF) and its simplification using Balanced Model Reduction (BMR) techniques. The ITPF is defined as the ratio between two HRTFs (same ear) representing the final and the initial point of a moving source ($IPTF_{i,f} = \frac{HRTF_f}{HRTF_i}$). The selection of the ITPFs is supported by the possibility of simplifying them to reduce the computational cost. The assumption in this method is that the measured HRTFs are close enough to allow the ITPFs to be represented by a low order model. The interpolation is made using the three closest HRTFs (triangular configuration) to the target point; however, two of them can be approximated by low-order IPTFs decreasing the computational cost. To evaluate the efficacy of the approach, they compared the results with the bilinear interpolation technique which had already been implemented in real-time audio applications [39]. The results showed that the triangular interpolation is more efficient than the bilinear interpolation. In addition, the simplification made with the low-order of IPTFs enables a higher reduction in the computational cost without sacrificing the accuracy at low and mid frequencies.

With the same spirit of the work done by Freeland, Matsumoto and collaborators [115] compared three different interpolation methods to predict a non-static source around the listener. The main difference with Freeland's research was the inclusion of a correction parameter due the variation in the arrival times when the source is moving [116]. A linear interpolation, a discrete Fourier transform and spline methods were compared. The assessment of the interpolation accuracy was made using the Signal to Distortion Ratio (SDR) which is defined as:

$$SDR = 10 \log \frac{\sum_{n=0}^{N-1} h^2(n)}{\sum_{n=0}^{N-1} [h(n) - \tilde{h}(n)]^2}, \tag{2.66}$$

where $h(n)$ is the measured binaural response, $\tilde{h}(n)$ is the interpolated bianural response and N indicates the number of response samples. Hence, a bigger SDR represents a better estimation. The outcomes indicate that the linear interpolation produces the best results, in contrast to the information reported by Hartung. Also,

the inclusion of the arrival time correction increases the SDR suggesting a better prediction. However, it was found that in angles where the HRTF changes quickly e.g. due to the diffraction of the head and pinna, the arrival time correction decreases the performance of the interpolation. The authors proposed to increase the resolution of the HRTF measurements in this angle interval to compensate for this drawback.

In a more recent research, Queiroz and Montesiao [117] evaluated the performance of two interpolation methods for rendering static and moving sources. A triangular linear interpolation of finite impulse response (FIR) filters corresponding to the HRTFs was compared with a spectral interpolation of infinite impulse response (IIR) filters (approximations of the FIR filters). The aim of this research was to reduce the computational load to provide a multi-user auralization platform. The weighting of the measured HRTFs for the interpolation was made based on the concept of Vector Base Amplitude Panning (VBAP) [118]. In the triangular method, the HRTF was approximated by the sum of the closest weighted neighbours. In the case of the spectral interpolation, the Z-transform was applied and then two methodologies were implemented, the filter coefficients interpolation and the pole-zero interpolation. The results showed that the triangular interpolation estimates the HRTFs with a very good accuracy in frequencies up to 5000 Hz. In addition, the IIR filters were a suitable approximation of the HRTFs allowing a trade-off between fidelity and computational cost.

The previous works rerpresent the movement of a source based only on the HRTF. This means that the simulations were done assuming an anechoic environment without taking into consideration the influence of an enclosure. This is a huge limitation in interactive auralizations where the acoustic of the room plays an important role in the immersive effect. A more robust application of the direct RIR rendering has been developed by Catt-Acoustics with the Catt-Walker tool. This module generates interactive auralizations based on the interpolation between impulse responses. The main concept is supported in the implementation of the B-format room impulse response and its respective downmix for binaural reproductions. The generation of a real-time auralization may be divided into two stages. First, the RIRs are calculated on a grid distributed around the enclosure. The resolution of the grid depends on the complexity of the sound field, e.g. if it changes rapidly (near field) a higher density of RIRs is required [31]. However, because this is an off-line step, there is no constraint with the number of RIRs. In the second stage, the RIRs are interpolated according to

the user's movements and convolved with an anechoic audio signal. Some important considerations can be drawn about the use of the interpolation. The computational cost of the real-time stage only depends on the length of the FIR filters (not the complexity of the room). Partitions inside the room can lead big errors in the interpolation if they are not taken into account. Also, the Doppler's effect is difficult to reproduce with this technique.

2.2.6 Spatial encoding of acoustic pressure meshed data

Although a monophonic impulse response allows for the main objective acoustic parameters of an enclosure to be estimated, spatial information is required when generating realistic auralizations. The outcomes from the methods based on the numerical solution of the wave equation are usually acoustic pressure data discretized over the domain. From that information, it is possible to extract spatial information, which can be used for auralization purposes using common sound field reproduction techniques. Different methodologies have been proposed to that end: Southern and colleagues [51] presented a method to obtain a spherical harmonic representation of FDTD simulations. The approach uses the Blumlein Difference Technique to create a higher order directivity pattern from two adjacent lower orders. This is performed by approximating the gradient of the pressure as the difference between two neighbouring pressure points on the grid where the solution was computed. The results suggest that this methodology allows for a correct estimation of the spherical harmonic coefficients that describes the sound field in that area, but in the current implementation only 2D cases (3D where the height information was not required) were analyzed.

Alternative sound field representations have been used by other researchers to obtain spatial information from acoustic meshed data. The plane wave expansion (PWE) is a common approach given the versatility when implementing different sound reproduction techniques. Støfrinsdal and Svensson [13] proposed the use of an inverse method to estimate a set of plane waves, which in turn, reconstruct the sound field in a given area determined by the position of a virtual microphone array. The findings indicate that the inverse approach is a suitable method to estimate properly the complex gain of the plane waves. The approach was implemented for 2-D cases but can be straightforwardly applied in 3-D.

One advantage of encoding spatial information from acoustic pressure meshed data is the possibility to control the sound field in terms of translation and rotation. Different approaches have been proposed, the most common of them being the plane wave and spherical harmonic representations. The plane wave expansion presents a suitable alternative for translating sound fields based on the application of delays in the time domain according to the relative position between the listener and the plane waves. This technique has been implemented in different studies [119–121], the main limitation being the size of the region where the reconstruction is accurate, which in turn depends on the frequency and on the number of plane waves used to reconstruct the acoustic field. In contrast, the spherical harmonic representation allows the sound field to be rotated by a simple matrix multiplication in this domain [87]. Due to the interoperability between these two approaches [17], translation and rotation can be applied together leading to a more sophisticated methodology to generate interactive auralizations. This combination has been implemented in the current research and its advantages, assumptions and limitations are discussed in the following chapters.

2.2.7 Conclusions about the state of the art and its relation to the current research

The review of the scientific literature points out that auralization is a relevant and an active area for research. Nowadays, the combination of methods for the numerical solution of the wave equation and geometrical acoustics is the most accurate approach to generate broadband auralizations from predicted room impulse responses. However, auralizations created under this approach usually correspond to a fixed position in space due to the computational cost demanded by the wave based methods.

Alternatively, the rendering of sound fields in real-time is a relevant part of the research that is being carried out in terms of auralization. The potential of reconstructing the acoustic response of an enclosure as the listener interacts with the environment has large number of applications in areas such as telecommunications, video gaming, teaching, consultancy, among others.

Taking into consideration the above, the generation of interactive sound fields whose room impulse responses have been numerically estimated using a combination of the methods for the numerical solution of the wave equation and geometrical acoustics is a relevant contribution for expanding the applications of auralization. In this thesis,

a methodology to create interactive auralizations of enclosures is presented. The approach is based on the representation of sound fields as a finite superposition of plane waves, which allows for interactive features such as translation and rotation. Several sound reproduction techniques can be straightforwardly implemented as well, thus providing a convenient method for rendering acoustic fields in real-time.

Chapter 3

Room Acoustic Simulations

This chapter addresses the simulation of room impulse responses based on a combination of the finite element method and geometrical acoustics. Firstly, the results of acoustic measurements that were conducted to validate the predictions of several enclosures are reported. Then, general modelling features such as the sound source representation, the geometrical detail of the room and the definition of the boundary conditions are considered for both simulation methods. The outcomes of each approach are compared to the measurements by means of the frequency response and time domain acoustic parameters. Finally, a methodology to combine the results from FE and GA simulations is presented. The results from the simulations are used in Chapter 4 to generate a plane wave expansion, which enables interactive operations such as translation and rotation of the acoustic fields.

3.1 Acoustic measurements

Measurements of Room Impulse Responses (RIRs) have been performed in two different rooms based on the ISO 3382-2 "engineering method" [122]. To that end, 3 receiver positions were recorded using 2 source locations for each enclosure. Two sets of measurements were taken for each receiver (two per source) leading to a total of 12 measurements per room. The measurements were carried out using an omnidirectional microphone (Brüel & Kjær Type 4189-L001) and a directional loudspeaker (Mackie 824 MK2). The excitation signal was an exponential sine-sweep from 20 Hz to 20 kHz.

The Impulse Response (IR), for this specific problem, is the transient response of the acoustic pressure captured at a specific observation point in free field due to the reproduction of a Dirac delta impulse emitted at a different position. The relation of the Dirac delta function and the inhomogeneous wave equation in free field is given by [123]

$$\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right) G\left(\mathbf{x}, t|\mathbf{y}, t_y\right) = -4\pi\delta(t - t_y)\delta(\mathbf{x} - \mathbf{y}). \tag{3.1}$$

Whose solution corresponds to a free field Green's function defined as

$$G\left(\mathbf{x}, t|\mathbf{y}, t_y\right) = \frac{\delta(t - t_y - r/c)}{r}, \tag{3.2}$$

where $\mathbf{x}$ is the fied point, $\mathbf{y}$ identifies the location of the acoustic source and $r = |\mathbf{x} - \mathbf{y}|$. In the case or a bounded propagation, as sound inside of a room, customized Green's functions can be formulated to satisfy the boundary conditions for a specific problem. Regarding the room impulse response, the transient measured response contains the information of the boundary conditions that are determined by the reflective surfaces of the enclosure. From the room impulse response, it is possible to estimate several objective acoustic parameters that define the acoustic characteristics of the space. Some of these parameters are the reverberation time (T_{20}), the Early Decay Time (EDT) and the Clarity index (C_{80}) [124].

The reverberation time is defined as the time necessary for the sound energy density in a room to decrease by 60 dB after the excitation signal has stopped. However, because measuring a decay of 60 dB requires a significant level/noise ratio, this value is usually extrapolated from shorter decay curves. The T_{20} refers a reverberation time that has been derived from a the decay curve between 5 dB and 25 dB below to the initial level. The EDT corresponds to the time required to obtain a decay in the decay curve of 10 dB evaluated between 0 dB and -10 dB. It is related to the perceived reverberance of the space, while the reverberation time is mainly associated with the physical properties of the space. Finally, the C_{80} provides an estimation of the ratio between the early and late energy of the room impulse response. The early limit adopted in this research is 80 ms, which is used when the results are intended to be associated with music. The clarity index is defined as

$$C_{80}(\mathbf{x}) = 10 \log \frac{\int_0^{80} p^2(\mathbf{x}, t)dt}{\int_{80}^{\infty} p^2(\mathbf{x}, t)dt} \text{dB}, \tag{3.3}$$

in which p is the instantaneous acoustic pressure of the room impulse response at the field point $\mathbf{x}$. From the measured data, the three previous acoustic quantities were calculated and compared to the simulations. This is an established methodology to verify the consistency of the simulations [125–128]. The selection of the two first parameters was motivated by the fact that humans are very sensitive to the reverberation time. The parameter C_{80} was taken into consideration because it provides an indication of the ratio between the early and late part of the room impulse response. This ratio can be related to the deterministic and the diffuse behaviour of the RIR which play an important role in the auralization process [129].

3.1.1 Meeting room

A typical meeting room (No. 4079 in building 13 at the Highfield Campus of the University of Southampton) was selected as a reference case for comparing the measured and predicted acoustic response of an enclosure. It is an L-shaped enclosure with a volume of approximately 88 m^3. The walls are formed by painted plaster over block; there are two large areas of glazing on adjacent walls, the lower parts of which are covered by rigid acrylic board. The floor is a concrete slab covered with a heavy traffic carpet and the ceiling has transverse rectangular beams made of concrete (see Figures 3.1 and 3.2). A large wooden table is located at the centre of the room. The room was selected because of its moderate size and significant number of edges, which present a challenge to the applicability of GA methods.

FIGURE 3.1: Meeting room 4079

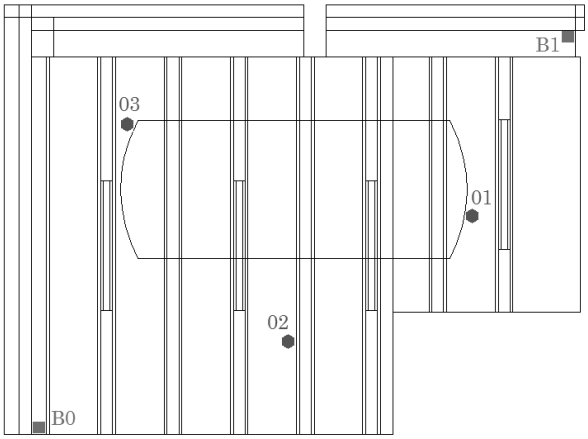


FIGURE 3.2: Plan view of meeting room. Source (B0, B1) and receiver (01, 02 and 03) positions are illustrated

Figure 3.3 shows the average reverberation time T_{20} in octave bands from 63 Hz to 4 kHz. Due to the spatial dependency of the EDT and C_{80}, the information related to these two parameters is reported for each source-receiver path in Figures 3.4 and 3.5

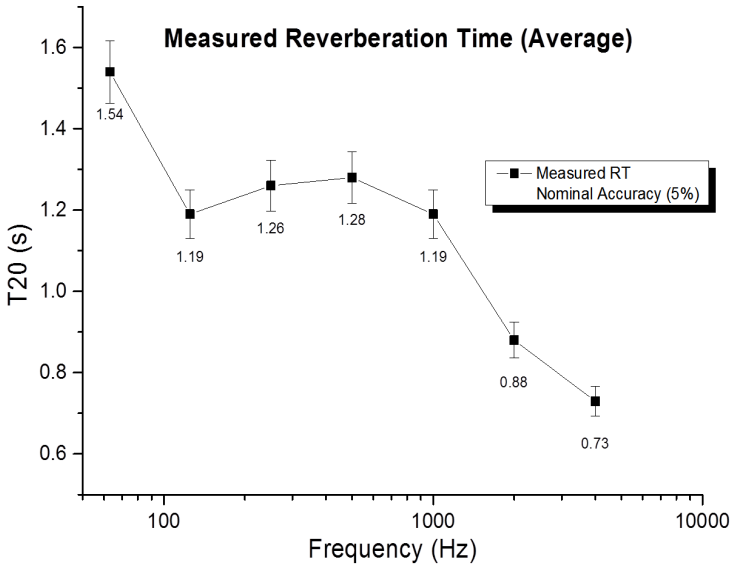


FIGURE 3.3: Average measured reverberation time, meeting room. The nominal accuracy is determined based on the guidelines given by the ISO 3382-2 using the engineering method

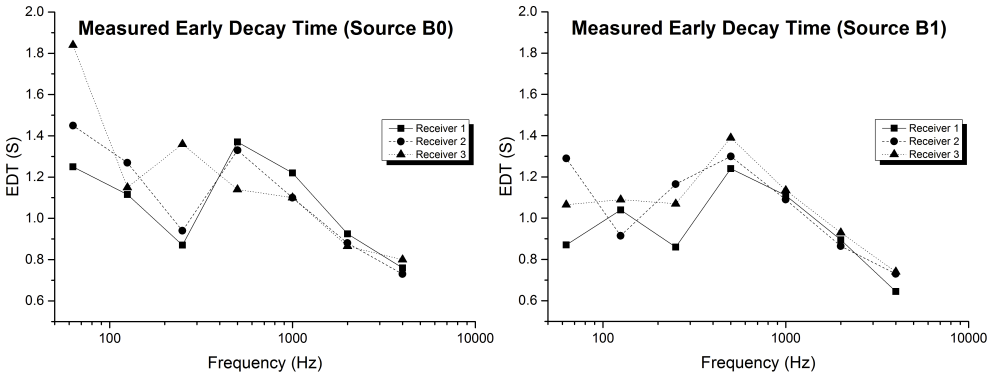


FIGURE 3.4: Measured early decay time, meeting room

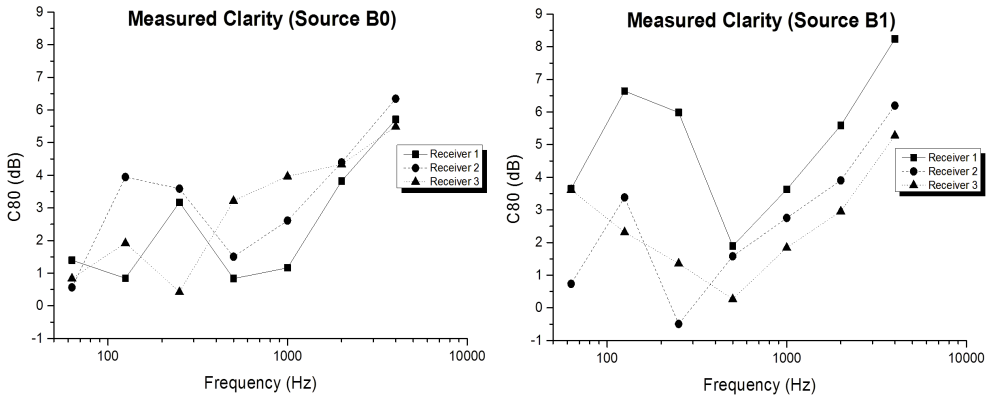


FIGURE 3.5: measured clarity index, meeting room

3.1.2 Ightham Mote

The Ightham Mote is a moated ancient house built in the 14th century that nowadays is maintained by the National Trust organization. The house consists of several spaces, the most relevant of which are the Great Hall, the Chapel and the Crypt. The Great Hall was selected for the experiments due to its interesting characteristics. It is a room with a volume of approximately 434 m^3 located on the first floor of the house that provides access to the other rooms. Its interior consists mainly of block and wood with two glass windows and a chimney. A carpet is located at the centre of the room with a large wooden table at the front. Two knight armours and other elements such as vases and paintings are part of the decoration elements. One relevant characteristic is that wood is highly engraved with fine details and the walls have a very rough finished (see Figures 3.6 and 3.7). These features enhance the diffusion of the enclosure.

FIGURE 3.6: Great hall of the Ightham Mote

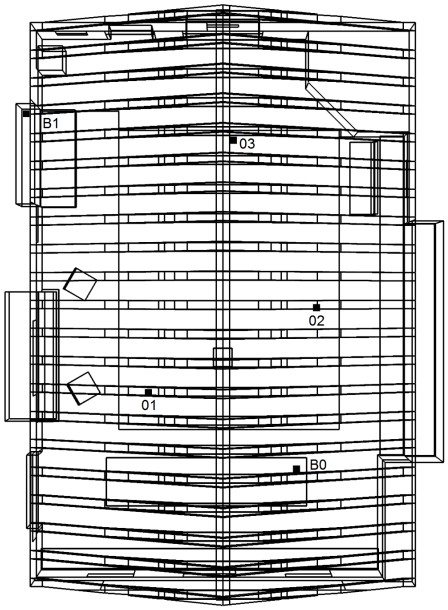


FIGURE 3.7: Plan view of the hall of the Ightham Mote. Source (B0, B1) and receiver (01, 02 and 03) positions are illustrated

Figure 3.8 shows the average reverberation time T_{20} in octave bands from 63 Hz to 4 kHz. The EDT and C_{80} are reported in Figures 3.9 and 3.10.

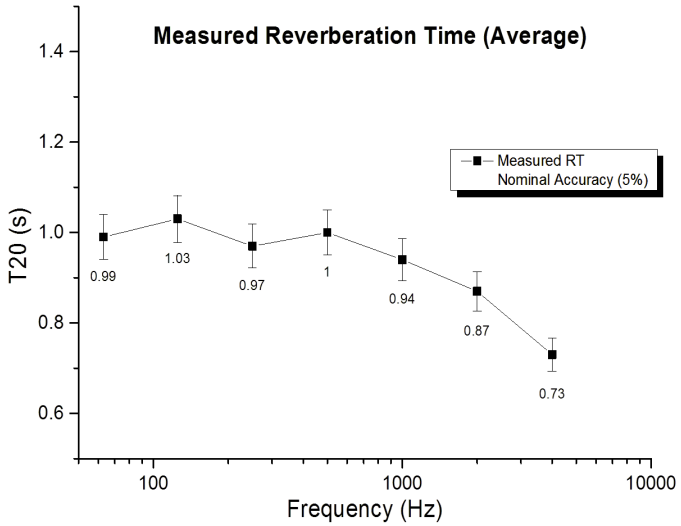


FIGURE 3.8: Average measured reverberation time, Ightham Mote. The nominal accuracy is determined based on the guidelines given by the ISO 3382-2 using the engineering method

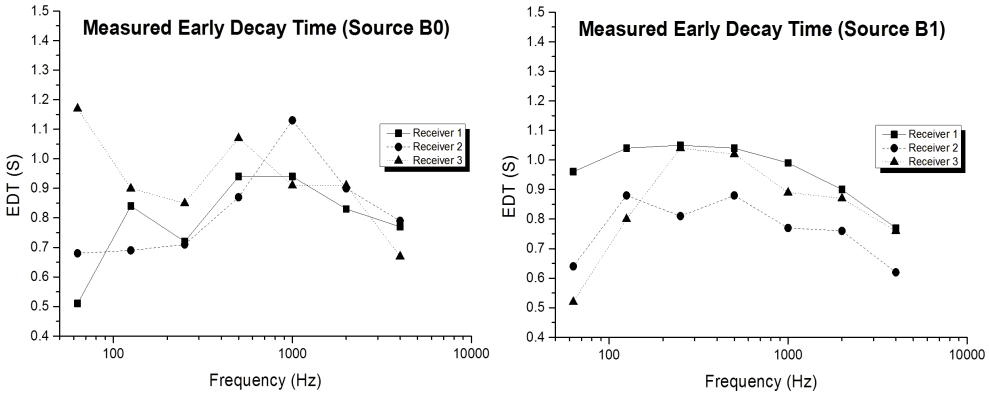


FIGURE 3.9: measured early decay time, Ightham Mote

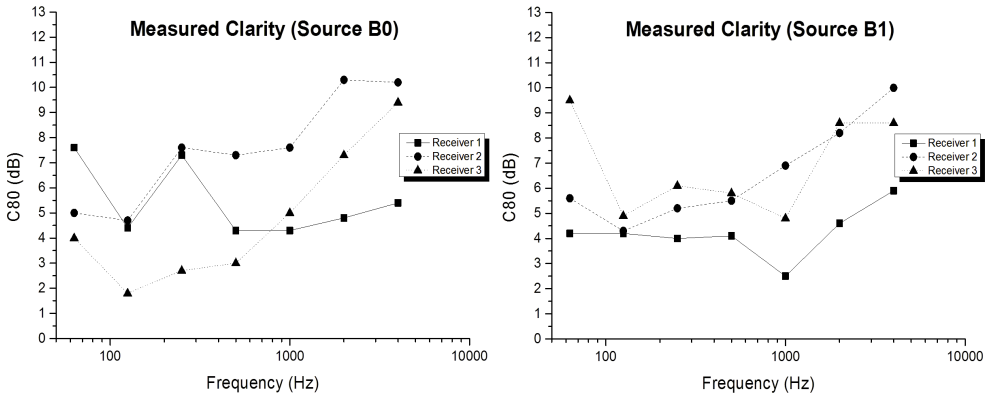


FIGURE 3.10: measured clarity index, Ightham Mote

The measured data indicates that in the case of the Ightham Mote, the reverberation time exhibits small variations with frequency up to 1 kHz. This result is unusual because the sound absorption properties of the materials are usually frequency dependent. However, the measurement procedure fulfilled the requirements established by the standard and the number of receivers was sufficient to characterize the room properly. Furthermore, nine additional positions were measured to analyze this trend in the reverberation time. It was found that this peculiar response was homogeneous in the set of 48 measurements. Table 3.1 presents the standard deviation for the reverberation time discretized in octave bands for all the measurements.

Src/Hz	**63**	**125**	**250**	**500**	**1000**	**2000**	**4000**
B0	0.11	0.12	0.06	0.05	0.04	0.03	0.01
B1	0.18	0.08	0.04	0.05	0.02	0.02	0.01

TABLE 3.1: Standard deviation of the reverberation time (s), Ightham Mote

This result may be caused by the transmission loss of the materials. When a sound wave impinges a wall, part of its energy is absorbed by the boundary, part is reflected into the room and part is transmitted through the element. If the transmission loss of the walls that compose the room is low, a relevant amount of energy will be transmitted to another environment through them, thus decreasing the reflected waves and thereby the reverberation time. The Great Hall has certain conditions that decrease the sound insulation: it is connected to other parts of the house via large wooden doors without acoustic sealing. Also, some partitions between the room and other spaces are made of thin wood. Finally, the Hall has two large single-glazing windows

to different outdoor spaces.

An explanation for the uniform reverberation time across several listener positions may be attributed to the significant number of irregular elements that lead to increase the diffusion. This phenomenon provides a uniform sound field where the reflections arrive in all directions with equal probability. A consequence of a high diffusion is a constant reverberation time over the diffuse field. The Great Hall has walls with rough finishes and elements of wood which are highly hard-carved. In addition, the roof contains a large number of exposed beams and there are many irregular elements as frames, vases, lamps, plates, armours, etc. All these elements help to generate non-specular reflections that increase the diffusivity of the room leading to a spatial uniform reverberation time.

3.1.3 Accuracy of the measurements

The uncertainty in the measured data must be taken into account to determine if the results are reliable and therefore can be used to validate or calibrate the simulations. This parameter depends on different factors such as frequency, size of the room, distribution of the absorption, the method used to excite the room and measurement equipment, among others. A set of measurements conducted by four different teams in the 3rd round robin on room acoustical computer simulation [125] indicated that the reverberation time for a specific receiver can vary between 8%-35% at 125 Hz. The conditions of the measurements corresponded to a room with closed curtains. When the curtains were open, the variation decreased significantly, which suggests that the amount of absorption plays a role in the accuracy of the measured reverberation time. A local dependence of the reverberation time was also found, as expected, due to the small size of the room. A different research carried out by Vorländer [5] showed that in order to obtain a just noticeable difference (5% according to ISO 3382-1) in the measured reverberation time it is necessary to characterize the absorption coefficients with a precision that is not possible using the reverberation chamber method (ISO 354) [56]. The standard deviation of the reverberation time according each source position is reported in Table 3.1 for the Ightham Mote. Table 3.2 illustrates the values for the meeting room.

Src/Hz	**63**	**125**	**250**	**500**	**1000**	**2000**	**4000**
B0	0.09	0.05	0.09	0.05	0.04	0.01	0.03
B1	0.20	0.09	0.09	0.02	0.01	0.02	0.02

TABLE 3.2: Standard deviation of the reverberation time (s), meeting room

Although the standard deviation is not directly related to the accuracy of the measurement procedure, it provides information about the spatial dependency of the reverberation time. The results indicate that the standard deviation is higher at low frequencies for both rooms. Nevertheless, the values are not significantly high, which suggests consistency in the measured data for both rooms. It is important to point out that by implementing an engineering method is expected to have a nominal accuracy of 5% in the estimation of the reverberation time. This parameter is different from the standard deviation, which only indicates how uniform is the reverberation time across measurement positions. The measurement procedure fulfilled the requirements established by the ISO standard and the standard deviation of the reverberation time for both enclosures is low. Therefore, it is assumed that the measurements provide a reasonably accurate description of the acoustic of the spaces and they can be used as a reference to calibrate the simulation models.

3.1.4 Office room

In addition to the enclosures considered before, a typical office room (No. 4091 in building 13 at the Highfield Campus of the University of Southampton) was chosen to evaluate the shielding effect within an enclosure. It is a rectangular room with a volume of approximately 46 m^3. The walls are formed by painted plaster over block and there is a large area of glazing at the back wall. The floor is a concrete slab covered with a heavy traffic carpet and the ceiling has transverse rectangular beams made of concrete. A wooden column was located between the source and receiver path to investigate the shielding effect (see Figure 3.11). One receiver-source receiver path was used for this reference case because the small volume of the enclosure in combination with the wooden column do not allow more source and receiver positions to be located keeping the minimum distance established by the ISO standard [122]. Nevertheless, five measurements were conducted to ensure the consistency of the information for this source-receiver path. Figure 3.12 illustrates the source and receiver locations.

FIGURE 3.11: Office room 4079

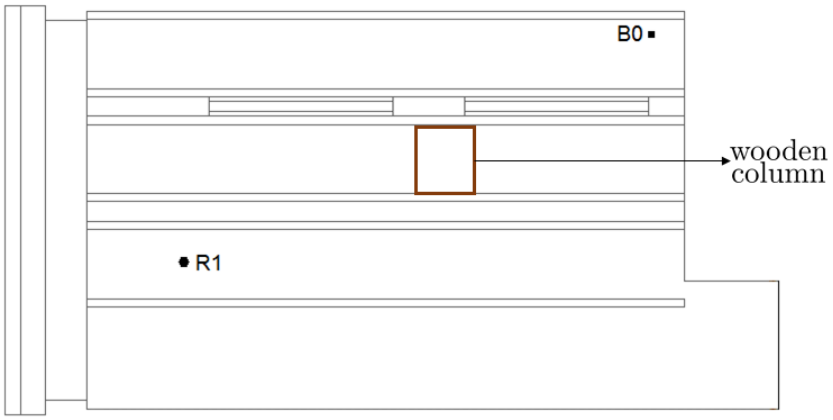


FIGURE 3.12: Plan view of office room. Source B0 and receiver 01 positions are illustrated

The reverberation time T_{20} is shown in Figure 3.13. The EDT and C_{80} are reported in Figures 3.14 and 3.15, respectively.

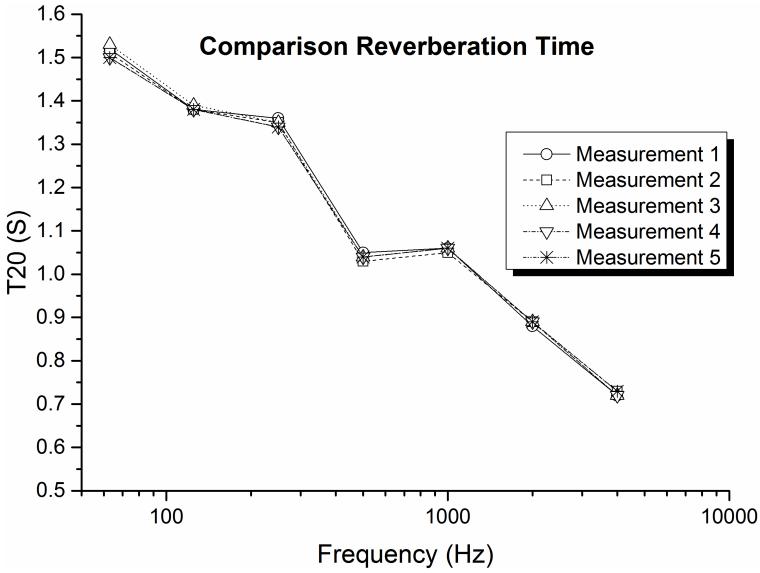


FIGURE 3.13: Measured reverberation time, office room

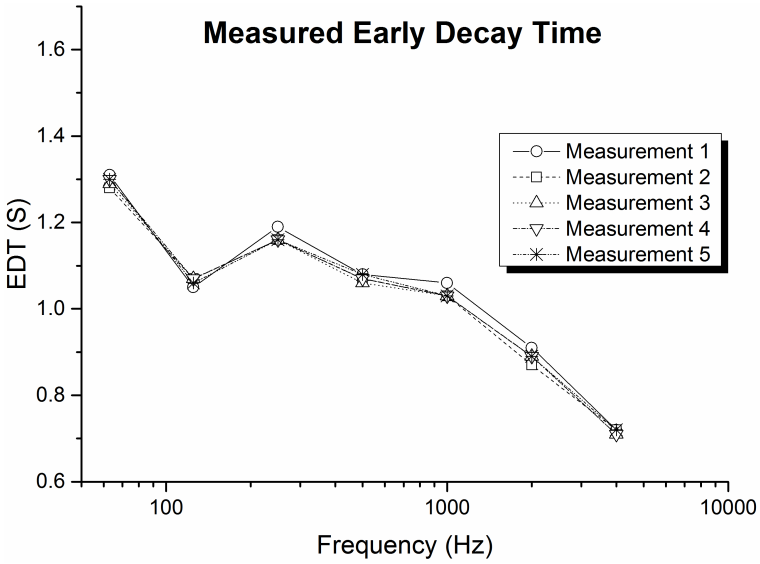


FIGURE 3.14: Measured early decay time, office room

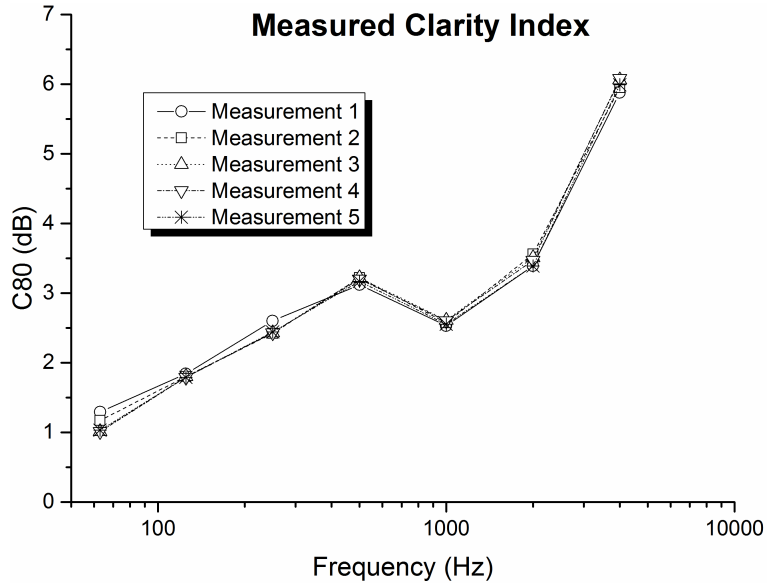


FIGURE 3.15: measured clarity index, office room

3.2 Room impulse response simulations using geometrical acoustics

3.2.1 Meeting room

Simulations were performed using the commercial package Catt-Acoustics V9. Although Catt-Acoustics is based on GA methods, it simulates the effect of diffusion through the implementation of scattering coefficients and includes an algorithm to estimate the edge diffraction using a secondary edge source method. Source-receiver paths from the measurements were duplicated in the GA models to compare the synthesized RIRs to the corresponding measured data. The number of rays selected for the simulation was 60000 (39054 was the recommendation of the software) and the truncation time was fixed at 1600 ms, that is longer than the measured reverberation time.

3.2.1.1 The geometric model of the room

In contrast to CAD models used for architectural purposes, a geometrical acoustic model must contain only details that are relevant for ray propagation. The detail in the geometry affects the time of calculation and the accuracy of the output data.

Siltanen et al. [126] proposed a method to reduce the level of geometry detail for GA simulations. Although the aim of the study was the development of an automatic tool to reduce the complexity of the geometry, the comparison between different levels of detail showed the significant effect of the geometrical detail in the final result.

Vorländer [3] recommends including surfaces with linear dimensions larger than 0.5 m. This value is suggested because smaller surfaces would affect only frequencies above 7 kHz, where the uncertainty of the simulations is a significant constraint on the accuracy of the results. Another reason is that the image source model assumes an infinite plane to produce an image source. If the size of the plane is decreased, this assumption is compromised due to the diffraction produced by the edges of the surface [5]. A more detailed geometry can be used to estimate the diffused reflections, however this effect is partially modelled by means of the scattering coefficients. Therefore, only surfaces and objects that are large compared to the wavelength should be taken into consideration.

In spite of the consideration above on the simplification in the geometry of the room, different studies have pointed out that in some cases a more detailed model produced better results. Foteinou et al. [127] simulated an ancient church using two geometric models. The first corresponded to a detailed model which contained the complex vaulting in the roof. In the second model, the vaults were replaced by planar sloped surfaces with higher scattering coefficients. The results indicated that the detailed model was closer to the measurements in terms of reverberation time, early decay time and clarity index. Following the same direction, Gade and collaborators [130] predicted the acoustic response of the Aspendos Roman theatre using different levels of complexity in the geometric model. The findings suggest that the detailed model provides a better agreement between the measured and predicted data. However, the results and the level of agreement were highly dependent of other parameters such as the transition order that is defined as the point where the ISM is changed by the RT model. Nowadays, the simplification of the room geometry requires more consideration due to the development of new algorithms to predict the effect of diffraction. Edges that were previously omitted (e.g. staircase steps) now can be modelled and therefore included in the geometry. Experiments related to the level of geometrical detail required for GA simulations are presented for the Ightham Mote (section). A model of the meeting room was created using the CAD software package AutoCAD and includes all the relevant edges. The most important surfaces were taken into

consideration as well the furniture inside of the room. Figure 3.16 illustrates the level of detail in the geometry used to simulate the enclosure.

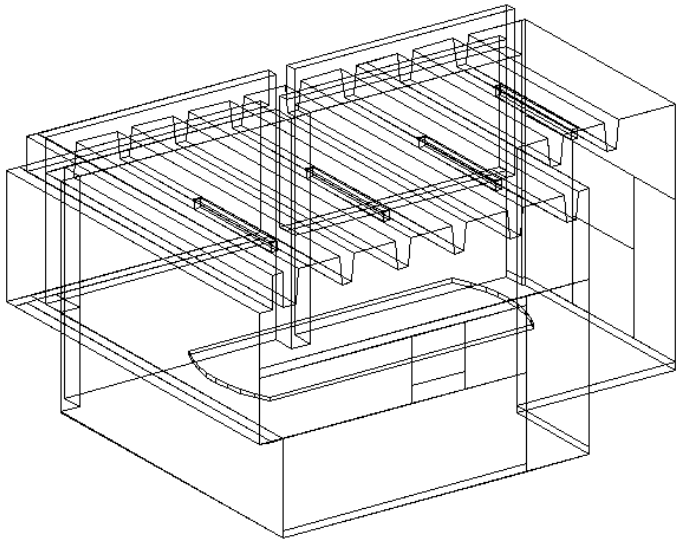


FIGURE 3.16: Geometric model of the meeting room

3.2.1.2 Characterization of the acoustic source

As indicated in the literature review, several studies have shown that the directivity of the acoustic source plays an important role in the prediction of RIRs for use in auralization. Therefore, the directivity of the loudspeaker used in the measurements was first characterized in a large anechoic chamber. An omnidirectional microphone was used to record the sound field generated by the loudspeaker (driven with white noise) at different azimuth and elevation angles. The sound field was measured on a sphere of 2 m radius by using an angular sampling of 15°. The frequency response of the loudspeaker was also estimated to evaluate whether frequency compensation was required. Figure 3.17 illustrates the magnitude of the radiation pattern plotted against the azimuthal angle for octave bands from 125 Hz to 4 kHz. Clearly the directivity is significant at higher frequencies. Information related to the directivity of the loudspeaker was included in the GA model by using the tool provided by the software for that purpose.

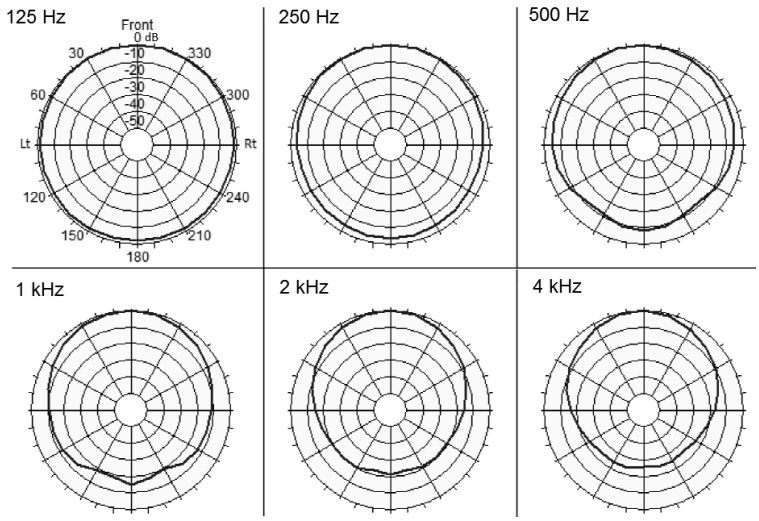


FIGURE 3.17: Measured loudspeaker directivity normalized to maximum value at each frequency

3.2.1.3 Characterization of the boundary conditions

A significant component of the creation of the simulation model is the definition of the boundary conditions. Catt-Acoustics characterizes them using two different coefficients in an octave band resolution. The energy in a reflected cone-ray is determined by the diffuse absorption coefficient, which specifies the amount of energy loss when a ray impinges a surface. The second coefficient is the random scattering coefficient, which determines if a given reflected cone-ray has a specular or non-specular direction. The diffuse absorption and scattering coefficients were selected from the Catt-Acoustics library and from scientific literature [3] according to the visual inspection of the surfaces of the enclosure. Then, the methodology proposed by Aretz [131] was implemented to calibrate the model. To that end, the average absorption coefficient of the room was calculated from the measured reverberation time using Eyring's reverberation equation. The values of the diffuse absorption coefficients were then modified to match this value.

Because the room is composed of several materials, different ratios exist between the diffuse absorption coefficients on different surfaces. These can be varied independently with different combinations yielding the same average absorption. The best agreement in terms of T_{20}, EDT and C_{80} was achieved (for the three rooms considered) when all the coefficients were modified proportionally without changing

their relative ratios. Finally, because GA simulations can predict a reverberation time slightly different from the measured one in the enclosure, an iterative algorithm has been implemented to calibrate the model. The algorithm iteration is expressed as

$$\bar{\alpha}_{n+1} = \bar{\alpha}_n \frac{T_{20s,n}}{T_{20m}}, \tag{3.4}$$

where $\bar{\alpha}$ is the average diffuse field absorption coefficient, n is the number of the iteration, T_{20s} and T_{20m} are the predicted and measured reverberation time, respectively. Equation (3.4) indicates that the target average diffuse absorption coefficient is modified based on the ratio between the predicted and the measured reverberation time.

The methodology proposed by Aretz [131] is based on the Sabine/Eyring equations to estimate the average absorption and, subsequently, the implementation of an iterative algorithm to calibrate the results. This means that the applicability of this technique depends on how well the rooms are fitted to the Sabine/Eyring criteria. Furthermore, the output value is the mean absorption coefficient which gives an average of the absorption of the room. However, an enclosure composed of different materials can be modified in different proportions leading to the same mean absorption value. In other words, the variation of the absorption coefficients can be performed in different proportions yielding the same average absorption.

The selection of the absorptive properties of the materials may produce an insignificant impact in the reverberation time if the same average absorption is achieved, but parameters such as EDT and C_{80} can be relevantly affected because of their spatial dependency. Due to this characteristic, three different distributions in the absorption coefficients that lead to the same average absorption were evaluated. Firstly, the materials were selected from the scientific literature. Then, some of them were modified to achieve the mean absorption coefficient. In the first distribution (model 1), the absorption of the floor (carpet) was changed because the other materials have very low absorption at high frequencies. In model 2, floor and windows (glass) were altered. Finally, in the last configuration (model 3), all materials were modified in such a way that their relative ratios were not modified.

3.2.1.4 Results

Objective room acoustic parameters

In the following the predicted reverberation time (T_{20}), early decay time and clarity index are presented for the three models. The red curves represent the estimation using energy addition and the blue curves the pressure addition, respectively.

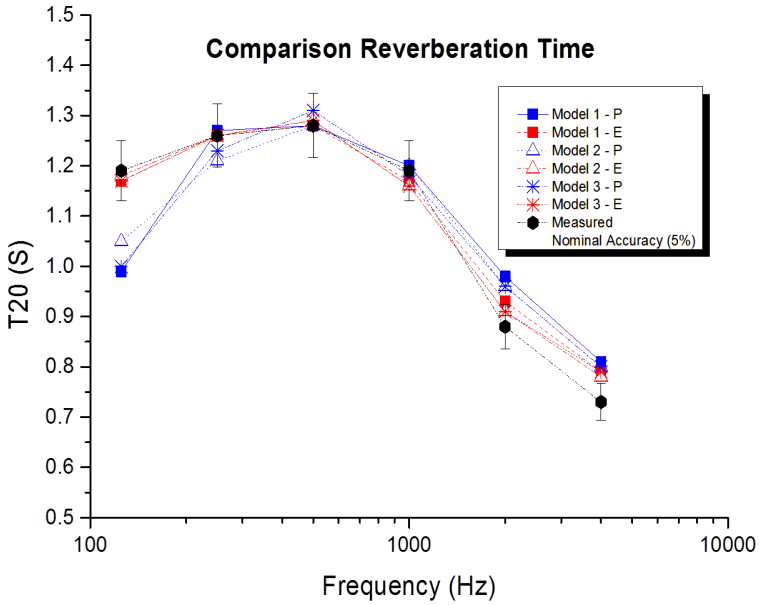


FIGURE 3.18: Predicted T_{20}, meeting room, GA

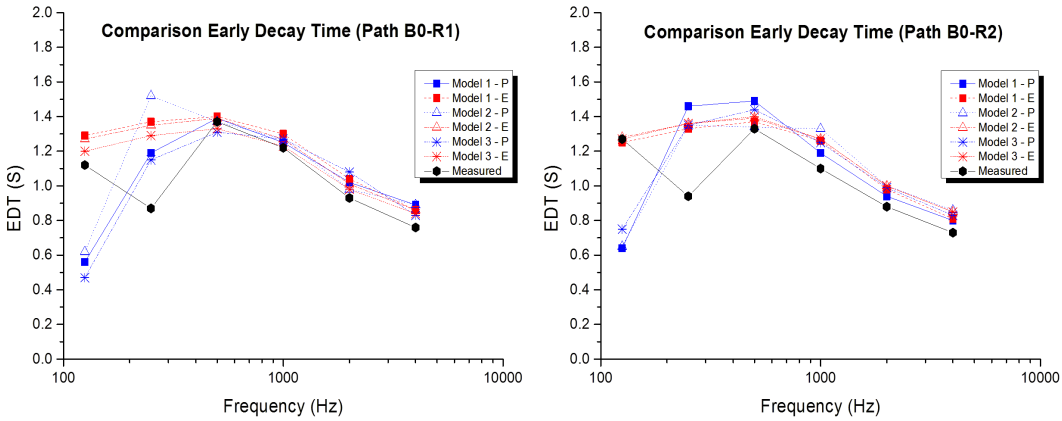


FIGURE 3.19: Predicted *EDT*, meeting room, GA, paths B0R1 & B0R2

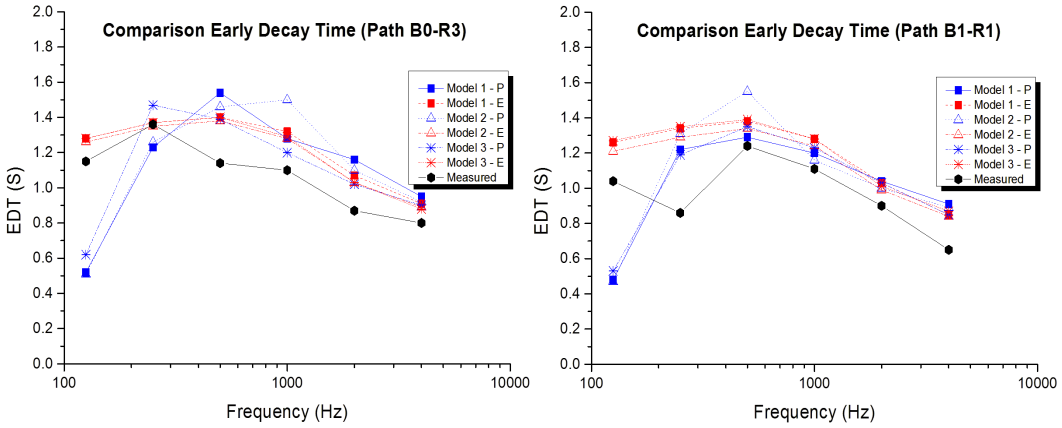


FIGURE 3.20: Predicted *EDT*, meeting room, GA, paths B0R3 & B1R1

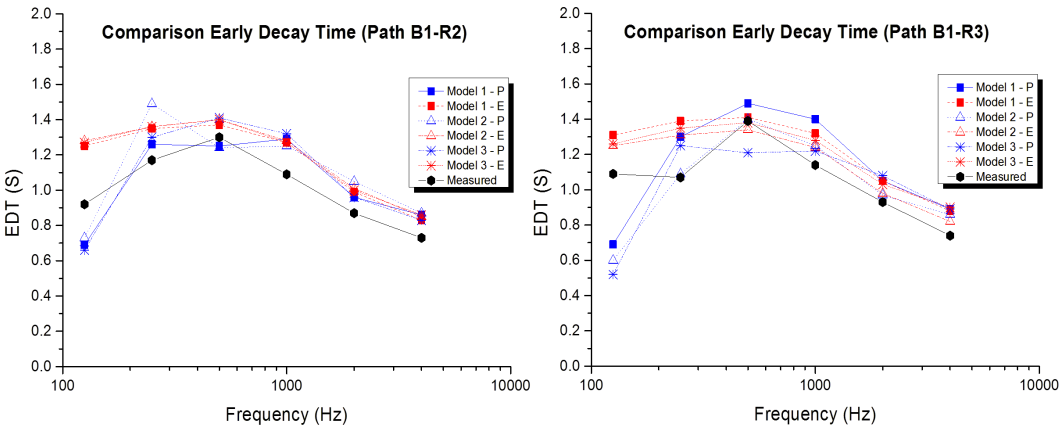


FIGURE 3.21: Predicted *EDT*, meeting room, GA, paths B1R2 & B1R3

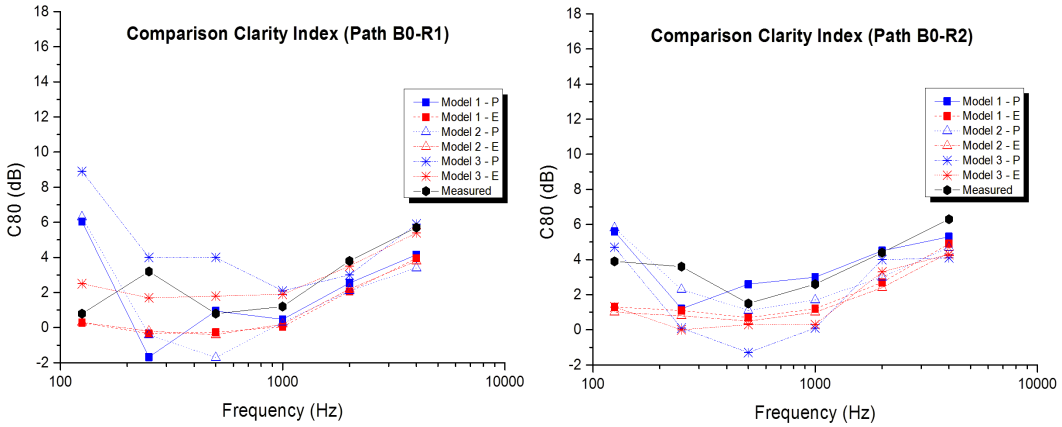


FIGURE 3.22: Predicted C_{80}, meeting room, GA, paths B0R1 & B0R2

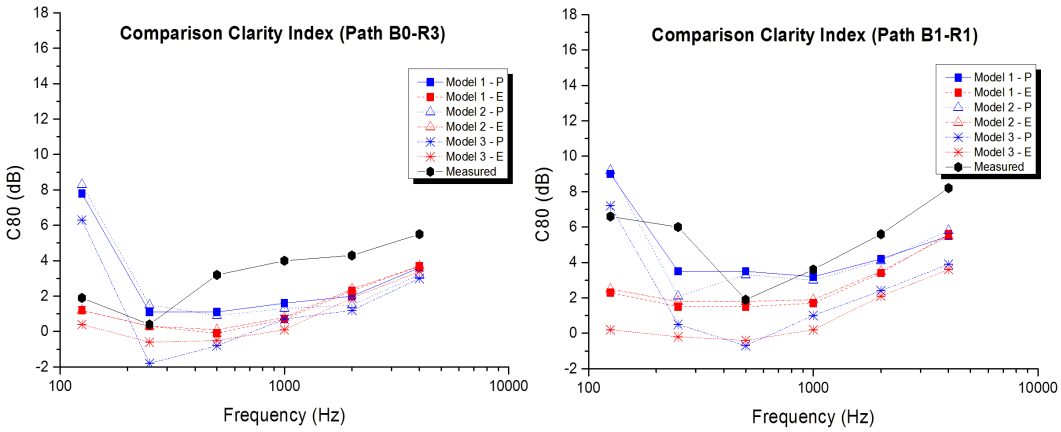


FIGURE 3.23: Predicted C_{80}, meeting room, GA, paths B0R3 & B1R1

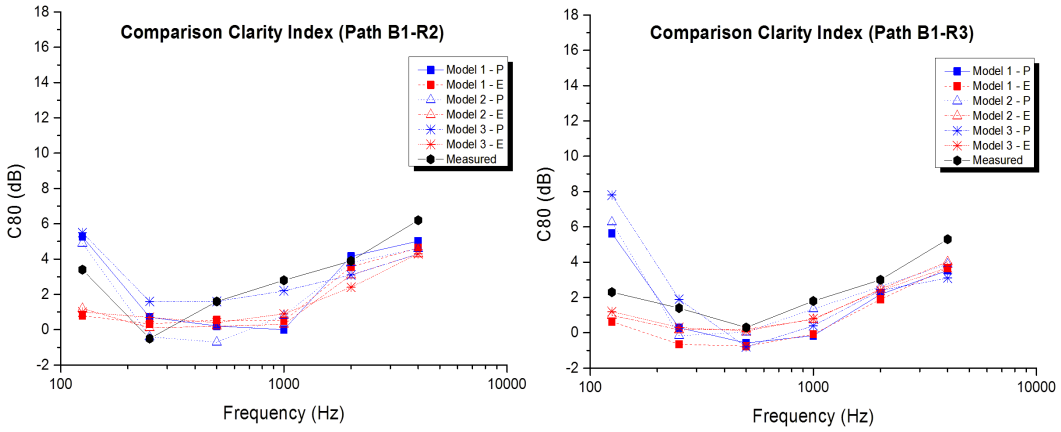


FIGURE 3.24: Predicted C_{80}, meeting room, GA, paths B1R2 & B1R3

The difference between the energy (red) and pressure (blue) curves at low frequencies can be used to determine the uncertainty of the calculation [132]. The reason is because through the use of absorption coefficients the phase of reflections is not considered, which is relevant at low frequencies. This leads to a mismatch between the values predicted using the energy echogram (energy addition) and the impulse response (pressure addition), which suggests how much inclusion of phase would matter. In Catt-Acoustics, the energy reflections are added directly to an energy echogram while with pressure, the phase is synthesized for each reflection using a minimum phase or a linear-phase octave FIR filter depending on the order of the reflection. The difference between these two approaches indicates a threshold where GA theory can be applied. Based on the previous results, the data suggests that GA

can be used from 500 Hz. Other approaches can be considered to determine the GA threshold. The most common is the Schroeder frequency [4]

$$f_s \approx \sqrt{\frac{c^3}{4 \ln 10}} \sqrt{\frac{T_{60}}{V}}, \tag{3.5}$$

where c is the speed of the sound, T_{60} is the reverberation time and V is the volume of the room. The Schroeder frequency is defined as the threshold above which, on average, three modes overlap generating a smoothed frequency response and the sound field tends to be diffuse. Above this frequency the assumption of statistical phase may be justified. Another method based on the spatial dependence is proposed by Kuttruff [133] who stated that the acoustic wavelength should be small compared to the mean free path of the room.

Catt-Acoustics recommends a more conservative value based on an empirical relation of $4f_s$ to determine the lower limit. The Schroeder frequency for the meeting room is $f_s \approx 237$ Hz, which leads to a frequency band limit of 1 kHz. On the other hand, the mean free path of the room is 1.95 m, therefore, establishing a relation of one order of magnitude of this value (0.195 m) the frequency limit can be defined to be 2 kHz approximately. Despite the recommendations given by the Schroeder frequency and the mean free path, the comparison between measured and predicted data suggests that the band of 500 Hz is a suitable threshold above which GA leads to reasonable results (for this specific room).

By applying the implementation of the iterative algorithm it is possible to achieve a good agreement between the predicted and measured reverberation time. Nevertheless, parameters such as EDT and C_{80} show important differences mainly in the octave bands of 125 and 250 Hz. This can be attributed to the size of the room which leads to a strong modal response at low frequencies that cannot be predicted correctly by GA methods. Path B0-R1 shows the best match between simulations and measurements, in contrast, the largest difference was observed in path B0-R3.

Frequency responses

The frequency response in narrow band and in 1/3 octave band resolution of the best (B0R1) and worst (B0R3) source-receiver path are illustrated in Figures 3.25

and 3.26. 0 dB corresponds to the acoustic pressure produced by a monopole source whose source strength generates 1 Pascal at 1 m distance in the octave band of 1 kHz.

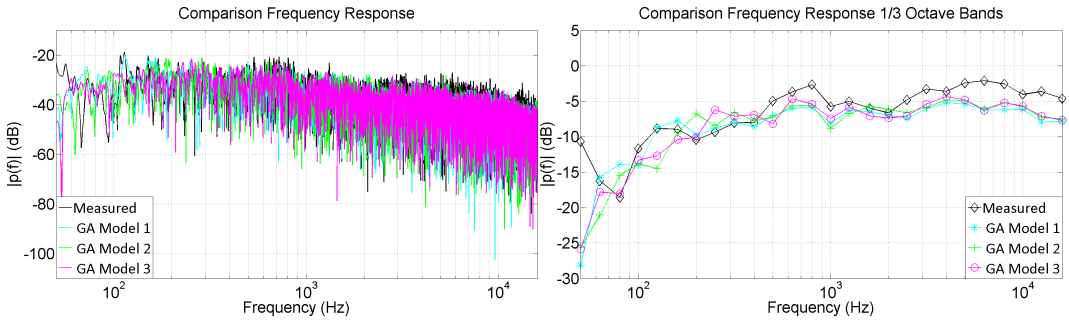


FIGURE 3.25: Predicted frequency response, meeting room, GA, path B0R1

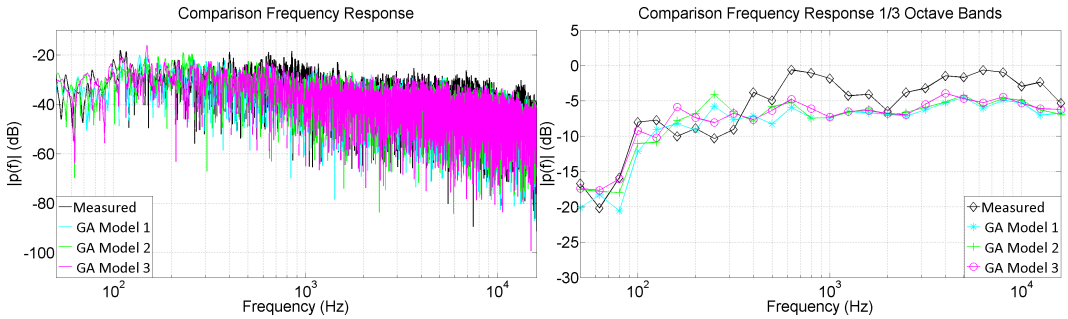


FIGURE 3.26: Predicted frequency response, meeting room, GA, path B0R3

The frequency responses displayed in Figures 3.25 and 3.26 are in accordance with the previous findings, which ratifies that path B0R1 presents a better agreement between the predicted and measured data. It is clear from the previous figures that the modal response of the enclosure cannot be correctly estimated, although a good agreement in terms of the 1/3 octave band resolution was obtained for the path B0R3 at low frequencies. The implementation of methods for the numerical solution of the wave equation to predict the low frequency content of the RIR will be addressed in further sections. Regarding the selection of the absorption coefficients, the mean square errors (MSE) of the frequency response in 1/3 octave band resolution, T_{20}, EDT and C_{80} were calculated to determine which configuration achieves results that are closer to the measurements. The general expression of the MSE is defined as

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(\widetilde{Y}_i - Y_i\right)^2, \tag{3.6}$$

in which n is the number of predictions, $\widetilde{Y}_i$ and Y_i are the predicted and measured data, respectively. The results suggest that the model 3 (all the materials are modified but keeping the initial ratio between them) is the closest to the measured data.

Parameter	**Model 1**	**Model 2**	**Model 3**
Frequency response path B0-R1 (dB)	2.6	2.8	2.4
Frequency response path B0-R3 (dB)	3.2	3.0	2.6
T_{20} (s)	0.01	0.01	0.01
EDT (s)	0.06	0.07	0.06
C_{80} (dB)	5.6	6.3	5.9

TABLE 3.3: Mean squared error for different absorption distributions, meeting room

3.2.2 Ightham Mote

As for the meeting room, a model of the Ightham Mote was created in Catt-Acoustics V9. The number of rays selected for the calculations is 60000 (42083 was the recommendation of the software) and the truncation time was fixed at 1500 ms. Because the procedure to predict the room impulse responses in the Ightham Mote is the same as the meeting room, only the modelling parameters whose characterization changed with respect to the previous enclosure are reported below.

3.2.2.1 The geometric model of the room

Evaluation of the level of geometrical detail

The geometry of the enclosure was created using a *Total Station* device, which is a precision instrument used in archeological surveying and building construction to obtain data related to distance and angles (see Figure 3.27). From the raw data, three different models were implemented and compared with the measured data. The first model corresponded to a basic geometry where the ceiling and some edges of the room were replaced by flat surfaces. In the second model, the edges of the walls were included but not the complex vaulting of the roof. Finally, the third model included details of the ceiling and furniture. Figure 3.28 shows the implemented geometries. Results related to these geometries are reported at the end of this section.

FIGURE 3.27: Adquisition of the geometry data, Ightham Mote

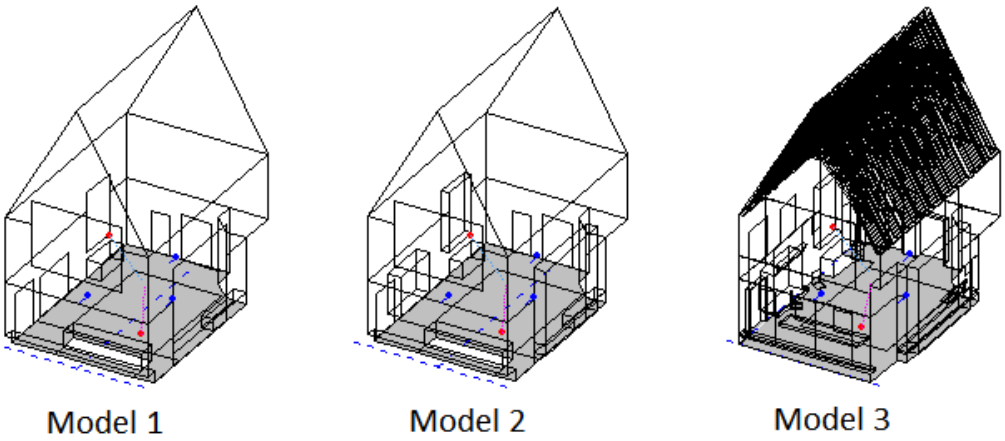


FIGURE 3.28: Comparison of geometrical models of the Ightham Mote

3.2.2.2 Characterization of the boundary conditions

The absorption and scattering coefficients were extracted from the Catt-Acoustics library data and scientific literature [3] following the same procedure as for the meeting room. The selection of these data was made according to the visual inspection of the surfaces of the room. However, the values of the wood absorption coefficient were selected by considering the possible sound transmission to other rooms. Based on that, the absorption coefficient of some wood elements corresponds to a *Hollow*

Wooden Podium (higher absorption at low frequencies). Subsequently, the methodology suggested by Aretz [131] was implemented to calibrate the simulations using the Sabine's equation. Sabine's criterion was selected because is more appropriate for enclosures of moderate size and diffuse behaviour.

3.2.2.3 Results

Objective room acoustic parameters

The predicted reverberation time (T_{20}), early decay time and clarity index are presented below in Figures 3.29 to 3.35 for the three geometric models.

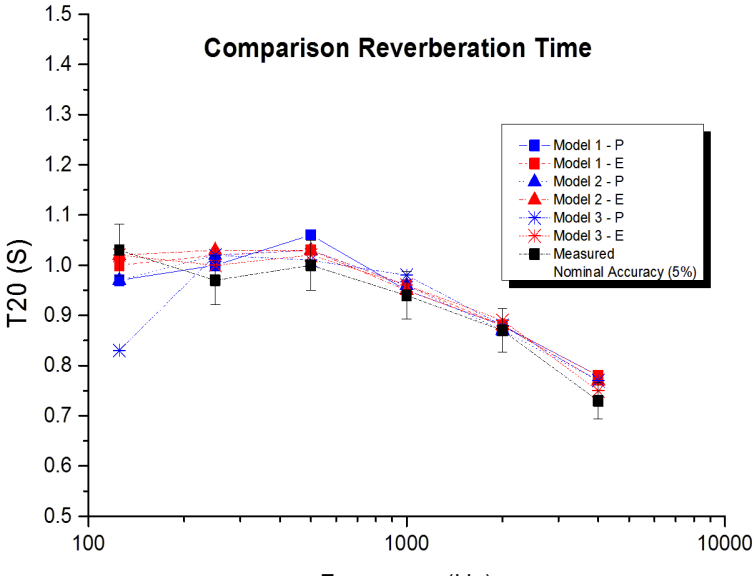


FIGURE 3.29: Predicted T_{20}, Ightham Mote, GA

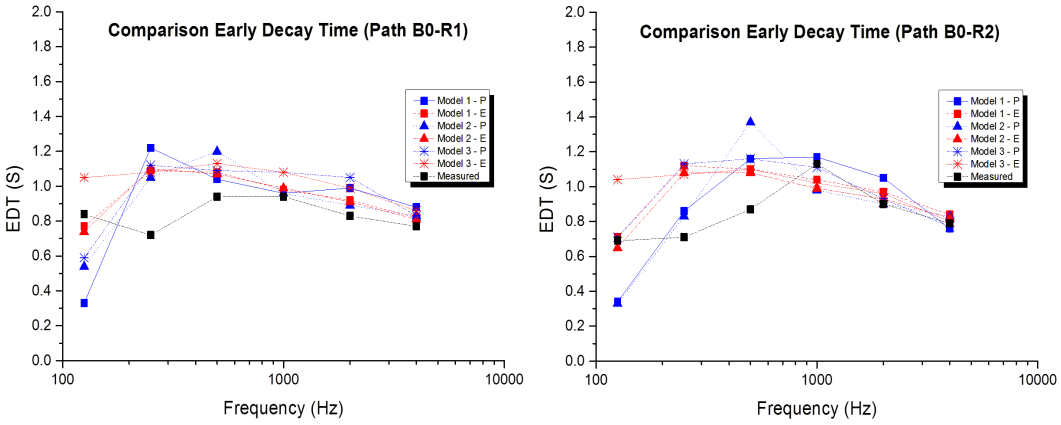


FIGURE 3.30: Predicted *EDT*, Ightham Mote, GA, paths B0R1 & B0R2

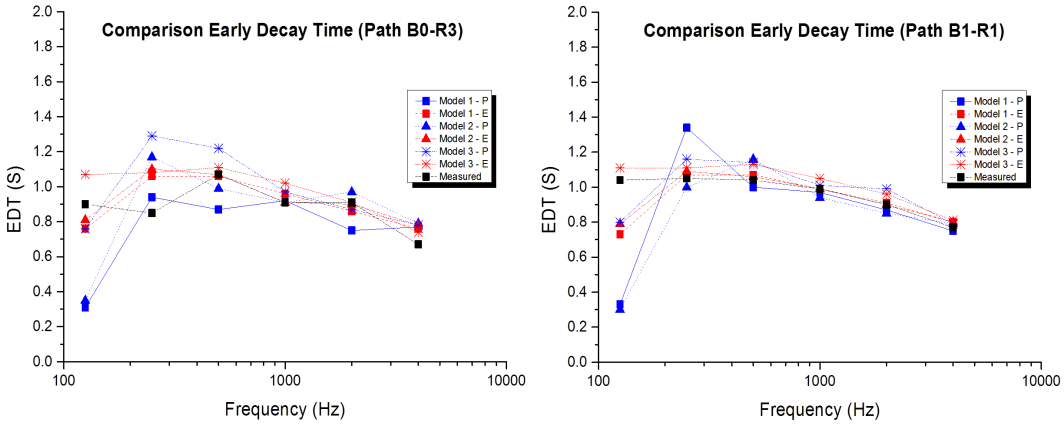


FIGURE 3.31: Predicted *EDT*, Ightham Mote, GA, paths B0R3 & B1R1

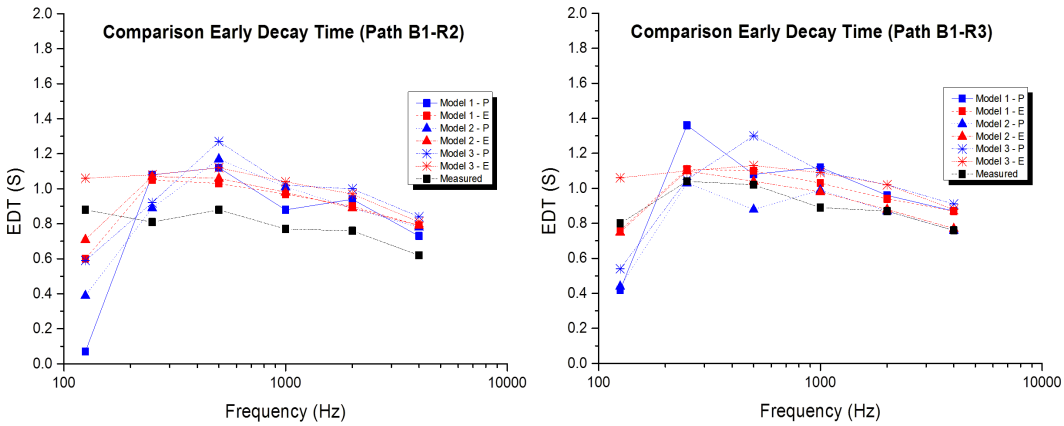


FIGURE 3.32: Predicted *EDT*, Ightham Mote, GA, paths B1R2 & B1R3

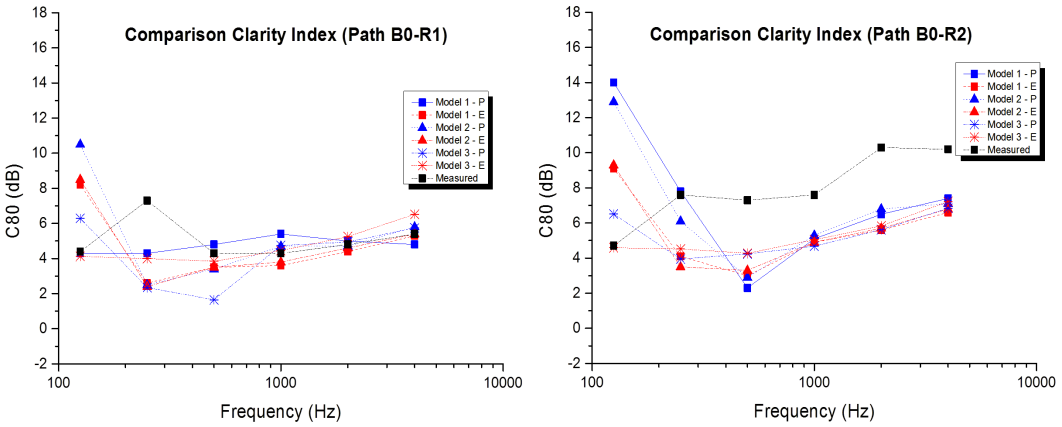


FIGURE 3.33: Predicted C_{80}, Ightham Mote, GA, paths B0R1 & B0R2

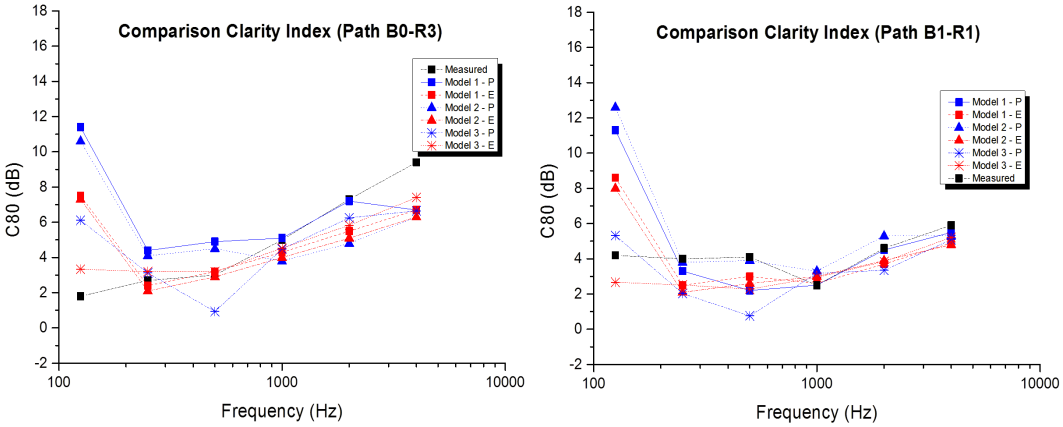


FIGURE 3.34: Predicted C_{80}, Ightham Mote, GA, paths B0R3 & B1R1

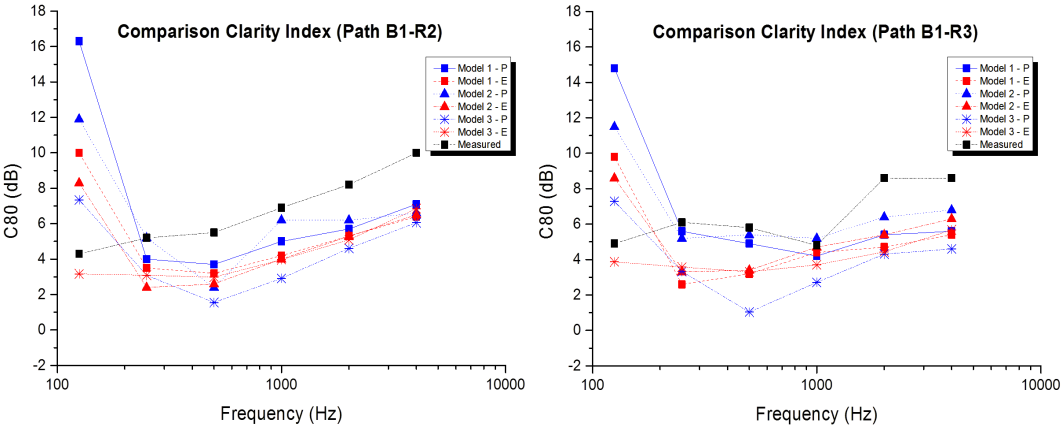


FIGURE 3.35: Predicted C_{80}, Ightham Mote, GA, paths B1R2 & B1R3

The Schroeder frequency of the Ightham Mote is low ($f_s \approx 100$ Hz) because the reverberation time at low frequencies is relative short and the volume is quite large. This yields a discrepancy between the volume of the room and the expected reverberation time in normal conditions (assuming a reverberation time which decays according to frequency) that suggests that the use the Schroeder's approach may be incorrect in this case. Based on Catt-Acoustics recommendations, the empirical relation of $4f_s$ leads to a frequency of 400 Hz. Finally, the mean free path of the room is 4.62 m, therefore, establishing a relation of one order of magnitude of this value (0.42 m) the frequency limit can be defined to be 800 Hz approximately. The comparison between the energy and pressure data indicates that GA can be applied from the 1 kHz octave band.

The best matching was achieved at paths B1R1 and B0R1 and the most significant differences were found at paths B1R2 and B0R2. Receiver two (R2) was the closest to the wall, it was located at the rigth-centre of the room near to the main window (see Figure 3.7). In this area of the enclosure, there are some components like edges and decorative elements that were simplified in the GA models. These simplifications could be the reason for the mismatch of the simulations at this position. Moreover, as expected, the values in the octave bands of 125 Hz and 250 Hz differ more significantly than at higher frequencies. Relative to the level of detail in the geometry of the room, the results of the objective room acoustic parameters do not indicate any relevant improvement due to the inclusion of the edges of the roof.

Frequency responses

The frequency response in narrow band and in 1/3 octave band resolution of the best (B1R1) and worst (B1R2) source-path are presented as follows.

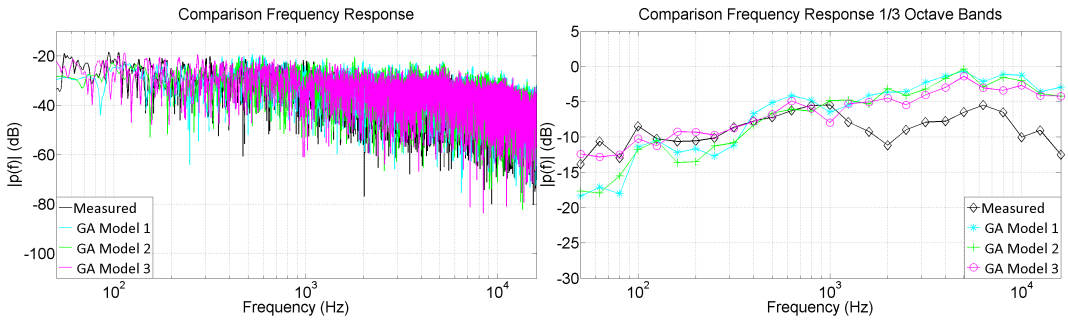


FIGURE 3.36: Predicted frequency response, Ightham Mote, GA, path B1R1

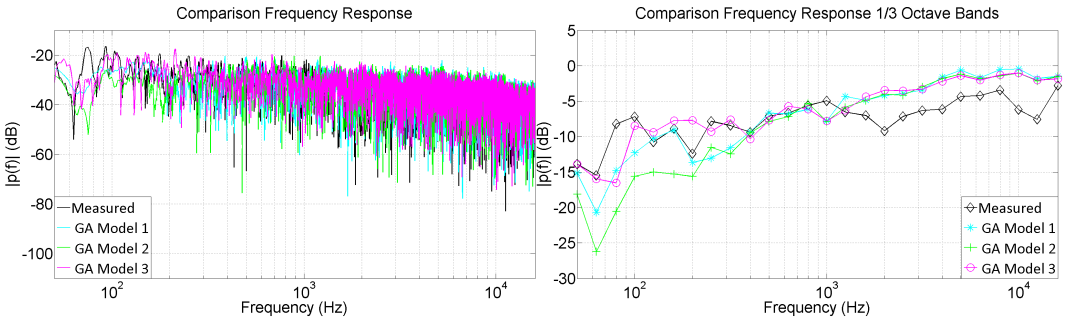


FIGURE 3.37: Predicted frequency response, Ightham Mote, GA, path B1R2

Figures 3.36 and 3.37 indicate that the frequency response of the synthesized and measured room impulse responses are different. These results are consistent with experiments made by Aretz [11, 131] in complex enclosures where the weaknesses of GA are highly enhanced. Nevertheless, it is important to emphasize that although the predicted frequency responses differ from the measured ones, parameters such as reverberation time are correctly predicted. Regarding the level of the detail in the geometry, the results suggest that the inclusion of edges (model 3) provides a slightly better result in terms of the mean squared error.

Parameter	**Model 1**	**Model 2**	**Model 3**
Frequency response path B1-R1 (dB)	4.0	3.7	2.8
Frequency response path B1-R2 (dB)	2.9	3.8	2.4

TABLE 3.4: Mean squared error for different geometric models, Ightham Mote

3.2.2.4 Listening test

A subjective test was conducted to identify if a predicted RIR with different frequency response but with similar acoustic parameters such as T_{20}, EDT and C_{80} can be used for auralization purposes and be subjectively acceptable. A sample of 26 people with normal hearing was used to evaluate the similarity between auralizations created from measured and synthesized RIRs of the Ightham Mote. Half of the sample (13 people) selected for the listening test had knowledge in audio or acoustics, whilst the rest corresponded to students from different Faculties of the University. The assessed information was the sense of space (reverberation time) and the timbre of the sound (frequency response). The experiment was performed in the Audio-Lab of the ISVR using a binaural reproduction technique (see Figure 3.38). The audio material was reproduced by headphones (Sennheiser HD 600) with non-individual

equalization to compensate for its response. A female and male voice recorded in an anechoic environment was used for the test.

The experiment was fully randomized between listeners and questions. It was based on rating different sets of two auralizations, one being the convolution with the measured RIR and the other with its respective prediction. An introduction about the meaning of sense of space and timbre of the sound was given to each participant before starting the test. The following questions were addressed:

- How would you compare with respect to sense of space or reverberation?
- How would you compare with respect to the timbre or tone?

FIGURE 3.38: Audio-Lab, set up of the listening test

Firstly, a comparison between the auralizations generated by the measured and the predicted RIR was carried out. Figures 3.39 and 3.40 show the results for the female and male voice. The left and right part of the figures correspond to the answers of the sense of the space and timbre, respectively.

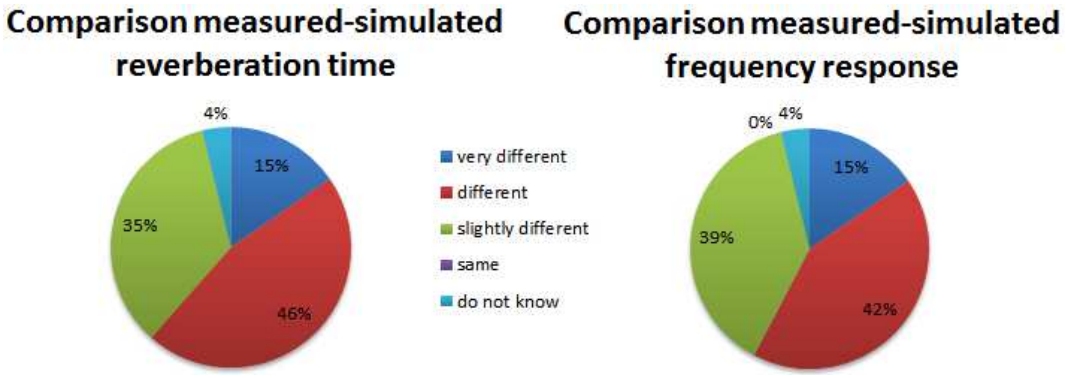


FIGURE 3.39: Comparison of auralized material (female voice) in terms of reverberation time and frequency response

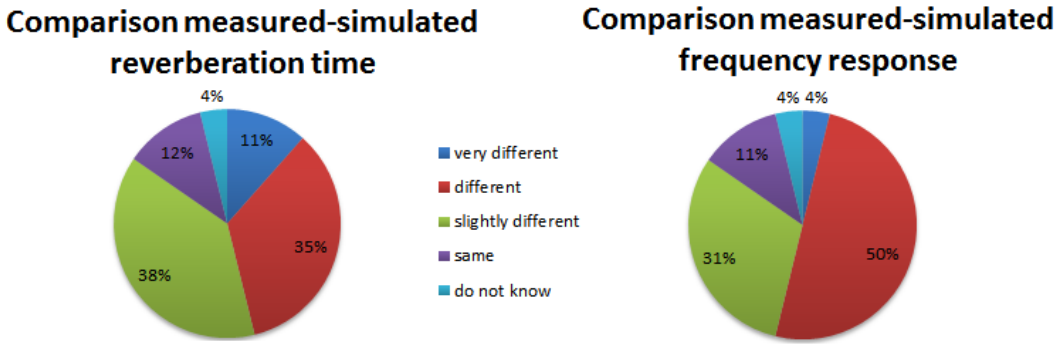


FIGURE 3.40: Comparison of auralized material (male voice) in terms of reverberation time and frequency response

Then, the auralized material was filtered using a 4th order Butterworth high-pass filter with a cut-off frequency of 355 Hz. The signals were filtered to verify if the perceptual differences were related to the low frequency content. Figures 3.41 and 3.42 show the results for the female and male audio respectively.

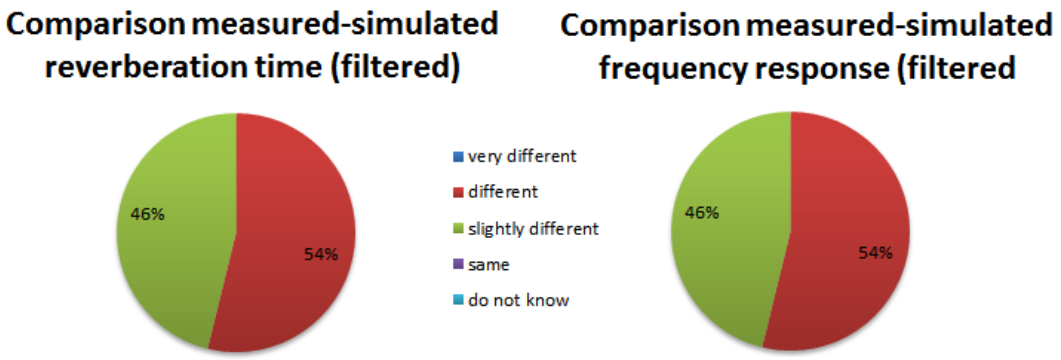


FIGURE 3.41: Comparison of auralized material (filtered female voice) in terms of reverberation time and frequency response

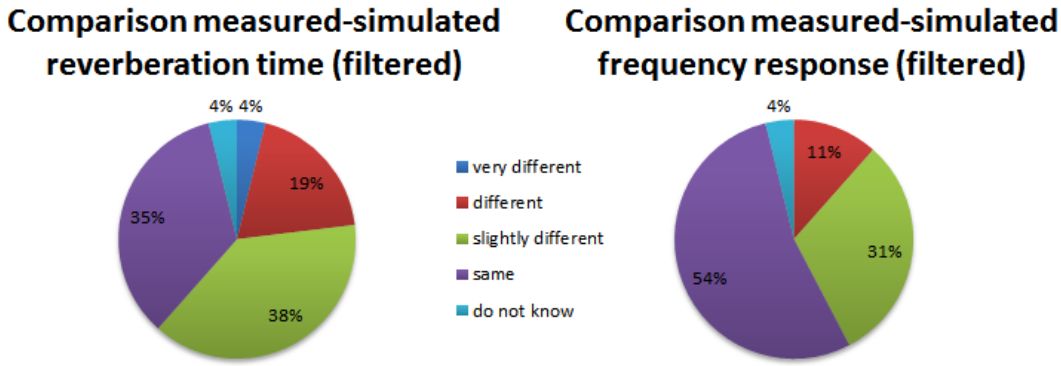


FIGURE 3.42: Comparison of Auralized material (filtered male voice) in terms of reverberation time and frequency response

The results indicated that the auralization made from the synthesized RIR is perceptually different from the auralization created with the measured RIR. However, when the signals were filtered, a better agreement between the auralizations was found. In addition, the male voice presented the best matching after the filtering process. These findings suggest that improving the low frequency calculation using the finite element method could lead to a more realistic sound field reconstructions.

3.2.3 Office room

Simulations were performed based on the guidelines given for the previous rooms. The number of rays corresponded to 50000 and a truncation time was fixed to 1600 ms. Figure 3.43 shows the model used for the simulations.

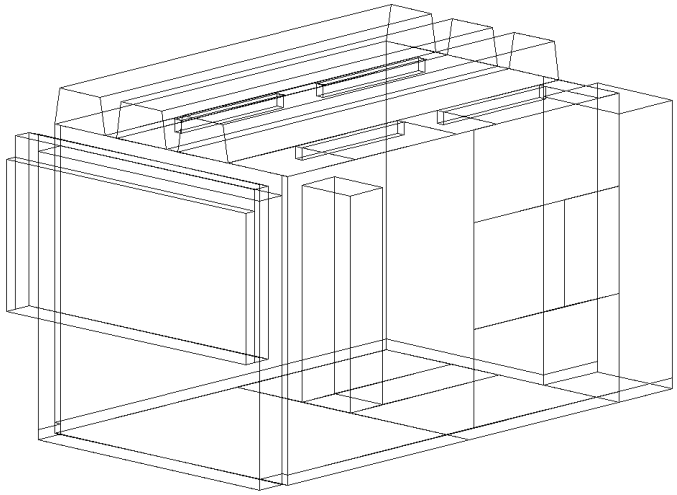


FIGURE 3.43: Geometric model of the office room

3.2.3.1 Results

Objective room acoustic parameters

The predicted reverberation time, early decay time and clarity index are presented in Figures 3.44 to 3.46 as follows. The measured value corresponds to the average of the 5 measurements.

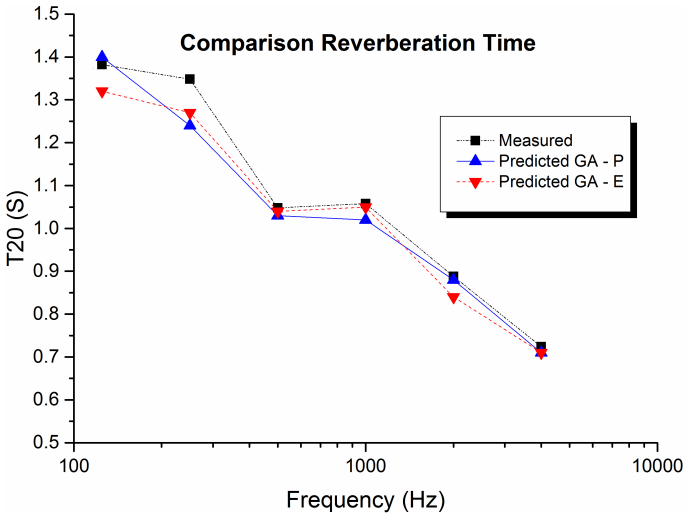


FIGURE 3.44: Predicted T_{20}, office room, GA

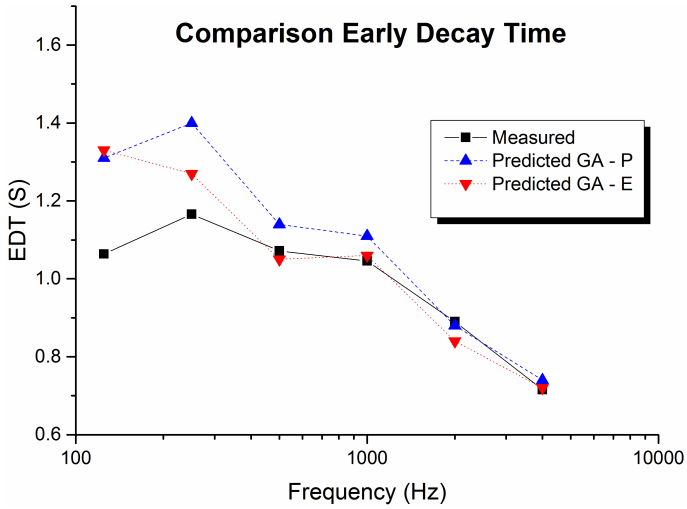


FIGURE 3.45: Predicted EDT, office room, GA

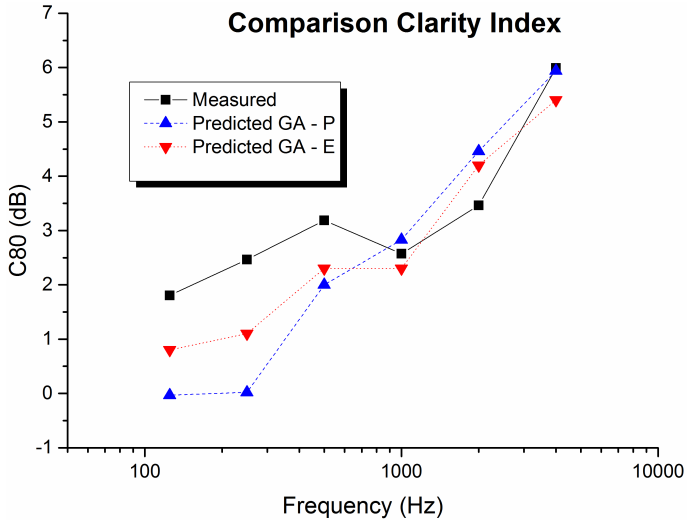


FIGURE 3.46: Predicted C_{80}, office room, GA

The comparison between the measured and predicted data indicates that GA can be applied for octave bands from 1 kHz. The Schroeder frequency of the room is 337 Hz, which leads to the same octave band based on Catt-Acoustics recommendations ($4f_s$).

Frequency responses

TThe frequency response in narrow band and in 1/3 octave band resolution are illustrated in Figure 3.47.

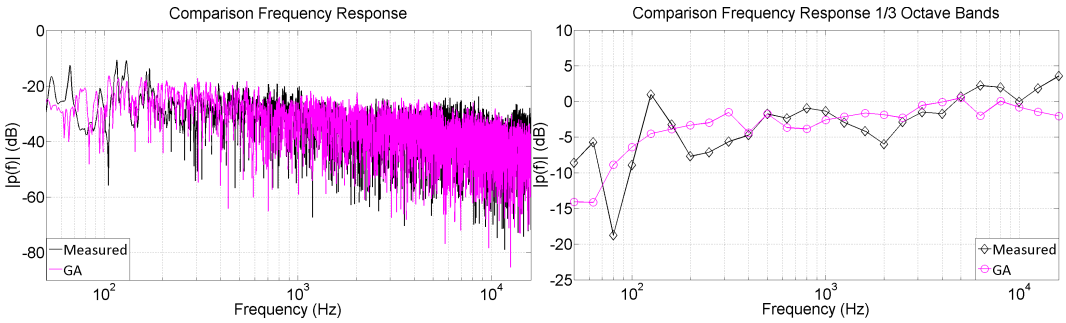


FIGURE 3.47: Predicted frequency response, office room, GA

The results are consistent with the previous findings where the fine structure of the measured frequency response cannot be replicated, but with a higher agreement in terms of the 1/3 octave band resolution at high frequencies.

The analysis for the three enclosures reveals that the implementation of geometrical acoustic methods to synthesize room impulse responses allows for an accurate prediction of the average reverberation time of the rooms. An agreement in terms of early decay time and clarity index is possible at some receiver points but not at all of them. Regarding the calculation of the frequency response, the results indicate that the fine structure cannot be recreated. However, a good matching in 1/3 octave band resolution can be achieved at middle frequencies.

3.3 RIR simulation using the finite element method

Room acoustic simulations based on geometrical acoustics revealed the inapplicability of this approach to predict the low frequency content of the room impulse response. In this section, the implementation of the finite element method using the commercial package Comsol V4.4 is discussed. In the following, the basic theory of FEM is considered. Then, the simulations of the three previous enclosures are carried out taking into account modelling parameters such as the geometry model of the room, the definition of the acoustic source and the characterization of the boundary conditions. The predicted frequency responses are presented for each enclosure.

3.3.1 Introduction

FEM is a numerical method designed to solve partial differential equations in complex geometries where the analytical solution is not achievable. In this case, the governing partial differential equation of the domain is expressed in its integral form (weak formulation), which leads to a system of algebraic equations which can be solved through computers. The weak formulation of the Helmholtz equation taking into consideration the boundary conditions can be expressed as [9]

$$\int_{\Omega}(-\nabla f\cdot\nabla p+k^2fp)d\Omega-jk\int_{S_u}f\rho_0cu_ndS-jk\int_{S_z}A(\omega)fpdS=0, \tag{3.7}$$

where f is an arbitrary weighting function, p is the acoustic pressure, k is the wave number, ρ_0 is the density of the medium, c is the speed of sound, u_n is the normal velocity and $A(\omega)$ is the acoustic admittance. The second term of expression (3.7) represents a vibrating surface with normal velocity u_n and the last term corresponds to a surface(s) with a local admittance. The use of the local admittance or impedance as a boundary condition is only appropriate in the case of *locally reacting* materials, which means that the specific acoustic impedance does not depend on the direction of the incidence wave [131]. Rigid surfaces or materials with high flow resistivity can be cataloged as locally reacting [4].

Subsequently, the domain Ω is discretized by elements in which the acoustic pressure is calculated at the nodes. Therefore, an approximate solution based on equation (3.7) is determined by choosing from all possible functions p and all possible weighting functions f a restricted set of them. This leads to a finite set of equations that depends on the number of nodes of the mesh. In each element, shape functions $N(\mathbf{x})$ that take the value of unity at a specific node and zero for all others are used to estimate the acoustic pressure at different points within the element through an interpolation process. Following the Galerkin's approach, the shape functions are selected as the weighting functions yielding a system of linear equations of the form

$$\left[\mathbf{K}+j\omega\mathbf{C}-\omega^2\mathbf{M}\right]\{\mathbf{p}\}=\{\mathbf{f}\}, \tag{3.8}$$

in which $\mathbf{M}$, $\mathbf{K}$ and $\mathbf{C}$ are the acoustic mass, stiffness and damping matrices. The vector $\mathbf{f}$ is the forcing term due to the structural or acoustic excitation.

3.3.2 Meeting room

3.3.2.1 The geometric model of the room

A similar resolution to that implemented in GA was used to define the geometry for FE simulations. In this case, the model must contain all details whose dimensions are comparable to the shortest wavelength resolved. In the current instance a common model of the geometry was used for both simulation methods. However, for FE simulations the domain must be discretized by using a mesh whose resolution is frequency dependent. A rule of thumb of 6 nodes per wavelength is suggested in the scientific literature [40, 41, 43] and this was used for the current instance. Different meshes were implemented with the mesh resolution of each defined by the octave band to be simulated. This approach reduces the computational cost of the calculations by using coarser meshes at lower frequencies. Figure 3.48 shows different meshes of the room corresponding to different octave bands. The number of nodes in each mesh is also shown.

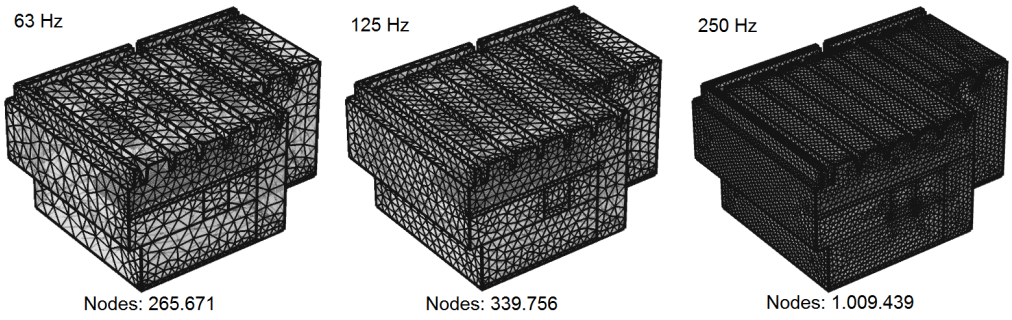


FIGURE 3.48: Number of nodes in the mesh according to the octave band to be simulated in FEM.

3.3.2.2 Characterization of the acoustic source

The directional characteristics of the loudspeaker used in the acoustic measurements should be included in the simulation to obtain accurate results. This involves a complex procedure to create a FEM model for the speaker with the correct directivity and frequency response. However, for the 125 Hz and 250 Hz octave bands (see Figure 3.17), it is reasonable to assume that the loudspeaker is omnidirectional, in which case the acoustic source can be characterized as a simple monopole at these frequencies. This simplification was implemented in the current FE model which

only contributes to the low frequency content of the RIR.

A calibration of the output level emitted by the source was performed to ensure compatibility between the prediction methods. In Catt-Acoustics the level of the source is characterized by the sound pressure level produced by the source at 1 m. White noise (94 dB in the 1 kHz octave band) at 1 m away from the source in a free field was taken as the reference. In Comsol a monopole source can be defined by its acoustic power. The relation between the sound pressure level and sound power level for spherical propagation is given in equation (2.35).

3.3.2.3 Characterization of the boundary conditions

In the case of the FE simulations, the boundaries of the domain are assumed to be locally reacting and the specific acoustic impedance is used to characterize the wave reflection and absorption at each boundary [4]. A locally reacting surface implies that its specific acoustic impedance does not depend on the direction of the incident wave and the particle velocity normal to the wall at any point of the surface depends only on the sound pressure at that point and not on the sound pressure of the surrounding region. This condition is satisfied in the case of rigid surfaces or porous absorbers with high flow resistivity. Non-locally reacting surfaces can be included in a FE calculation but it requires a more complex model that takes into account propagation within the absorbing layer of materials forming the boundary. Locally reacting boundaries are assumed in all of the results presented in this research.

Information on the values of the specific acoustic impedance available in the existing scientific literature is insufficient for many room acoustic simulations. These values are often determined by measurements based on laboratory or in-situ methods. The impedance tube is a well-known technique to estimate the specific acoustic impedance and the absorption coefficient of materials [134]. A limitation of this technique lies in the fact that it requires a sample of the material to be measured in the laboratory which in turn, can lead to significant differences compared to the behaviour of the material in-situ [131]. Different in-situ methods based on pressure-pressure (p-p) [135] and pressure-particle velocity (p-u) [136] probes have been proposed, but the frequency range where the estimation is reliable (above 300 Hz) constrains their use for low frequency FE simulations.

A comparison between two methods to determine the specific acoustic impedance of three types of absorbers has been carried out by Aretz [131]. The materials were measured using an impedance tube and a p-u probe. Two samples of the absorbers were measured in the impedance tube, and the p-u probe was used to characterize the same type of absorbers located at different places of the enclosure. The results revealed large differences between the two methods at frequencies below 400 Hz. Moreover, different values of specific acoustic impedance were obtained with the same method under different conditions. These findings confirm the presence of significant uncertainty in many measurements of specific acoustic impedance at low frequencies.

In the current FE model, diffuse field absorption coefficients that have been calculated in the same way as for GA are used to estimate the specific acoustic impedances. Initially, these are assumed to be purely resistive. This means that the phase in the reflected wave is not accurately represented, only the change in amplitude being modelled correctly. This assumption is valid for hard and reflective surfaces and allows for an easy and practical estimation of impedance from diffuse field absorption coefficients derived from measurements of the reverberation time. Equation (2.48) establishes the relation between the real part of the specific acoustic impedance and the absorption coefficient depending on the angle of incidence. The relation which has been assumed between the specific acoustic impedance and the diffuse field absorption coefficient [137] is given below.

$$\begin{aligned}\alpha_d = 8\Gamma\{1 - \Gamma \ln\left[\frac{r_n}{\Gamma} + 2r_n + 1\right] + \left(\frac{x_n}{r_n}\right)\Gamma\times \\ \left(\left(\frac{r_n}{x_n}\right)^2 - 1\right)\tan^{-1}\left(\frac{x_n}{(r_n+1)}\right)\},\end{aligned} \tag{3.9}$$

in which r_n and x_n are the real and imaginary part of the specific acoustic impedance Z_n non-dimensionalized by $\rho_0 c$, and $\Gamma = r_n/(r_n^2 + x_n^2)$. By setting the imaginary part of the acoustic impedance to zero it is possible to calculate a resistive part which can be used in FE simulations. For a specific value of α_d, r_n is given by the solution of

$$\alpha_d = \frac{8}{r_n}\left\{1 - \frac{1}{r_n}\ln\left(1 + 2r_n + r_n^2\right) + \frac{1}{1+r_n}\right\}. \tag{3.10}$$

In the FE model, the initial values of r_n for each surface were chosen to satisfy equation (3.10).

The effects of assuming purely resistive acoustic impedance

By using equation (3.10), purely resistive specific acoustic impedance values are calculated from the diffuse field absorption coefficients. Here, the consequences of that assumption are investigated with reference to two simple models. First, the effect of an impedance surface on the indirect transmission of sound from a spherical source at A to a receiver point at C is represented by an image source at $A^{'}$ as shown in Figure 3.49. The strength of the image source is defined by a reflection factor R which depends upon the surface impedance. The use of an image source attenuated by a reflection factor to describe a wave reflection has been evaluated by Suh and Nelson [138]. The results indicate that this approach is a good approximation when the source and receiver are located at least 1 wavelength from the reflecting surface.

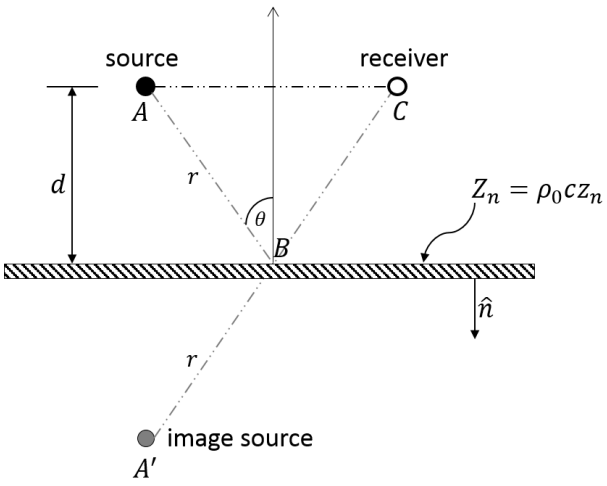


FIGURE 3.49: Diagram of image source model used to represent a wave reflection phenomenon

The relation between the non-dimensional specific acoustic impedance z_n at the surface and the reflection factor R for spherical propagation is determined by the locally reactive boundary condition at B given by:

$$z_n \frac{\partial p}{\partial n} + jkp = 0, \tag{3.11}$$

in which z_n is the acoustic impedance normalized by the characteristic impedance of the medium $\rho_0 c$, j is the imaginary unit, p is the acoustic pressure and k is the wavenumber. The total acoustic pressure at the point B on the boundary (see Figure 3.49) due to the monopole source at A and the image source at $A^{'}$ (attenuated by a reflection factor R) is given

$$p_B = \frac{qe^{-jkr}}{r} + \frac{Rqe^{-jkr}}{r} = (1+R)\frac{qe^{-jkr}}{r}, \tag{3.12}$$

where $q = (j\rho_0\omega q_0)/4\pi$, q_0 is the volume velocity of the acoustic monopole and $r(= d/\cos\theta)$ is the distance from the sources to the point B. The total normal derivative of the acoustic pressure at the point B is

$$\frac{\partial p_B}{\partial n} = -(1-R)\frac{-jkqe^{-jkr}}{r}\left[1 - \frac{j}{kr}\right]\cos(\theta). \tag{3.13}$$

Inserting (3.12) and (3.13) into equation (3.11) gives, after some rearrangement of terms

$$R = \frac{z_n\left[1 - \frac{j}{kr}\right]\cos(\theta) - 1}{z_n\left[1 - \frac{j}{kr}\right]\cos(\theta) + 1} \tag{3.14}$$

Equation (3.14) shows that for a given complex non-dimensional specific acoustic impedance z_n the image source strength, Rq, depends on the incidence angle θ and on the distance of the source from the boundary. It is therefore different for each receiver point. Figure 3.50 illustrates the absolute value of the acoustic pressure plotted against frequency at the receiver point C for three different types of specific acoustic impedance (purely resistive and complex with positive and negative imaginary part) that give the same diffuse field absorption coefficient according to equations (3.9) and (3.10). The coordinates of the points A, B and C correspond to (4.5, 9.5, 1.5), (0, 7.25, 1.5) and (4.5, 5, 1.5), respectively. The reflective boundary is located at $x = 0$. The results suggest that the frequency responses are similar in all three cases, but that in the case of the positive imaginary impedance (i.e. positive reactance) the spectrum has moved to the right compared to the purely resistive impedance. For negative values of the reactance, the frequency response has moved to the left. Small changes in the magnitude of the peaks and troughs are also present but the general envelope is preserved.

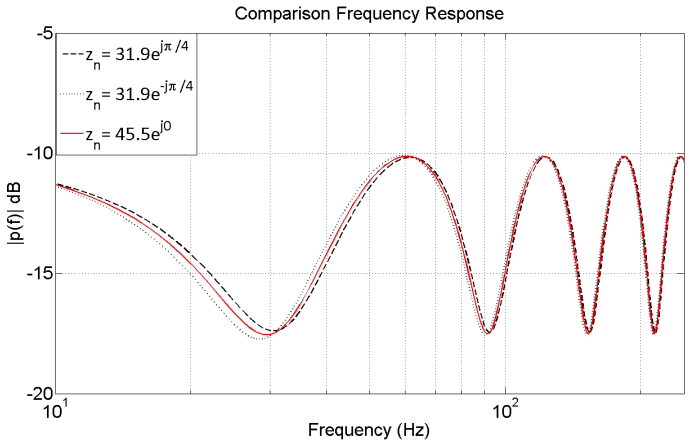


FIGURE 3.50: Comparison of frequency responses of a first order image source model with complex and purely resistive specific acoustic impedance

The second model corresponds to a slightly more complicated problem where a rectangular enclosure of dimensions (5m x 10m x 3m) has been considered. The origin (0, 0, 0) is located at the left-bottom corner of the room. A FE model is used to determine the frequency response at a receiver point (1, 2.5, 1.5) due to a monopole source located at (4.5, 9.5, 1.5). Computed solutions are obtained by using frequency independent purely resistive and complex specific acoustic impedance values uniformly applied to all boundaries. Figure 3.51 shows the results for specific acoustic impedances of varying phase, which give the same value for the diffuse field absorption coefficient.

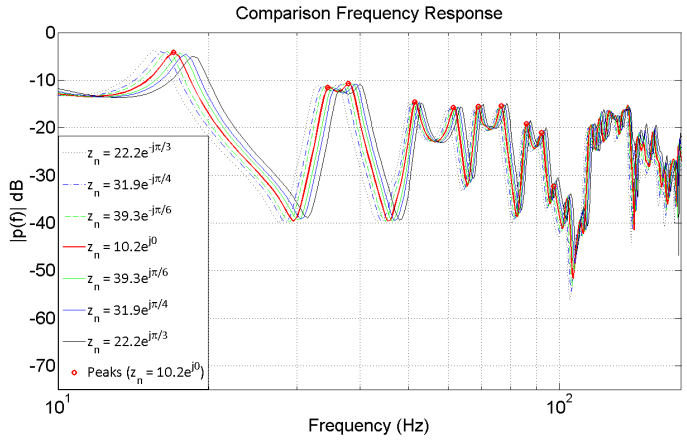


FIGURE 3.51: Comparison between purely resistive and complex impedances. Frequency responses have been calculated in an arbitrary point of the rectangular room using FEM

The findings are consistent with the analytical model of Figure 3.49. The frequency response spectrum is shifted to the right for positive reactances and to the left for negative reactances when compared to the purely resistive case. The frequency shift depends on the magnitude of the imaginary part. The frequency of the peaks (in Hz) are shown in Table 3.5 ($z_n = |z_n|\, e^{j0}$ was selected as reference). This suggests that the frequency shift tends to be consistent keeping the relative frequency difference between peaks.

φ	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}	6^{th}	7^{th}	8^{th}	9^{th}	10^{th}
$-\frac{\pi}{3}$	15.5	33	36	50	59.5	67.5	74.5	84.5	90	95.5
$-\frac{\pi}{4}$	16	33.5	37	50.5	60.5	68	75.5	85	91	96
$-\frac{\pi}{6}$	16.5	34	37.5	51	61	68.5	76	75.5	91.5	96.5
0	17	34.5	38	51.5	61.5	69	76.5	86	92	97.5
$\frac{\pi}{6}$	17.5	35.5	38.5	52	62.5	69.5	77	86.5	93	98
$\frac{\pi}{4}$	18	36	39	52.5	63	69.5	77.5	87	93.5	98.5
$\frac{\pi}{6}$	18.5	36.5	40	53	63.5	70.5	78.5	87.5	94	99.5

TABLE 3.5: Frequency (Hz) of the peaks according to the angle of the impedance $z_n = |z_n|\, e^{j\varphi}$

Nevertheless, it is important to point out that the use of the specific acoustic impedance to characterize the boundary conditions presents a constraint at frequencies equal or close to 0 Hz. In this case, the last term of equation 3.7 tends to 0 leading to a resonant system without attenuation. The predicted acoustic pressure at these frequencies will contain a significant amount of energy, which require to be reduced by means of a filtering process.

3.3.2.4 Results

Comparison with geometrical acoustics

A comparison between the GA and FE results was conducted to determine if the latter approach provides a better accuracy at low frequencies. Figures 3.52 to 3.57 show the frequency response in narrow band and in 1/3 octave band resolution for each source-receiver path. In addition, the mean errors between the 1/3 octave bands of the predicted and measured data have been estimated to evaluate which simulation

method provides a more accurate energy representation. These are indicated in each figure. The ME is defined as

$$\mathrm{ME(dB)} = \frac{1}{n}\sum_{i=1}^{n}\left|10\log_{10}(|\widetilde{p}_i|^2) - 10\log_{10}(|p_i|^2)\right|, \tag{3.15}$$

in which n is the number of 1/3 octave frequency bands, $|\widetilde{p}_i|^2$ and $|p_i|^2$ are the predicted and reference energy of the acoustic pressure in the 1/3 octave band i, respectively. This error considers an equal contribution of all the 1/3 frequency bands, thus being similar to a model in which pink noise have been used as input signal. It was selected to provide an insight of how dissimilar on average the reconstructed signal is with respect to the reference one.

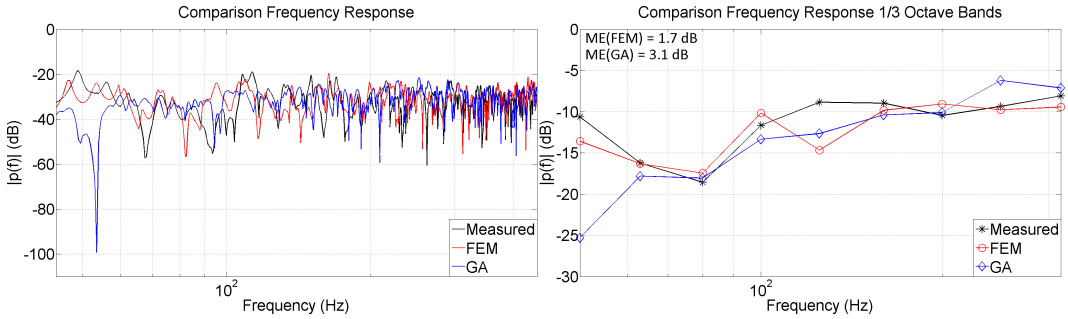


FIGURE 3.52: Comparison frequency response, FEM and GA meeting room, path B0-R1

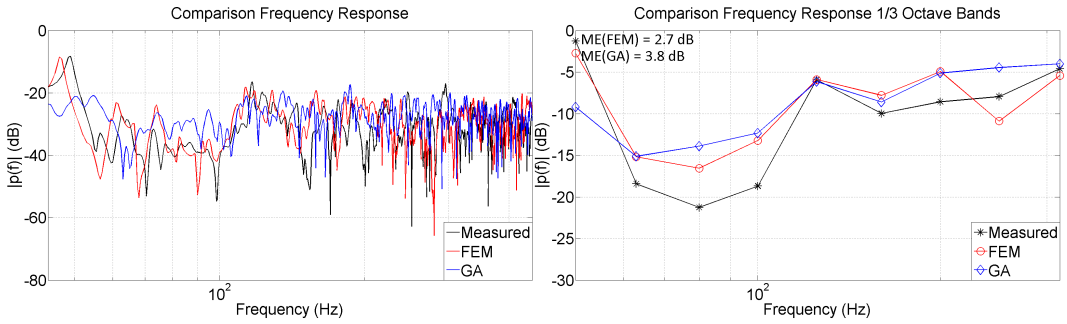


FIGURE 3.53: Comparison frequency response, FEM and GA, meeting room, path B0-R2

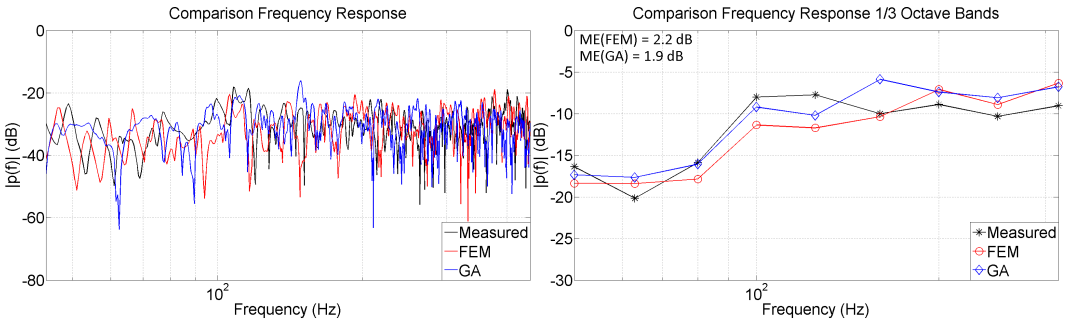


FIGURE 3.54: Comparison frequency response, FEM and GA, meeting room, path B0-R3

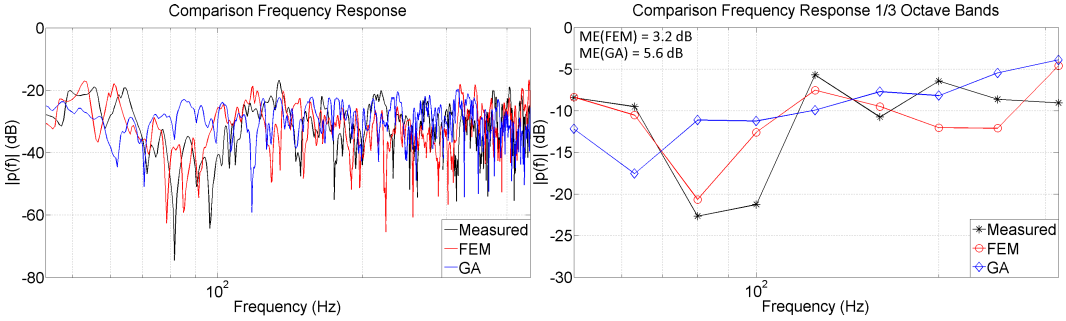


FIGURE 3.55: Comparison frequency response, FEM and GA, meeting room, path B1-R1

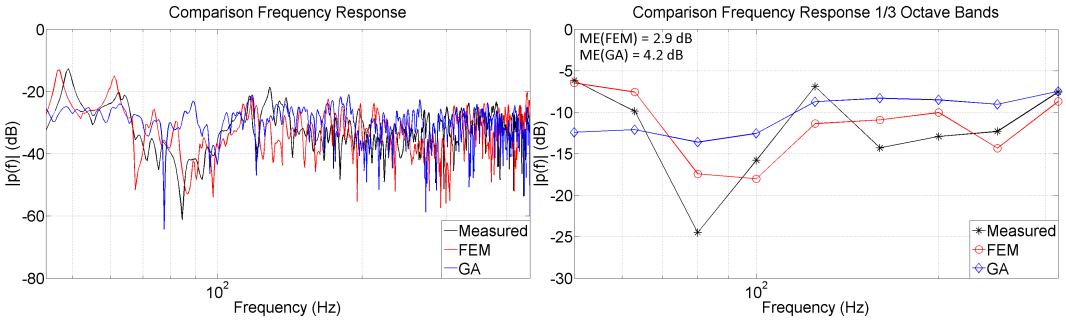


FIGURE 3.56: Comparison frequency response, FEM and GA, meeting room, path B1-R2

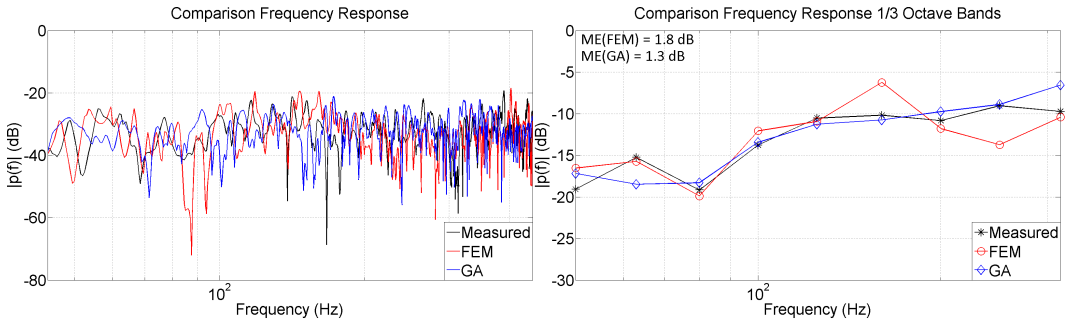


FIGURE 3.57: Comparison frequency response, FEM and GA, meeting room, path B1-R3

The results indicate that as expected, FEM yields a more accurate frequency response than GA for frequencies below 150 Hz. In general, the envelope of the frequency response is correctly predicted, but shifted with respect to frequency. The extent to which the frequency shift can be attributed to the assumption of a purely resistive specific acoustic impedance is now investigated, although other factors such as inaccuracy in the room dimensions and in the source-receiver positions may also contribute to this effect.

Figures 3.58 to 3.60 show the frequency response at the same receivers, but using complex specific acoustic impedances for all surfaces. The complex values of the impedance were obtained by an iterative adjustment of the phase of the impedance; i.e. a complex impedance was written as

$$z_{(complex)} = \frac{z_{(purelyresistive)}}{\beta}\left(\gamma + j\gamma^{'}\right), \tag{3.16}$$

where γ and $\gamma^{'}$ represent the ratio of resistance to reactance ($\gamma+\gamma^{'} = 1$) and the factor β is chosen so that $z_{complex}$ gives the same value of the diffuse field absorption coefficient as for the purely resistive case.

The response was computed for a range of different values of the ratio $\gamma : \gamma^{'}$ for the glazed/plastered surfaces and for the other materials. The best agreement in terms of frequency response was achieved for complex specific acoustic impedances with proportions of $\gamma = 0.2$ and $\gamma^{'} = 0.8$ in the glass and plaster and $\gamma = 0.4$ and $\gamma^{'} = 0.6$ in the remaining materials. The results indicate a much better agreement between the measured and predicted data for all source-receiver paths. This supports the

hypothesis that the frequency shift observed in Figures 3.52 to 3.57 is associated with the phase of the specific acoustic impedance and the "correct" choice of this quantity can lead to improved results at low frequencies. This emphases the importance of including phase information for impedance whenever this is available.

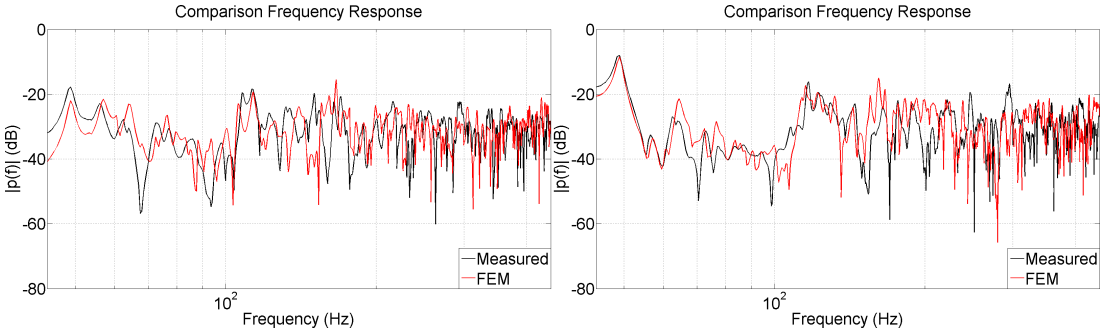


FIGURE 3.58: Comparison frequency response, FEM (complex impedance) and GA meeting room, Paths B0-R1 and B0-R2

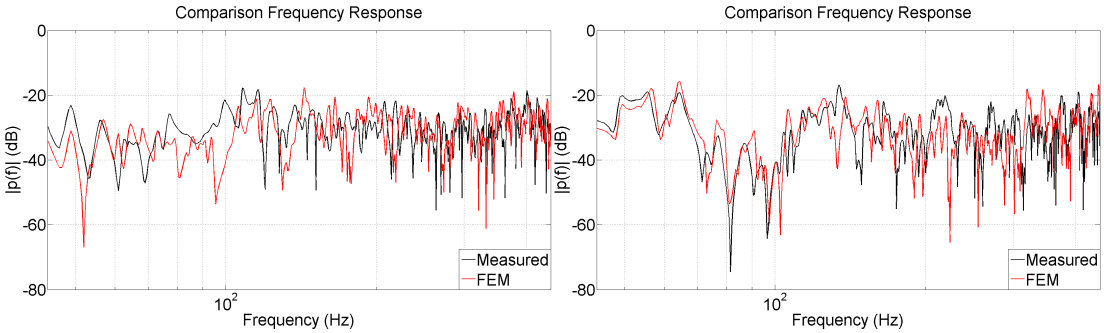


FIGURE 3.59: Comparison frequency response, FEM (complex impedance) and GA meeting room, Paths B0-R3 and B1-R1

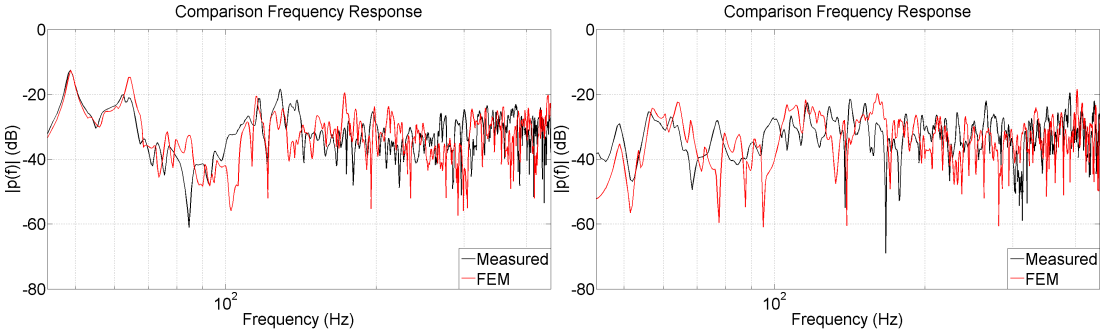


FIGURE 3.60: Comparison frequency response, FEM (complex impedance) and GA meeting room, Paths B1-R2 and B1-R3

Objective room acoustic parameters are presented in section 3.4 when the two simulation methods (FEM+GA) are combined.

3.3.3 Ightham Mote

3.3.3.1 The geometric model of the room

The model implemented in the simulations for the Ightham Mote corresponded to the most complex one used for GA. It contains all the relevant details of the geometry such as the edges of the ceiling, door frames and furniture. As an example, Figure 3.61 illustrates the complexity of the roof, which was replicated from the data provided by the total station. Nevertheless, a relevant amount of decorative elements and small furniture were neglected due to their complexity.

FIGURE 3.61: Detail of the ceiling, Ightham Mote

The same methodology was used for the discretization of the domain creating the meshes according to different octave bands. Figure 3.62 shows the meshes of the enclosure with the number of nodes according to the octave band to be simulated.

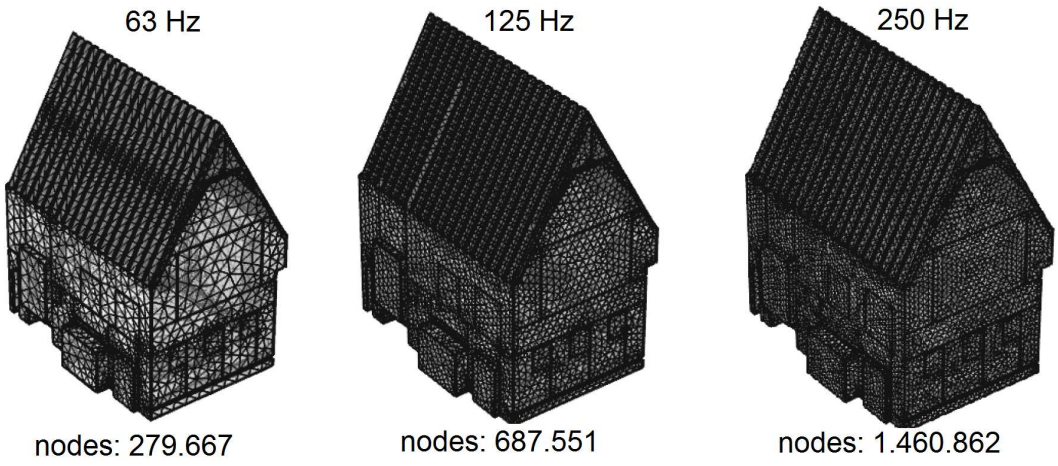


FIGURE 3.62: Resolution of meshes, Ightham Mote

3.3.3.2 Characterization of the boundary conditions

Based on the approach implemented for the meeting room, specific acoustic impedance values, which are first assumed to be purely resistive were calculated. A particular condition was found in the windows because they are subdivided by a small glass pieces in a stained glass style. This circumstance benefits the assumption of local reacting surfaces and the use of the real part of the specific impedance due to the reduction of the bending waves which are common in normal windows.

3.3.3.3 Results

Figures 3.63 to 3.68 compare the frequency response in narrow band and in 1/3 octave band resolution obtained from FEM and GA simulations for each source-receiver path with the measurements.

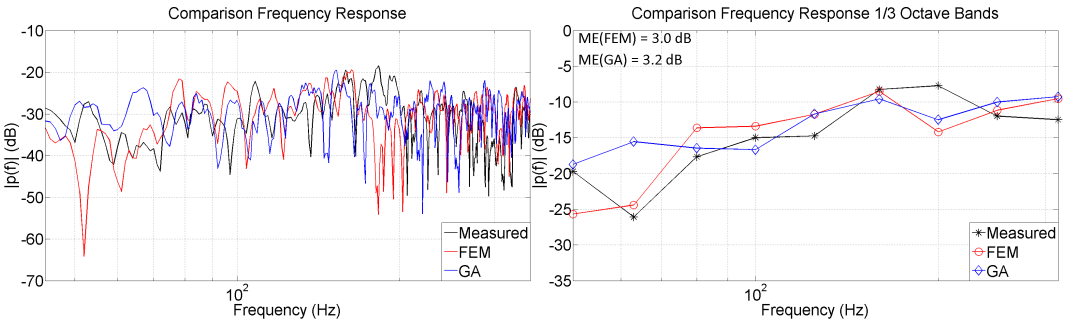


FIGURE 3.63: Comparison frequency response, FEM and GA, Ightham Mote, path B0-R1

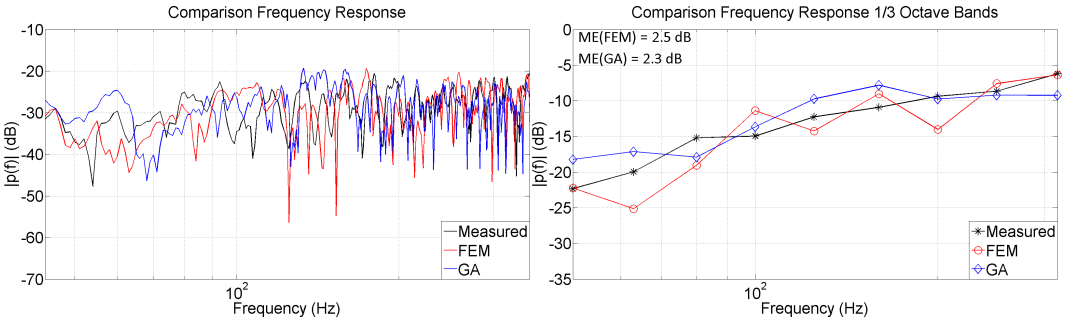


FIGURE 3.64: Comparison frequency response, FEM and GA, Ightham Mote, path B0-R2

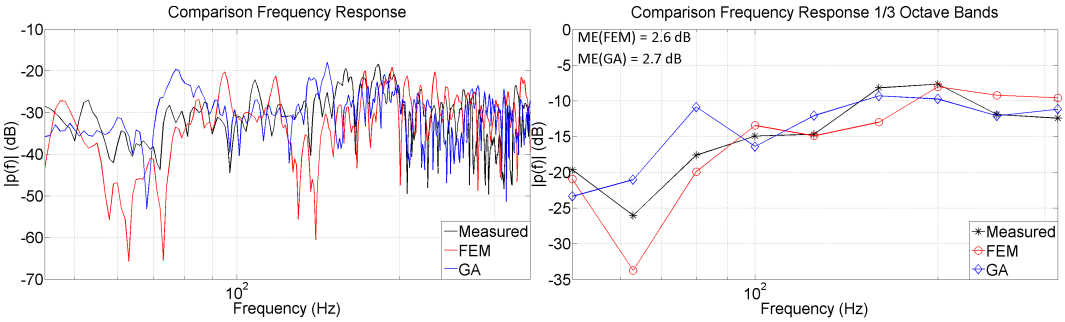


FIGURE 3.65: Comparison frequency response, FEM and GA, Ightham Mote, path B0-R3

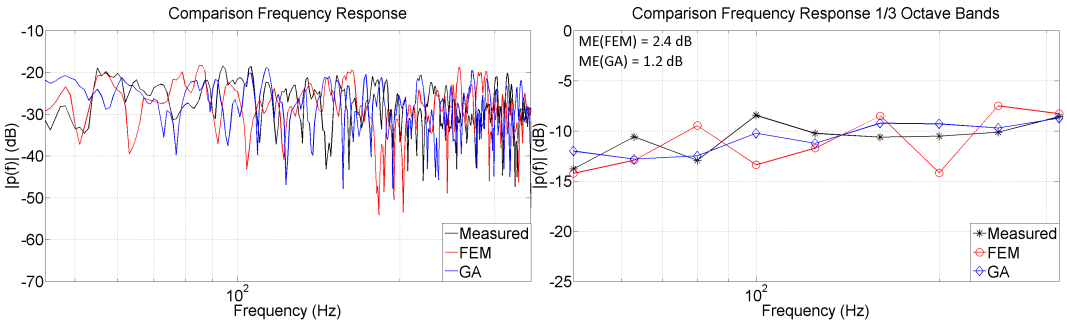


FIGURE 3.66: Comparison frequency response, FEM and GA, Ightham Mote, path B1-R1

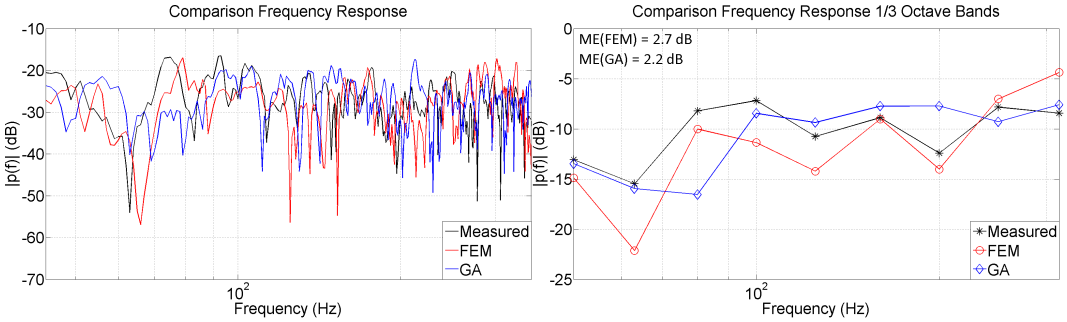


FIGURE 3.67: Comparison frequency response, FEM and GA, Ightham Mote, path B1-R2

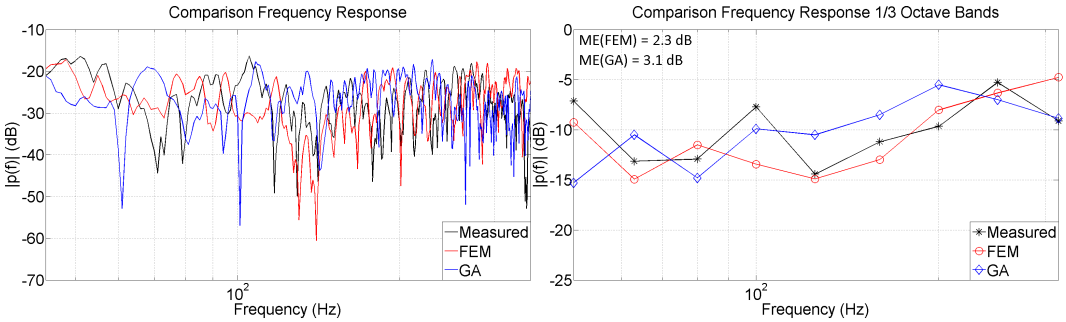


FIGURE 3.68: Comparison frequency response, FEM and GA, Ightham Mote, path B1-R3

Despite all the considerations made in the simulations, the frequency responses could not be reconstructed with the same level of accuracy as for the meeting room. A better agreement was found for the source position B1, which suggests that the discrepancy for the source position B0 may be due to a mismatch in the representation of the source-receiver locations. Regarding the mean errors, the finite element method

leads to slightly better results compared to geometrical acoustics for some source-receiver paths, but it is not the general case for all of them. It is considered that the main reason for the disagreement is due to the complexity of the enclosure. Predictions were conducted with a simplified model in which elements that are part of the decoration and furniture of the room were omitted. In some areas of the Ightham Mote the walls correspond to thin panels of wood, which in the case of GA, its transmission loss is represented as a higher absorption coefficient. However, in FEM this condition constrains the assumption of locally reacting surfaces due to the bending waves. An improvement of the results would demand a detailed analysis about the characterization of the geometric model and the boundary conditions, which is beyond of the scope of this research. Objective room acoustic parameters computed from the combination of FEM and GA simulations are illustrated in the next section when the two simulation methods are combined.

3.3.4 Office room

FE simulations of the office room were conducted following the procedure established for the meeting room. A common model with respect to GA simulations was implemented using coarser meshes at low frequencies. The criteria of 6 nodes per wavelength in the mesh resolution was adopted. The predictions were carried out up to 708 Hz. Although at the octave band of 500 Hz the directivity of the acoustic source is not omnidirectional, its location within the room (at the corner) decreases its effects. Figure 3.69 illustrates the frequency response in narrow band and in 1/3 octave band resolution of the measured and predicted RIRs by means of FEM and GA. As expected, the results indicate that the finite element method is more accurate compared to geometrical acoustics. A shift in the frequency response was also found using real specific acoustic impedance values.

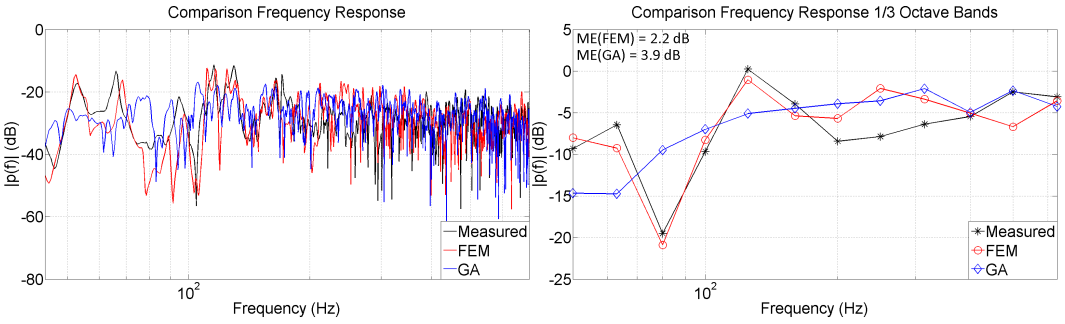


FIGURE 3.69: Comparison frequency response, FEM and GA office room, path B0-R1

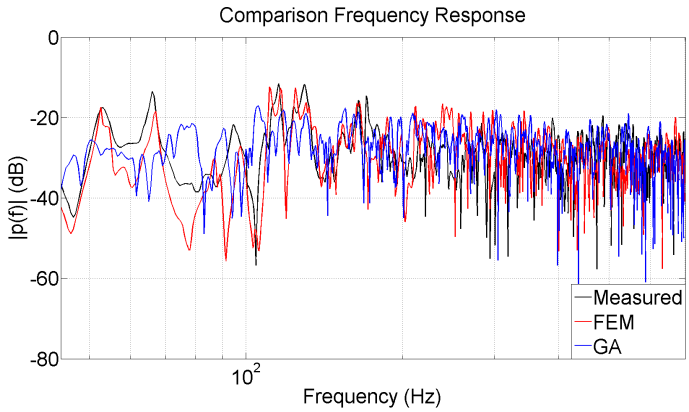


FIGURE 3.70: Comparison frequency response, FEM and GA office room (complex impedance), path B0-R1

Objective room acoustic parameters are presented in the next section when the two simulation methods (FEM+GA) are combined.

3.4 Combination of simulation methods

Previous sections provide the general framework to synthesize room impulse responses based on geometrical acoustics and the finite element method. In this section, a procedure is now proposed to combine both methods to generate unified RIRs. The unification of the results from each method was conducted in the frequency domain. For this, the RIRs obtained from the GA model with zero padding up to 2 s were transformed into the frequency domain by applying a Fourier transform. The zero padding was implemented to obtain a frequency resolution of 0.5 Hz. The maximum frequency for the FE simulations was chosen according to the specific enclosure. The values were 447 Hz, 355 Hz and 708 Hz for the meeting room, Ightham Mote, and office room, respectively. The frequency resolution used for the FE calculations corresponded to 0.5 Hz ensuring the same resolution as the GA data.

A crossover formed by a low pass and a high pass filter was designed to unify the results. The central frequency of the crossover was always selected at one 1/3 octave band lower than the maximum FE simulated frequency allowing the low pass filter to reduce the sharp discontinuity due to the truncation of the FE data. The crossover frequencies were 355 Hz, 282 Hz and 562 Hz for the meeting room, Ightham Mote, and office room, respectively. The low and high pass filters were designed with the

aim of obtaining a uniform magnitude and phase in the crossover point avoiding coloration in the frequency response. 8th order Butterworth filters were used for this purpose; high order in the filters was required to ensure rapid decay in the crossover area. Figure 3.71 illustrates the magnitude and phase of the crossover at the central frequency for the meeting room reference case. The low and high pass filters were applied to the FE and GA data, respectively. Then, the filtered signals were added together and transformed into the time domain by applying an inverse Fourier transform.

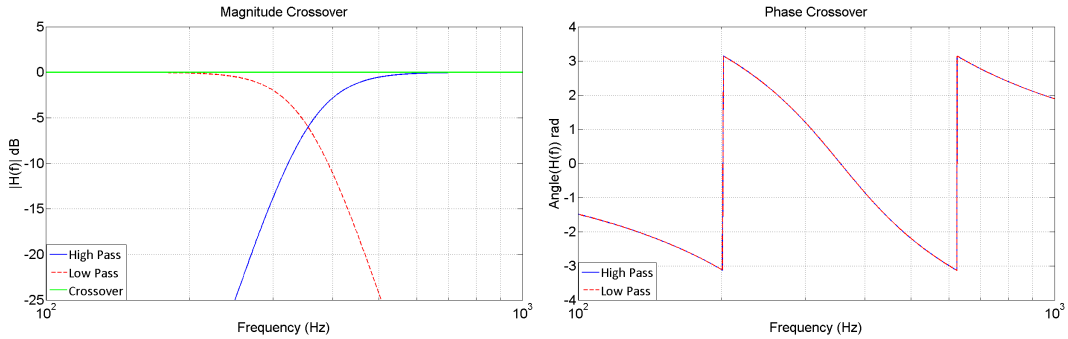


FIGURE 3.71: Magnitude and phase of the crossover network.

3.4.1 Objective room acoustic parameters computed from unified (FEM+GA) room impulse responses

The T_{20}, EDT and C_{80} estimated from impulse responses whose frequency content have been predicted using a combination of the finite element method and geometrical acoustics is presented below for each enclosure. The source-receiver paths correspond to the locations defined in Figures 3.2, 3.7 and 3.12, respectively.

3.4.1.1 Meeting room

Figure 3.72 shows the comparison between the predicted and measured spatially averaged reverberation time. The early decay time and clarity index are illustrated for each receiver-source path in Figures 3.73 to 3.78. The comparison of the results to the acoustic parameters derived from only GA simulations (Figures 3.18 to 3.24) indicates that the inclusion of the finite element method improves the accuracy at low frequencies.

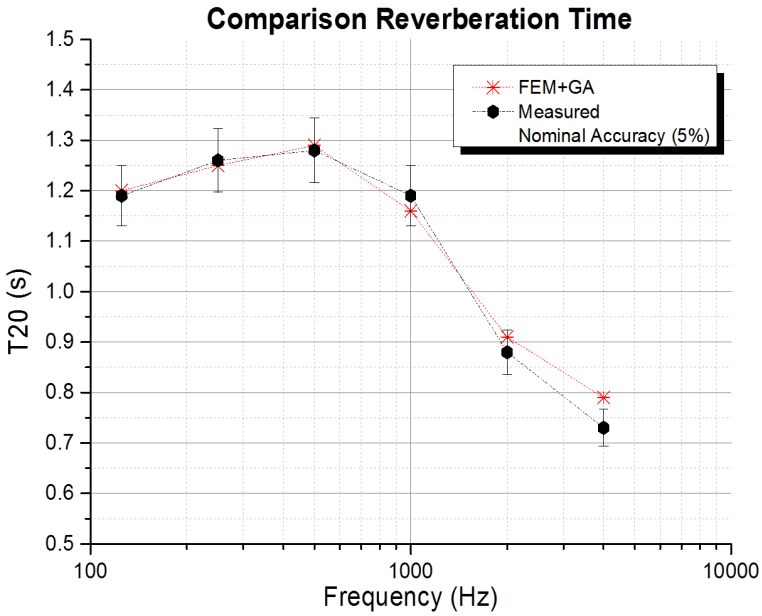


FIGURE 3.72: Predicted T_{20} meeting room, FEM + GA

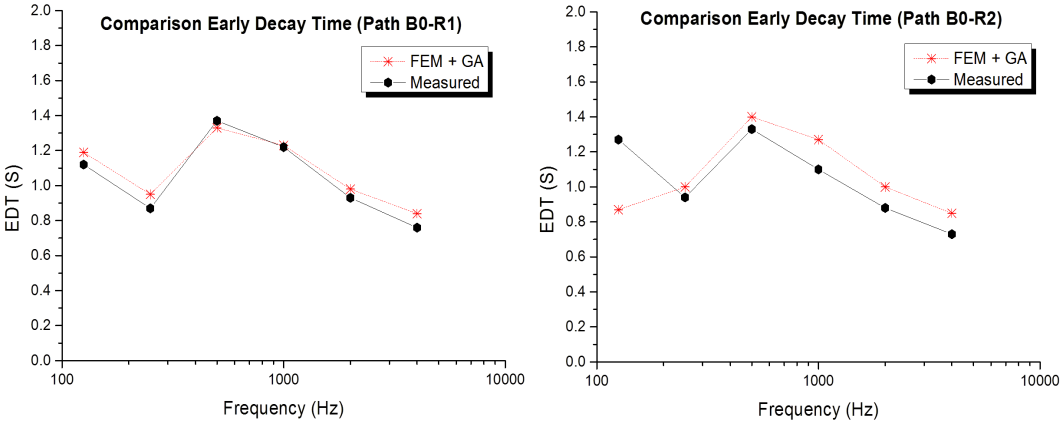


FIGURE 3.73: Comparison *EDT* meeting room, FEM + GA, paths B0R1 & B0R2

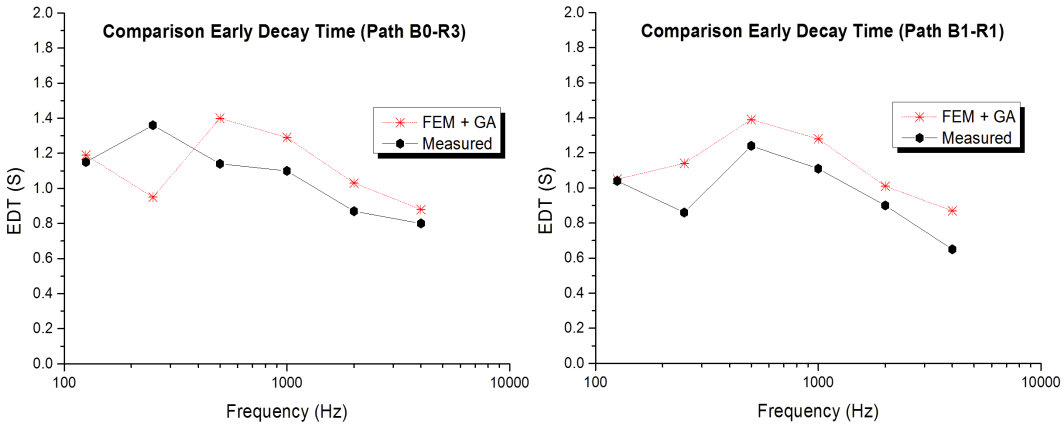


FIGURE 3.74: Comparison *EDT* meeting room, FEM + GA, paths B0R3 & B1R1

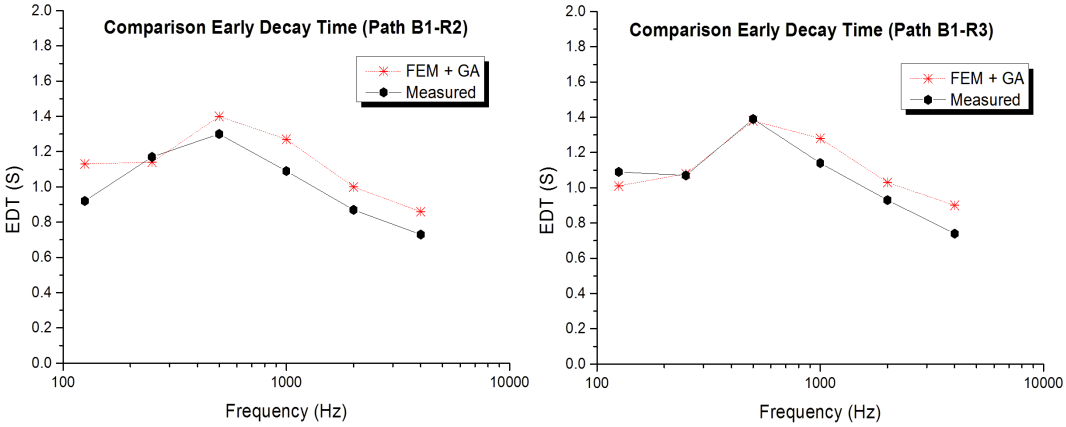


FIGURE 3.75: Comparison *EDT* meeting room, FEM + GA, paths B1R2 & B1R3

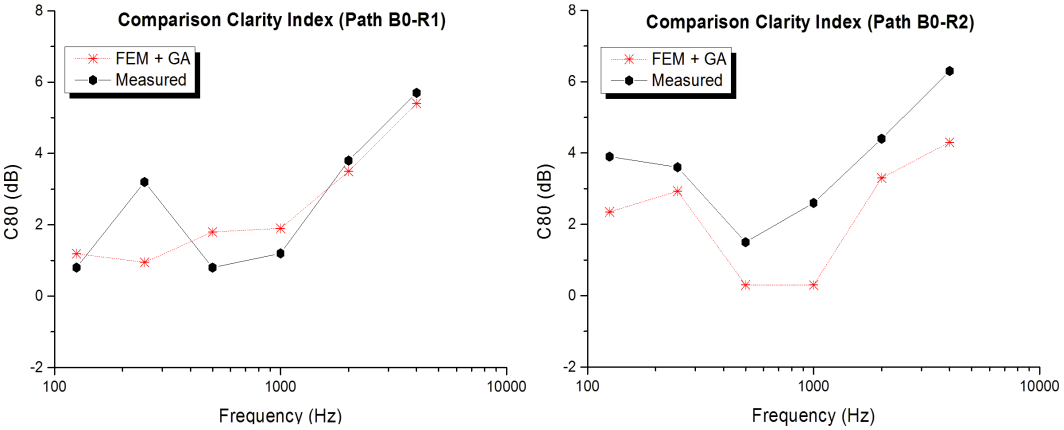


FIGURE 3.76: Comparison C_{80} meeting room, FEM + GA, paths B0R1 & B0R2

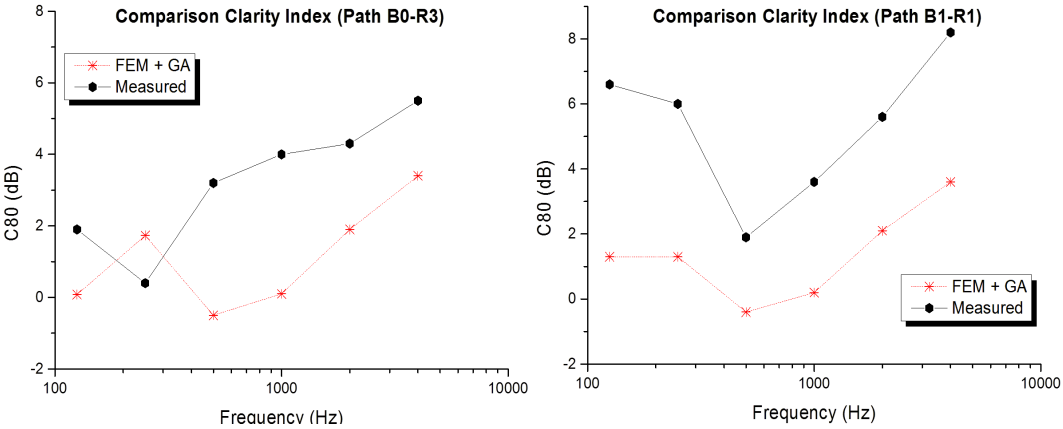


FIGURE 3.77: Comparison C_{80} meeting room, FEM + GA, paths B0R3 & B1R1

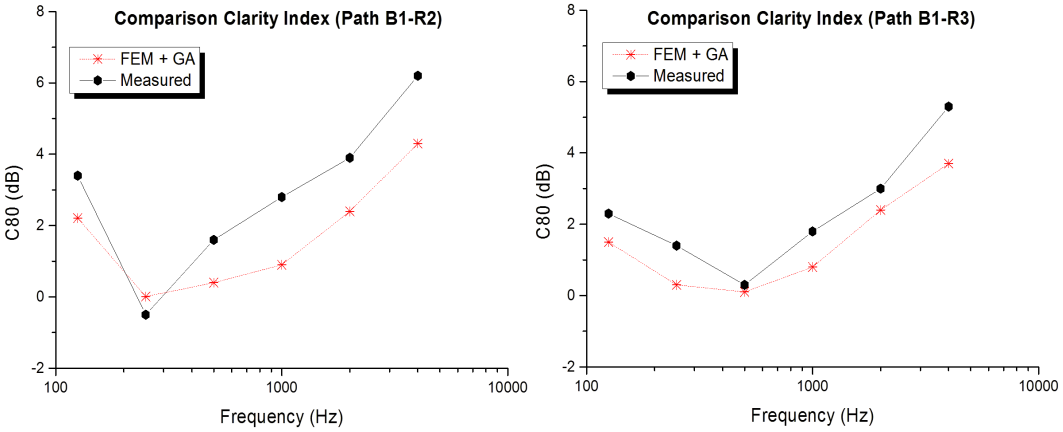


FIGURE 3.78: Comparison C_{80} meeting room, FEM + GA, paths B1R2 & B1R3

3.4.1.2 Ightham Mote

For the Ightham Mote the spatially averaged T_{20} is presented in Figure 3.79. The EDT and C_{80} for each source-receiver combination is displayed in Figures 3.80 to 3.85.

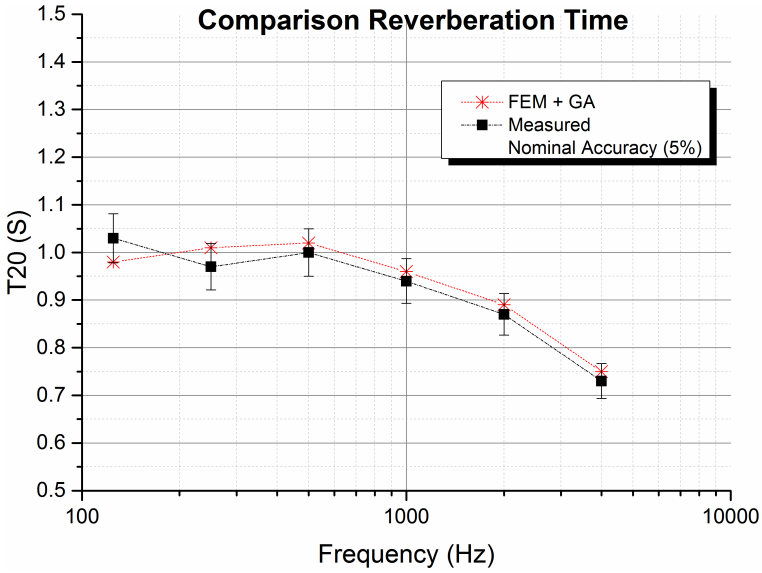


FIGURE 3.79: Predicted T_{20} Ightham Mote, FEM + GA

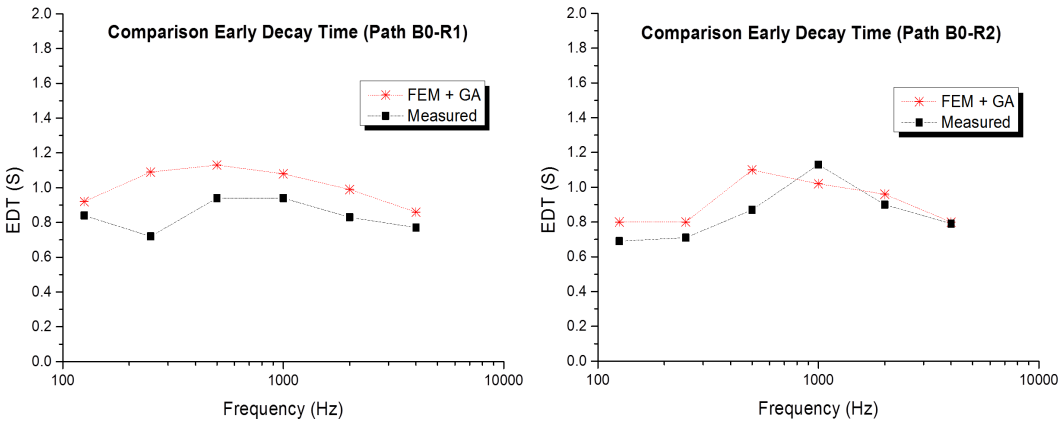


FIGURE 3.80: Comparison *EDT* Ightham Mote, FEM + GA, paths B0R1 & B0R2

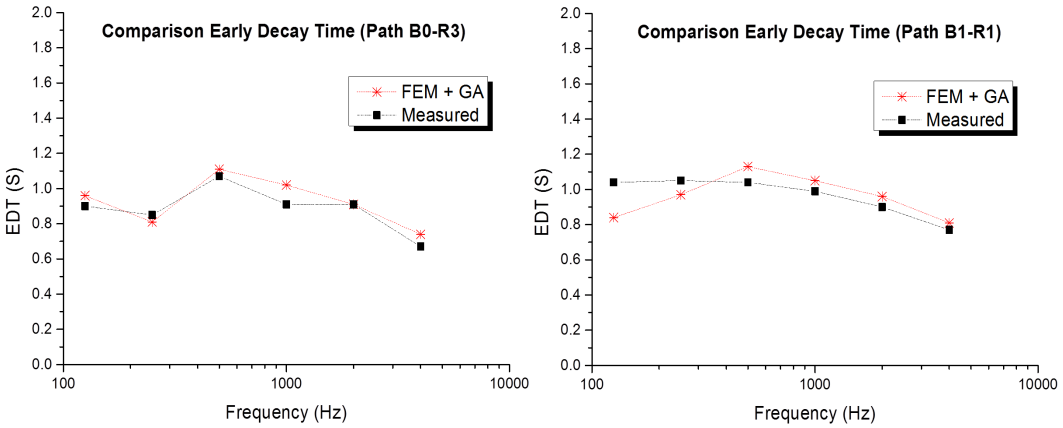


FIGURE 3.81: Comparison *EDT* Ightham Mote, FEM + GA, paths B0R3 & B1R1

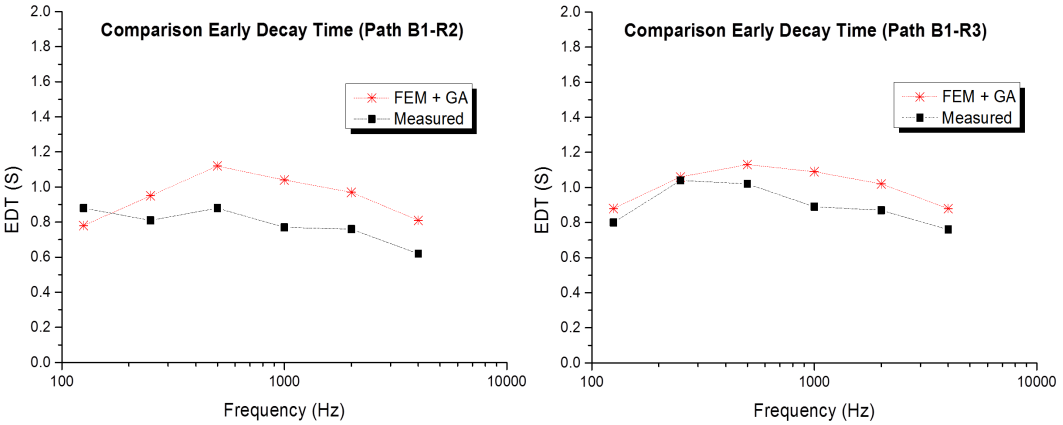


FIGURE 3.82: Comparison *EDT* Ightham Mote, FEM + GA, paths B1R2 & B1R3

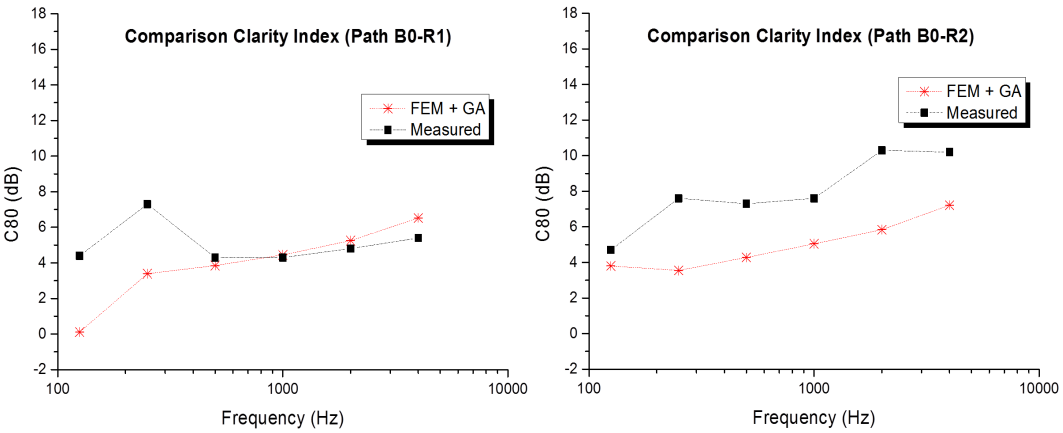


FIGURE 3.83: Comparison C_{80} Ightham Mote, FEM + GA, paths B0R1 & B0R2

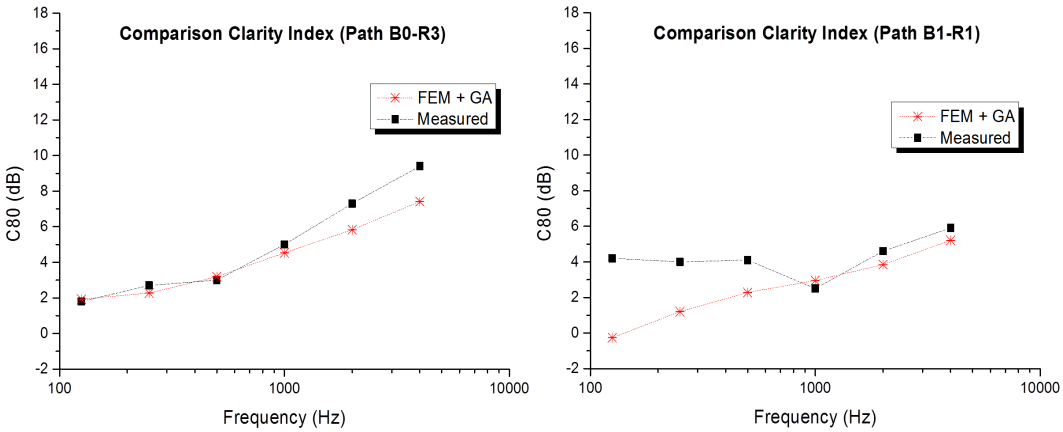


FIGURE 3.84: Comparison C_{80} Ightham Mote, FEM + GA, paths B0R3 & B1R1

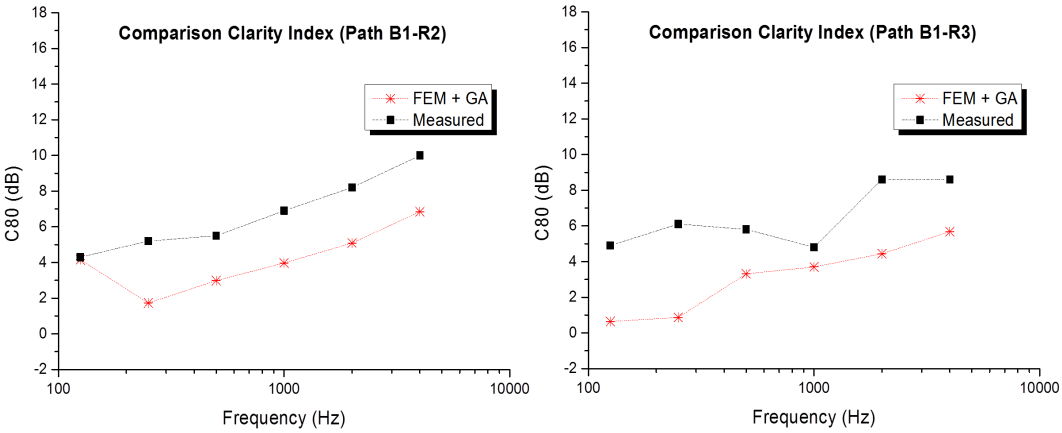


FIGURE 3.85: Comparison C_{80} Ightham Mote, FEM + GA, Paths B1R2 & B1R3

The analysis of the objective room acoustic parameters indicates that the inclusion of the finite element method leads to a good agreement in terms of reverberation time and early decay time. In particular, the EDT is predicted with higher accuracy compared to GA (Figures 3.30 to 3.32) in most of the source-receiver paths for both source positions. In contrast, the estimation of the C_{80} shows an improvement only for the path B0R3. In the other paths the C_{80} results do not indicate any benefit for this specific this enclosure.

3.4.1.3 Office room

Figures 3.86, 3.87, and 3.88 show the T_{20}, EDT and C_{80}, respectively for the office room. The outcomes confirm that the inclusion of FEM in the prediction of the low

frequency content of the room impulse response yields more accurate results at low frequencies compared to the use of only GA methods (Figures 3.44 to 3.46).

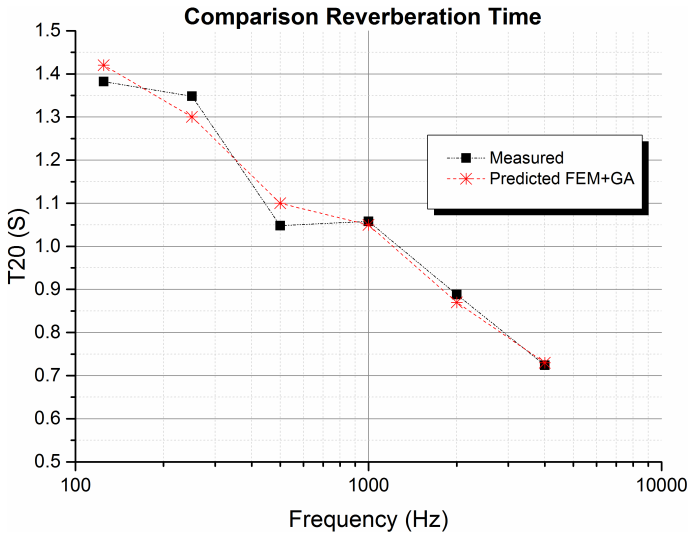


FIGURE 3.86: Predicted T_{20}, office room, FEM+GA

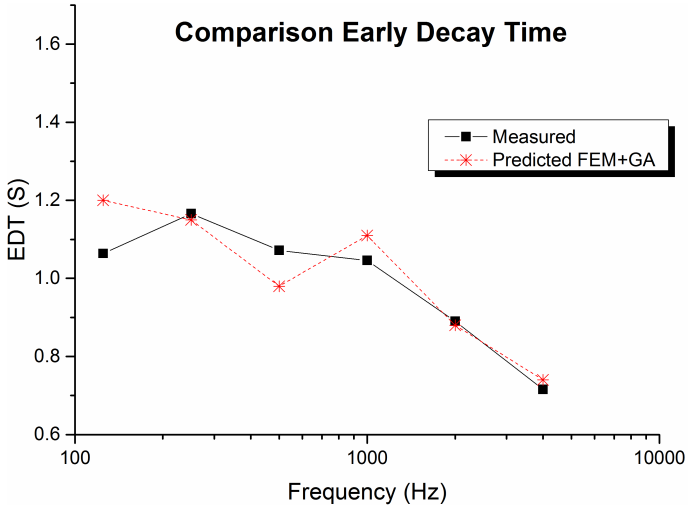


FIGURE 3.87: Predicted EDT, office room, FEM+GA

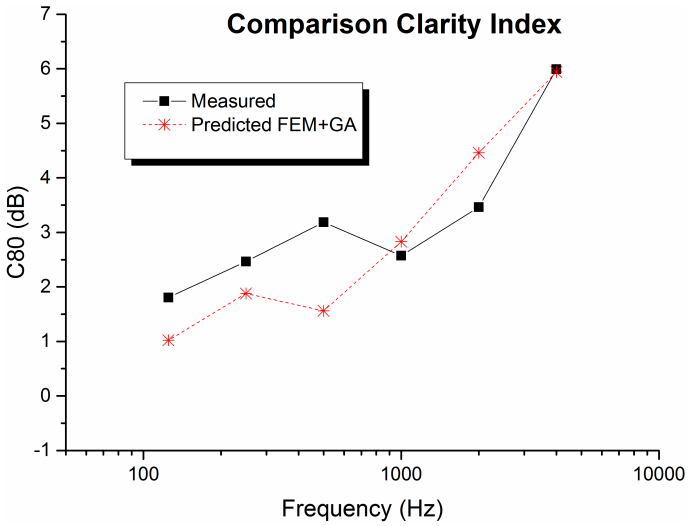


FIGURE 3.88: Predicted C_{80}, office room, FEM+GA

3.5 Discussion

An overall basis for the prediction of room impulse responses using a combination of the finite element method and geometrical acoustics has been discussed. Modelling parameters such as the definition of the acoustic source, the level of geometrical detail in the models and the characterization of the boundary conditions have been discussed for each approach. A method for the combination of FE and GA results has been proposed based on the application of IIR filters in the frequency domain. The results indicate that the use of reverberation time measurements is a suitable approach to determine the absorption coefficients and the effective specific acoustic impedance values for simple geometries such as the meeting room and the office room. In these cases, the inclusion of the finite element method yielded more accurate results in the prediction of the room impulse responses. In complex enclosures (e.g. Ightham Mote), further investigations are required to determine a more effective methodology to perform FE-GA simulations. The difficulty to obtain an accurate prediction of the room impulse response in complex environments has been already pointed out by Aretz [131] who tried to synthesize RIRs of a recording studio. The acoustic simulation of complex spaces remains as an open question to be solved in future.

Chapter 4

Spatial Encoding of Numerical Simulations

This chapter provides a framework about the generation of a plane wave representation of sound fields that have been predicted by means of the finite element method and geometrical acoustics. The plane wave representation has been selected because it allows for interactive features that will be discussed in Chapter 5. In addition, several sound reproduction techniques can be implemented for the rendering of the acoustic field, which makes this approach convenient for auralization purposes. Firstly, an evaluation about the suitability of an inverse method to estimate the complex amplitude of a set of planes waves that reconstruct the sound field around the receiver position is carried out. This method is convenient to tackle the lack of directional information about the field provided by the methods based on the numerical solution of the wave equation. The relation of this sound field representation with a spherical harmonic basis is also considered, which is useful to perform interactive transformations in different domains. Limitations of the current approach are discussed and techniques to improve the sound field reconstruction are presented. Differently from FEM, results from the GA simulations contain directional information related to the incoming sound reflections. However, this data must be processed in order to obtain a suitable representation which is common to the FE results. Several approaches are addressed to discretize the direction of reflections into a finite number of directions determined by the number of plane waves used to process the FE data.

4.1 Implementation of an inverse method to synthesize spatial acoustic information

4.1.1 Plane wave expansion

The aim of this section is to achieve a plane wave representation of predicted sound fields based on the theoretical background given in section 2.1.7.1. To this end, results from FE simulations are processed using the concept of a virtual microphone array and an inverse problem formulated to estimate the complex amplitude of the plane waves [139]. Issues related to the size of the array, number of plane waves and ill-conditioned propagation matrices are addressed. The suitability of this technique to obtain spatial information is also discussed making emphasis on parameters such as distance to the acoustic source, area of accurate reconstruction and the complexity of the target sound field.

The complex acoustic pressure at the point $\mathbf{x}$ based on a plane wave expansion is given by:

$$p(\mathbf{x}) = \int_{\widehat{\mathbf{y}}\in\Omega} e^{jk\mathbf{x}\cdot\widehat{\mathbf{y}}} q(\widehat{\mathbf{y}}) d\Omega(\widehat{\mathbf{y}}), \tag{4.1}$$

in which $\widehat{\mathbf{y}}$ indicates the different incoming directions of the plane wave density and $q(\widehat{\mathbf{y}})$ is the amplitude density function. Due to the implementation of the method, equation (4.1) must be discretized into a finite number of L plane waves, namely

$$p(\mathbf{x}) = \sum_{l=1}^{L} e^{jk\mathbf{x}\cdot\widehat{\mathbf{y}}_l} q(\widehat{\mathbf{y}}_l) \Delta\Omega_l, \tag{4.2}$$

where $\Delta\Omega_l$ is the area attributed to each direction $\widehat{\mathbf{y}}_l$ as result of the discretization of the integral representation. The discretization of equation (4.1) is performed using a predefined uniform distribution of L plane waves over a unit sphere [140]. Nevertheless, the use of a finite number of plane waves generates artifacts in the sound field reconstruction. Figure 4.1 shows the real part of the acoustic pressure in Pascals for a plane wave of 250 Hz coming from an elevation $\theta = 90°$ and azimuth angle of $\phi = 45°$. The number of plane waves used in the expansion corresponds to 64 and 144 (see Figure 4.2), respectively. The black circle represents the central point of the expansion. The results indicate that the area where the acoustic field is

accurate depends on the number of plane waves implemented for the synthesis, being wider at a higher number of plane waves.

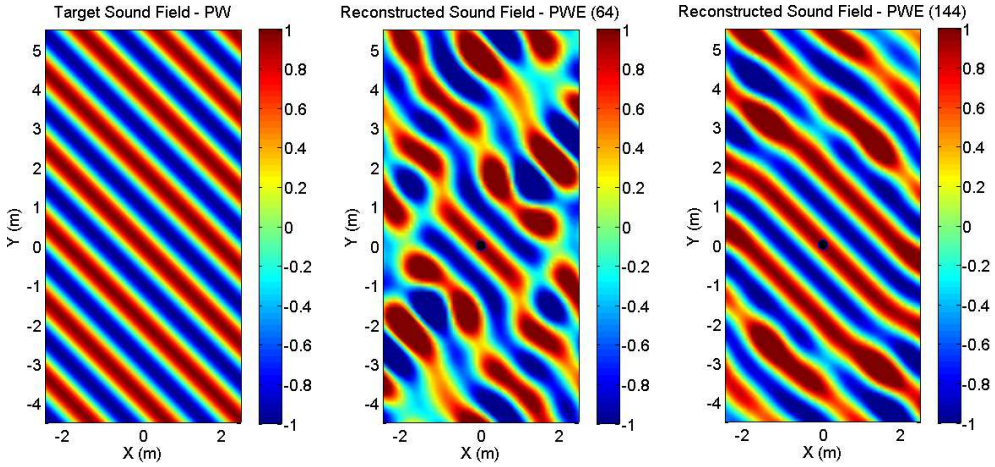


FIGURE 4.1: Comparison of sound field reconstruction at 250 Hz using 64 and 144 plane waves

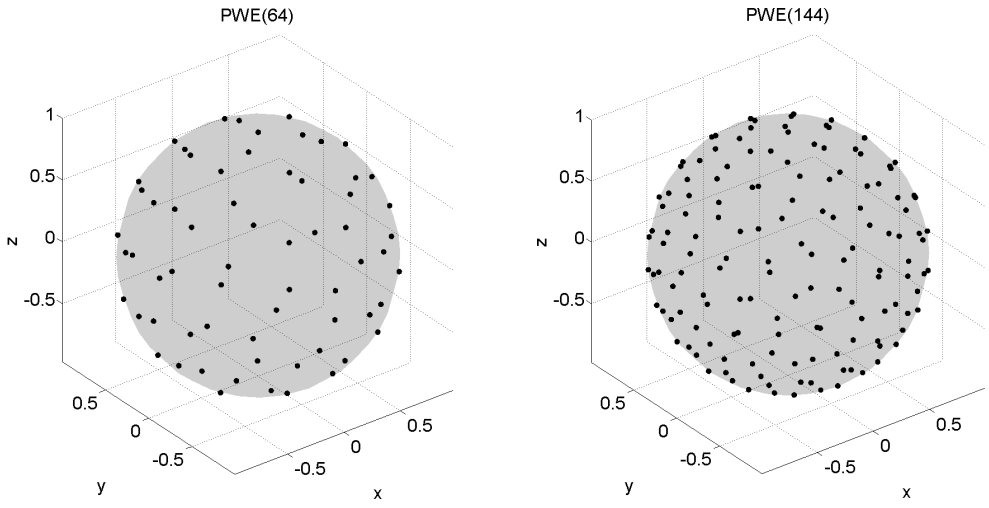


FIGURE 4.2: Directions of the PWE using 64 and 144 plane waves

4.1.1.1 Formulation of the inverse problem

An implementation of an inverse method is proposed to calculate the amplitudes q of a set of L plane waves that reconstruct a selected target acoustic field. To that end, the acoustic pressure estimated from FEM at a specific location of the domain can be understood as corresponding to the output from an omnidirectional microphone. The combination of different acoustic pressure points generates a virtual microphone

array which can be used to extract spatial information of the sound field. Based on that information, the amplitude q of each plane wave is determined by the inversion of the transfer function matrix between microphones and plane waves [139]. This technique has been successfully implemented by Støfringsdal and Svensson for 2D cases [13]. The principle is explained as follows: the complex acoustic pressure at M virtual microphone positions is denoted using vector notation as

$$\mathbf{p}(\omega) = [p_1(\omega), p_m(\omega), \ldots, p_M(\omega)]^T, \tag{4.3}$$

where p_m is the acoustic pressure at the m-th virtual microphone. Likewise, the complex amplitudes of L plane waves used to reconstruct the sound field are represented by the vector

$$\mathbf{q}(\omega) = [q_1(\omega), q_l(\omega), \ldots, q_L(\omega)]^T. \tag{4.4}$$

Finally, the transfer function that describes the sound propagation from each plane wave to each virtual microphone can be arranged in matrix notation as:

$$\mathbf{H}(\omega) = \begin{vmatrix} h_{11}(\omega) & \cdots & h_{1L}(\omega) \\ \vdots & h_{ml}(\omega) & \vdots \\ h_{M1}(\omega) & \cdots & h_{ML}(\omega) \end{vmatrix}$$

in which $h_{ml} = e^{jk\mathbf{x}_m \cdot \hat{\mathbf{y}}_l}$. Consequently, the relationship between the plane wave amplitudes and the virtual microphone signals is

$$\mathbf{p}(\omega) = \mathbf{H}(\omega)\mathbf{q}(\omega). \tag{4.5}$$

The amplitude of the plane waves is calculated by solving equation (4.5) for $\mathbf{q}(\omega)$. This is carried out in terms of a least squares solution, which minimizes the sum of the squared errors between the reconstructed and the target sound field [139]. In the case of an overdetermined problem (more virtual microphones than plane waves), the error vector can be expressed as

$$\mathbf{e}(\omega) = \tilde{\mathbf{p}}(\omega) - \mathbf{p}(\omega), \tag{4.6}$$

where $\tilde{\mathbf{p}}(\omega)$ is the pressure reconstructed by the plane wave expansion and $\mathbf{p}(\omega)$ is the target pressure from the FEM model. The least squares solution is achieved by the minimization of a cost function $J(\omega) = \mathbf{e}^H(\omega)\mathbf{e}(\omega)$ in which $(\cdot)^H$ indicates the Hermitian transpose. The minimization of the cost function $J(\omega)$ is given by [139]

$$\mathbf{q}(\omega) = \mathbf{H}^{\dagger}(\omega)\mathbf{p}(\omega), \tag{4.7}$$

in which $\mathbf{H}^{\dagger}(\omega)$ is the Moore-Penrose pseudo-inverse of the propagation matrix $\mathbf{H}(\omega)$ [139]. A preliminary analysis was performed in Matlab to establish the size of the microphone array and the number of plane waves required to generate an inverse matrix whose condition number is smaller than 10^6. The condition number is defined as the ratio between the largest and the smallest singular value of the propagation matrix $\mathbf{H}(\omega)$. It has been shown in [139] that the stability of the solution provided by the inverse method is determined by the condition number, therefore high values of this parameter indicate that errors in the model such as noise or non-linearity of the system will affect the result for $\mathbf{q}(\omega)$ significantly. Firstly, a simple incoming plane wave of 63 Hz ($\theta = 90°$, $\phi = 45°$) in free field was selected as a target. The sound field was analytically calculated in a rectangular domain of dimensions (5m, 10m, 3m) and captured by 4 different virtual cube arrays with linear dimensions of 0.8 m, 1.2m, 1.6m and 2 m, respectively. The spatial resolution between microphones corresponded to 0.1 m. 63 Hz was selected as a reference because the condition number decreases with frequency, thus providing a reasonable lower threshold. Table 4.1 shows the condition number of matrix $\mathbf{H}(\omega)$ for different sizes of array and number of planes waves uniformly distributed over a unit sphere [140].

Length of the array	$L = 64$	$L = 144$	$L = 324$
0.8 m (729 mics)	5.07e+08	2.91e+16	1.73e+17
1.2 m (2197 mics)	3.12e+07	1.80e+16	5.39e+16
1.6 m (4913 mics)	4.06e+06	1.21e+12	3.85e+16
2 m (9261 mics)	7.90e+05	8.53e+10	4.05e+16
2.4 m (15625 mics)	1.98e+05	9.47e+09	3.61e+16

TABLE 4.1: Condition number of the matrix $\mathbf{H}(\omega)$ as function of the size of the microphone array and the number of plane waves, 63 Hz

As expected, the condition number decreases as the size of the array increases. The reason for that may be attributed to the wavelength of the plane wave which is approximately 5.4 m at this frequency and a larger array captures more information

of the sound field. Regarding the number of plane waves, the results suggest that increasing its number does not improve the situation, instead, it compounds the condition number significantly. This can be explained because at 63 Hz there is not enough information in the sound field that is captured by the virtual microphone array, therefore additional plane waves only make the inversion of the propagation matrix $\mathbf{H}(\omega)$ more difficult. An optimal relation of 64 plane waves, a microphone array of 1.6 m of length with virtual microphone spacing of 0.2 m (729 microphone positions) was found at 63 Hz. This spatial resolution (0.2 m) yields an aliasing frequency of $\approx$ 850 Hz, which is sufficient for the range of the FE simulations. This high frequency limit is calculated based on the Nyquist theorem for sampling signals [141].

4.1.1.2 Sound field reconstruction

Acoustic simulations were conducted to analyze several aspects that affect the accuracy of the sound field reconstruction. Two sound fields corresponding to a plane (PW) and spherical (SW) wave propagation were analytically synthesized in a free field domain with dimensions of (5m, 10m, 3m) (see Figure 4.3). The incoming direction of the plane wave does not match with any incoming direction of the elements of the plane wave expansion. Samples of the sound fields were extracted by using a cubic virtual microphone array with linear dimensions of 1.6 m and a spatial resolution of 0.2 m (729 microphone positions). This information was used to estimate the complex amplitude of a set of plane waves. The number of plane waves was chosen to be ($L = 64$) because this, in conjunction with the number of virtual microphones, provides a low condition number of the matrix $\mathbf{H}(\omega)$. In addition, 64 is the number of modes for a spherical harmonic expansion of 7th order, which facilitates the comparison between approaches that is explained in section 4.1.3. Three type of figures are used to analyze the sound field reconstruction by means of the inverse method. The first one corresponds to the comparison between the real part of the target (analytical) and the reconstructed acoustic pressure (Pa) in a cross-section of the domain (z = 1.5m). Figure 4.4 provides an example of this kind of figure. The origin of the coordinate system in the figures corresponds to the central point of the expansion.

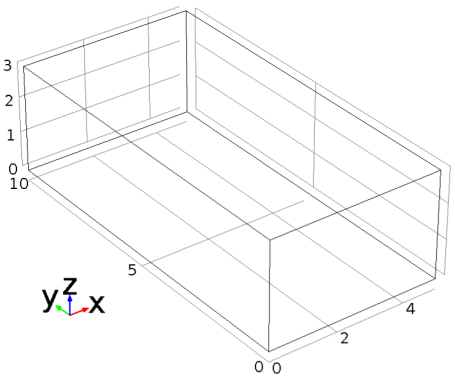


FIGURE 4.3: Domain used to analytically predict the free field sound propagation

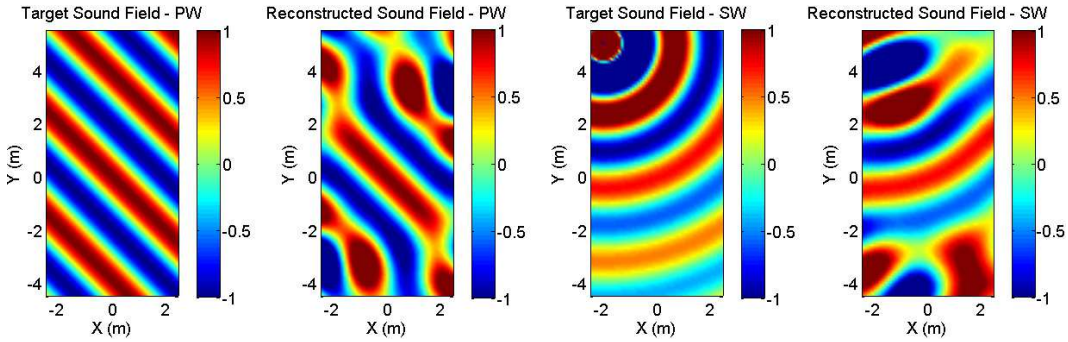


FIGURE 4.4: Example of figure type 1, the frequency corresponds to 125 Hz

The second type of figure illustrates an interpolated distribution of the complex amplitude of the plane waves. This is performed by unwrapping the unit sphere in a 2D plot whose axis represent the elevation and azimuth angle in degrees, respectively. This information is useful to assess the energy distribution of the plane wave density. An example is presented in Figure 4.5. The white circles identify the position of the plane waves. The total energy of the plane wave expansion is reported at the legend of this type of figure.

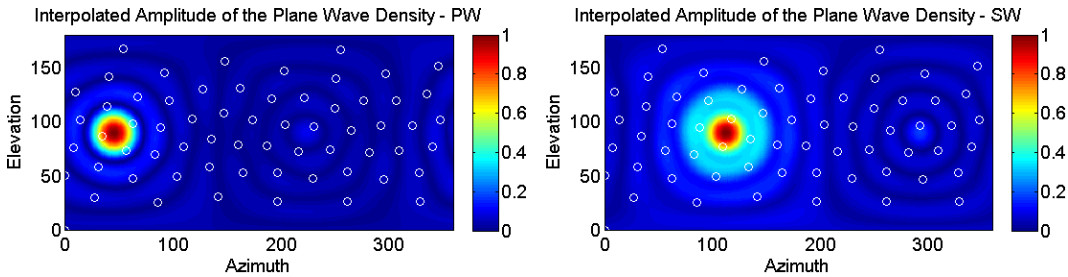


FIGURE 4.5: Example of figure type 2

The total energy of the plane wave expansion is calculate as

$$\mathrm{q}_{\mathrm{total}} = \sum_{L}^{l=1} |q_l|^2 \tag{4.8}$$

in which q_l is the complex amplitude of the plane wave l-th. Finally, the last type of figure illustrates 3 acoustic errors that have been adopted as a metric to evaluate the accuracy of the reconstructed sound field by means of the inverse method. The plots correspond to a cross-section of the domain (z = 1.5 m). The selected acoustic errors are:

Amplitude error:

$$E_{pa}(\mathbf{x}) = 20 \log_{10} \left(\frac{|\tilde{p}(\mathbf{x})|}{|p(\mathbf{x})|} \right), \tag{4.9}$$

Phase error:

$$E_{pp}(\mathbf{x}) = \angle \left(p(\mathbf{x}) \tilde{p}(\mathbf{x})^* \right), \tag{4.10}$$

where $\tilde{p}(\mathbf{x})$ is the reconstructed pressure, $p(\mathbf{x})$ is the target pressure and $(\cdot)^*$ indicates the complex conjugate operator. The amplitude error provides an insight about whether the reconstructed acoustic field is louder or quieter compared to the target one. The phase error allows to determine the phase differences between the reconstructed and target acoustic fields. Nevertheless, because the amplitude of the plane and spherical waves are different, a normalized error was implemented, which allows these two types of sound fields to be compared. The normalized error is defined as

$$E_{npa}(\mathbf{x}) = 10 \log_{10} \left(\frac{|p(\mathbf{x}) - \tilde{p}(\mathbf{x})|^2}{|p_t(\mathbf{x})|^2} \right). \tag{4.11}$$

An example of this type of figure is presented in Figure 4.6.

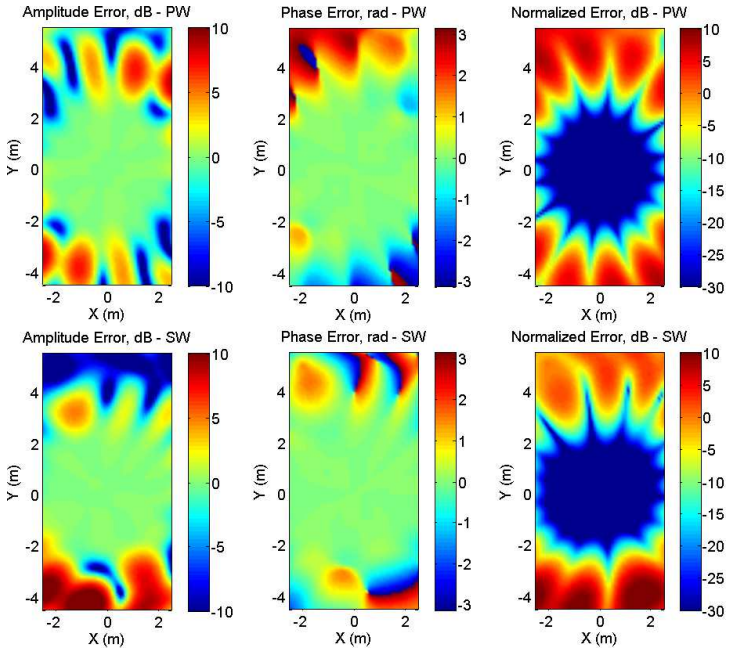


FIGURE 4.6: Example of figure type 3

Figures 4.7 to 4.23 show the three set of figures for multiple positions of the array and for two frequencies (63 Hz and 250 Hz). Three microphone array positions have been selected for the analysis being the central point of the arrays at (2.5, 1.5, 1.5), (2.5, 4.5, 1.5) and (2.5, 5.5, 1.5) based on the origin established in Figure 4.3. The direction of propagation for the plane wave is $\theta = 90°$ and $\phi = 45°$. The point source is located at (0.5, 9.5, 1.5). The black squares presented in figures type 1 represents the position of the microphone array. Regarding figures type 3, the white contour defines the region within which the normalized error is smaller than -20 dB and the black circle corresponds to the area of accurate reconstruction based on the relation proposed by Kennedy et al. [142]

$$L = \left(\left\lceil e\pi \frac{R}{\lambda} \right\rceil + 1 \right)^2, \tag{4.12}$$

in which L is the number of plane waves, e is euler's number, λ is the wavelength and R is the radius of a sphere within which the reconstruction is supposed accurate.

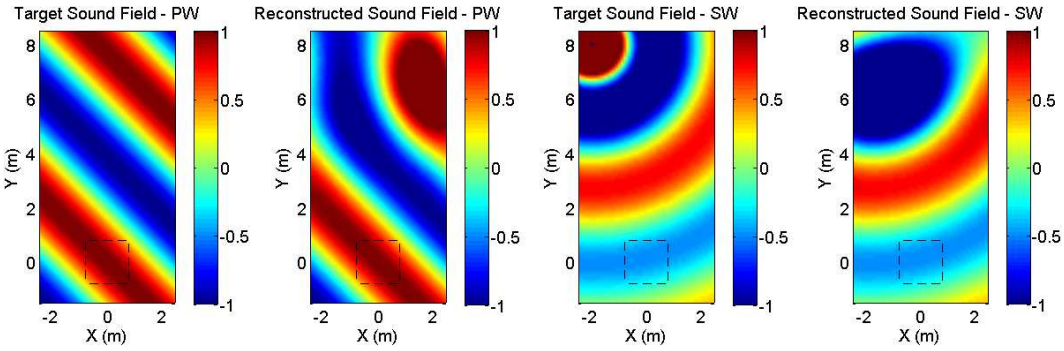


FIGURE 4.7: Target and reconstructed field at 63 Hz, array position 1, PWE

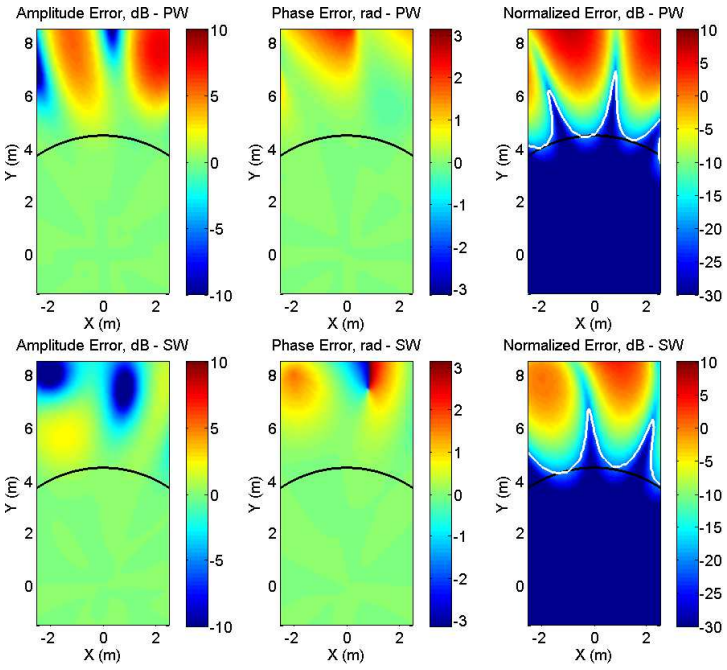


FIGURE 4.8: Errors at 63 Hz, array position 1, PWE

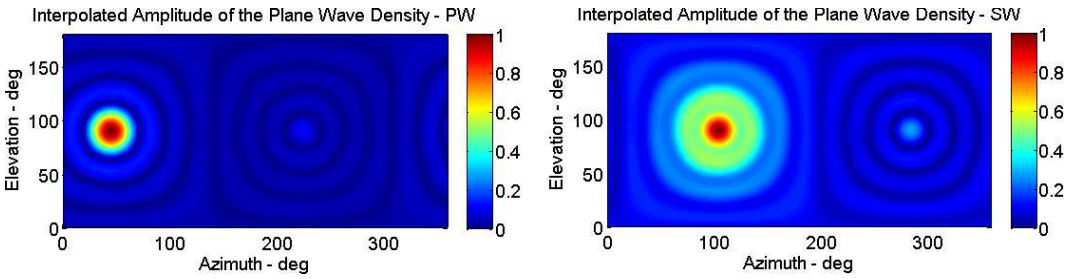


FIGURE 4.9: Normalized gain PWE at 63 Hz, array position 1, q_{total}(PW)=1.34, q_{total}(SW)=0.44

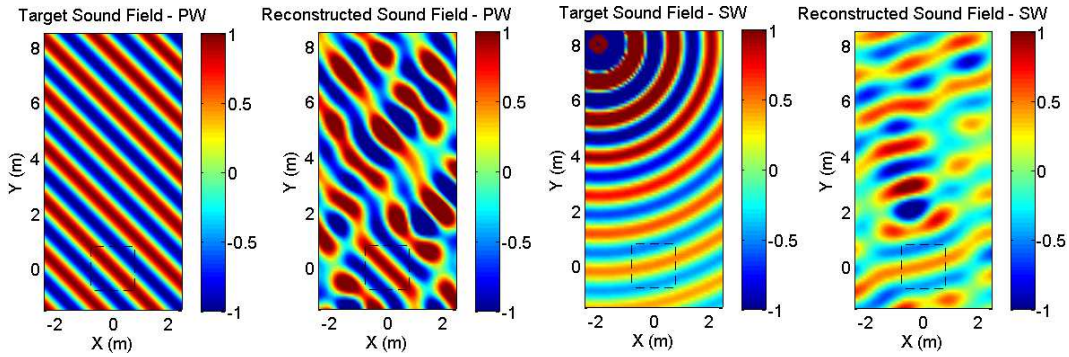


FIGURE 4.10: Target and reconstructed field at 250 Hz, array position 1, PWE

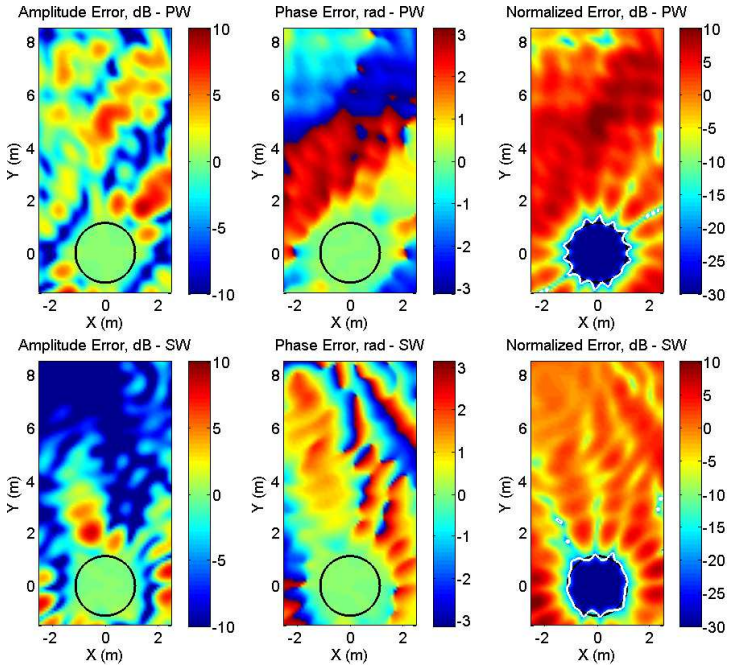


FIGURE 4.11: Errors at 250 Hz, array position 1, PWE

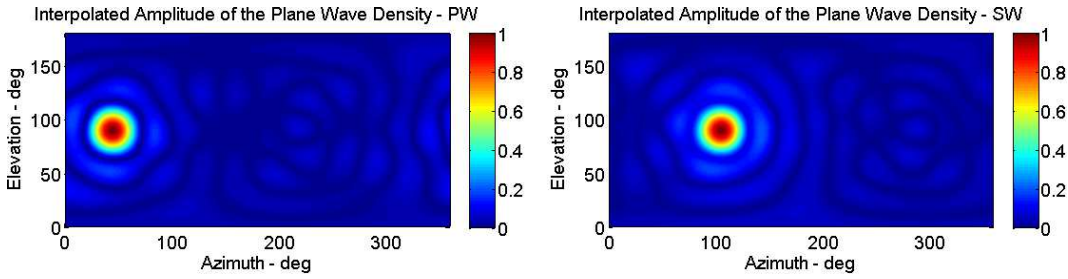


FIGURE 4.12: Normalized gain PWE at 250 Hz, array position 1, q_{total}(PW)=1.04, q_{total}(SW)=0.27

These figures indicate that the inverse method is a suitable approach to synthesize acoustic fields. Figures 4.7 and 4.10 (type 1) reveal that as expected, the reconstruction is accurately performed around the virtual microphone array. In terms of the amplitude of the plane wave density (figures type 2), Figures 4.9 and 4.12 illustrate that the energy is directionally concentrated at the plane wave positions closer to the direction of propagation of the plane wave. For the spherical wave case, similar results are obtained based on the location of the point source. Nevertheless, at 63 Hz the energy is less directionally concentrated in the plane wave density, which is a consequence of the wave curvature that is more predominant compared to 250 Hz for this specific source-array distance. Finally, the acoustic errors presented in Figures 4.8 and 4.11 (type 3) confirm that the area of accurate reconstruction depends on the frequency of the acoustic field, being broader at low frequencies. A good agreement between the area of accurate reconstruction and the radius predicted by equation (4.12) was also found for both acoustic fields for this specific case.

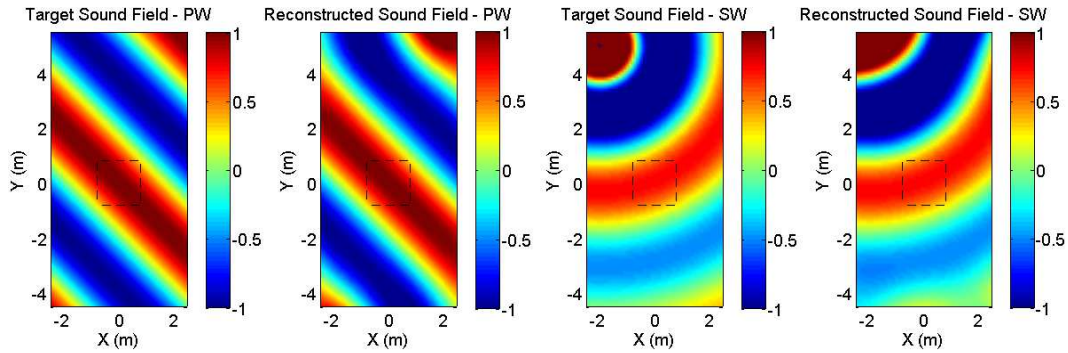


FIGURE 4.13: Target and reconstructed field at 63 Hz, array position 2, PWE

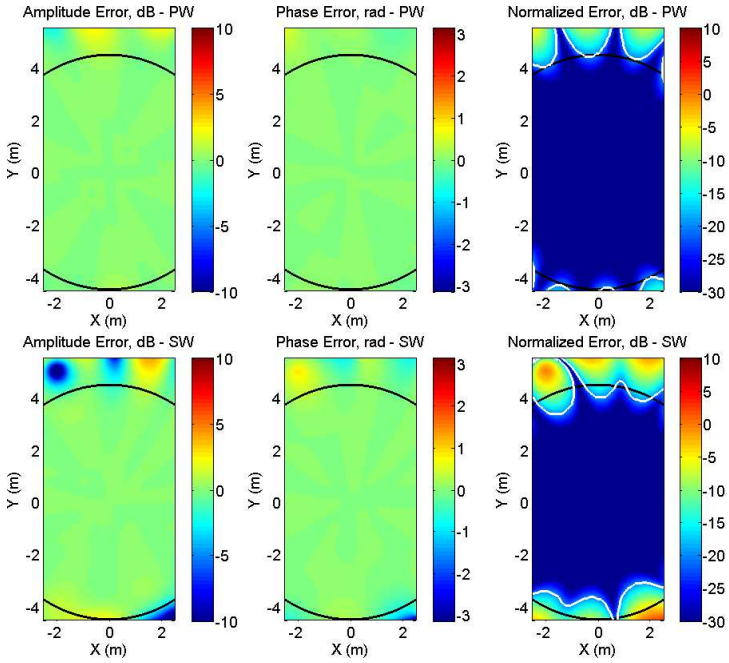


FIGURE 4.14: Errors at 63 Hz, array position 2, PWE

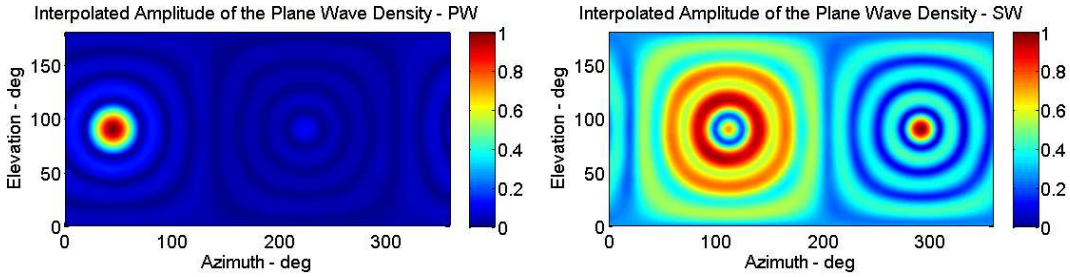


FIGURE 4.15: Normalized gain PWE at 63 Hz, array position 2, q_{total}(PW)=1.34, q_{total}(SW)=1.74

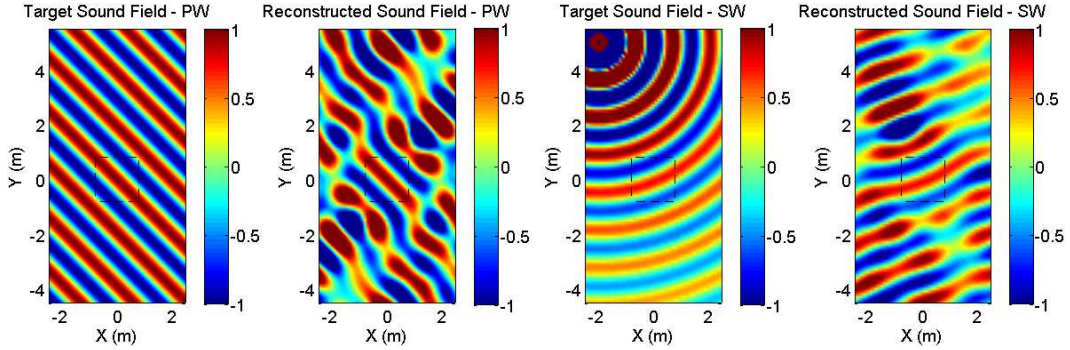


FIGURE 4.16: Target and reconstructed field at 250 Hz, array position 2, PWE

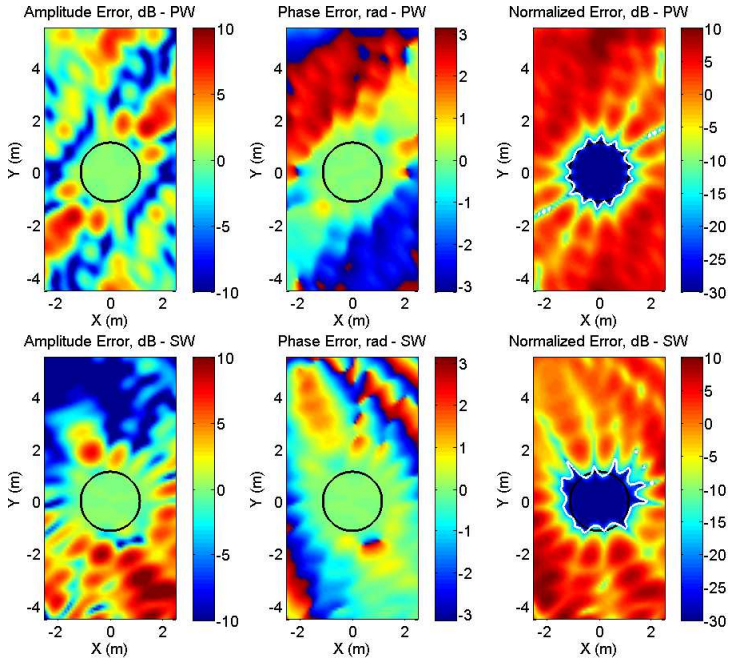


FIGURE 4.17: Errors at 250 Hz, array position 2, PWE

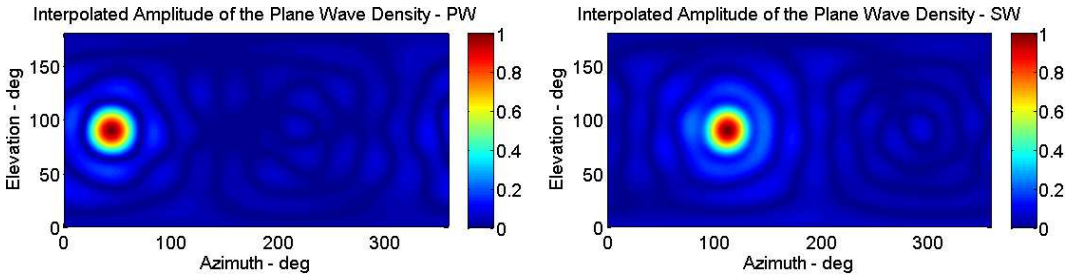


FIGURE 4.18: Normalized gain PWE at 250 Hz, array position 2, q_{total}(PW)=1.04, q_{total}(SW)=0.61

Results for the array position 2 suggest that the acoustic near field produced by the point source start to constrain the sound field reconstruction at 63 Hz. The energy of the plane wave density has been distributed around the plane waves for the spherical acoustic field at 63 Hz, which is expected as the curvature of the spherical wave becomes more prevalent. Although the radius predicted by equation (4.12) does not reach the position of the acoustic source, its proximity decreases the area where the normalized error is smaller than -20 dB and there is not an agreement between the two parameters in contrast with the results obtained for the array position 1. Regarding plane wave propagation, the results are consistent with the array position

1 due to the location of the array does not have any effect in the inversion for this type of propagation.

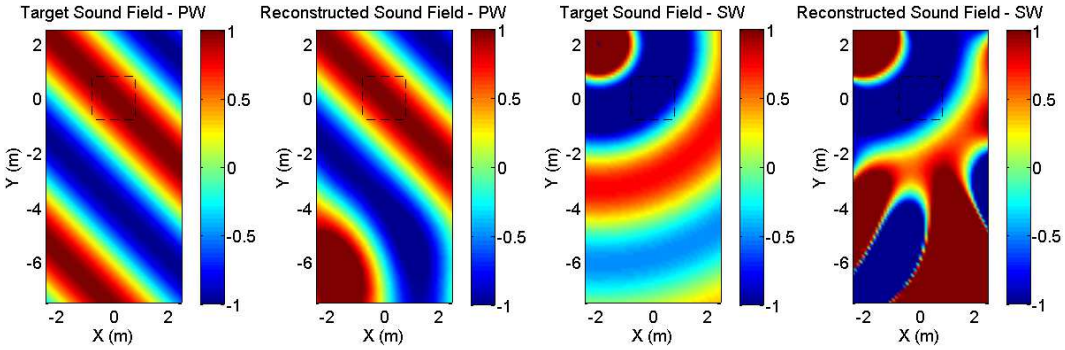


FIGURE 4.19: Target and reconstructed field at 63 Hz, array position 3, PWE

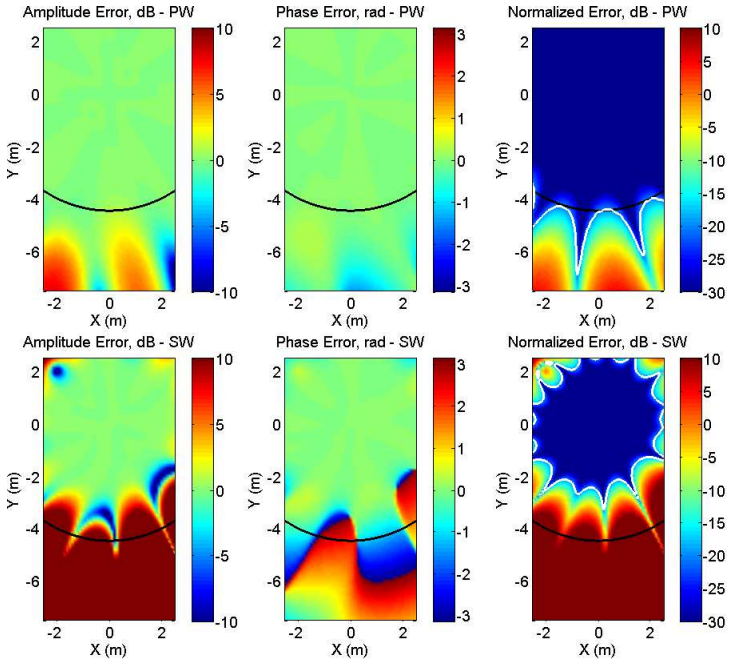


FIGURE 4.20: Errors at 63 Hz, array position 3, PWE

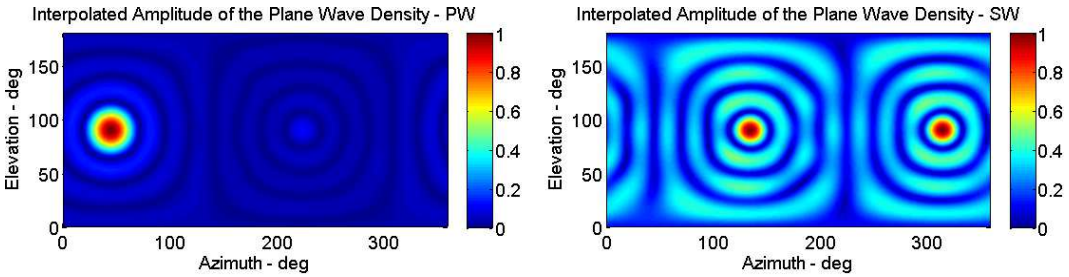


FIGURE 4.21: Normalized gain PWE at 63 Hz, array position 3, q_{total}(PW)=1.34, q_{total}(SW)=3.03×10^3

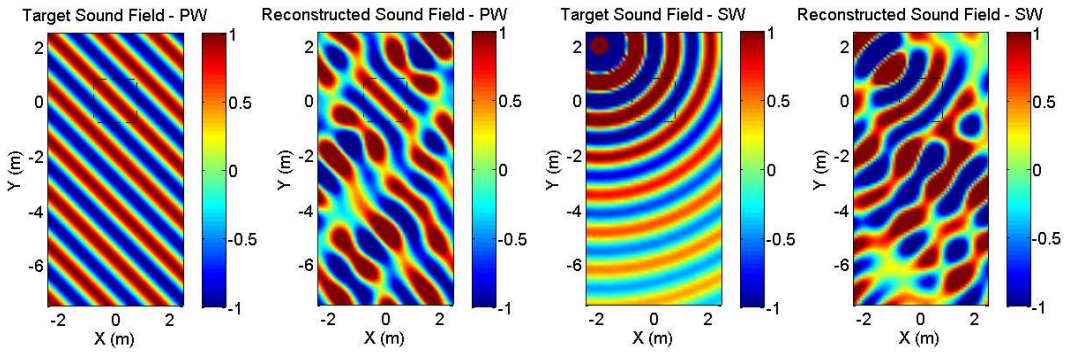


FIGURE 4.22: Target and reconstructed field at 250 Hz, array position 3, PWE)

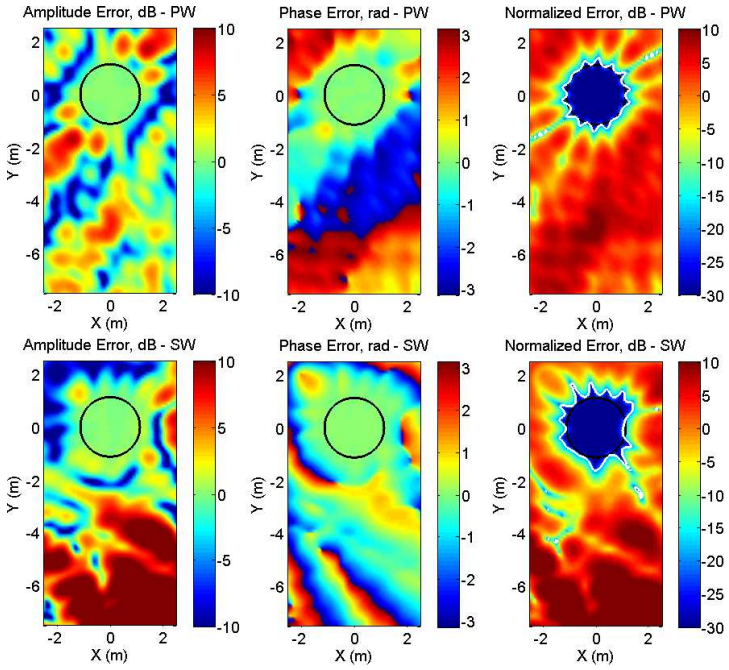


FIGURE 4.23: Errors at 250 Hz, array position 3, PWE

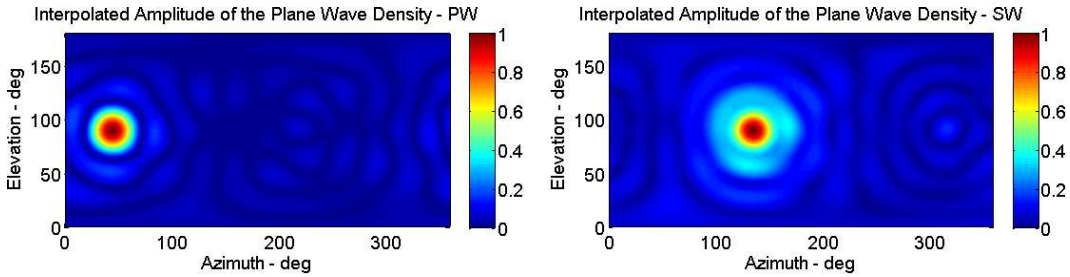


FIGURE 4.24: Normalized gain PWE at 250 Hz, array position 3, q_{total}(PW)=1.04, q_{total}(SW)=2.23

The results are consistent which the outcomes found for the array position 2. Nevertheless, an interesting result was found based on the acoustic errors at 63 Hz for the spherical acoustic field (see figure 4.20). The amplitude error indicates that outside the region where the reconstruction is accurate the sound field is much louder than the target field. This suggests that a significant part of energy of the plane wave density is used to reconstruct the sound field outside the region of interest. Regarding the normalized error, the area of accurate reconstruction is compounded up to the extent where the point source is located.

In general, the results from Figures 4.7 to 4.24 show that the implementation of an inverse method is a suitable methodology to estimate the amplitude of the plane waves. The area of accurate reconstruction produced by the plane wave expansion depends on frequency and on position. Furthermore, an analysis of the errors suggests a high congruence between the area with a low normalized error (smaller than -20 dB) and the radius of validity computed from equation (4.12) for the plane wave propagation case. For spherical wave propagation, this relation holds as long as the plane wave assumption remains valid.

Effects of the sound field representation

The reconstruction of sound fields based on plane waves reveals constraints when a spherical wave is intended to be reconstructed. In this case, the region of accurate reconstruction shrinks and the relation given by equation (4.12) does not hold if the far field assumption is not fulfilled, namely

$$r < R(L, \omega) \tag{4.13}$$

where r is the distance to the point source and $R(L, \omega)$ is the radius predicted by equation (4.12). Moreover, the distribution of the energy in the plane wave density (figures type 2) indicates that for the case of a plane wave the energy is directionally concentrated on the plane waves located closest to the direction of propagation of the plane wave to be reconstructed. In contrast, a sparse distribution was found at 63 Hz when the virtual microphone array is positioned closer to the source (5 m and 2 m) for the spherical wave propagation. This result is consistent with the Weyl's integral, which describes a point source as an infinite superposition of propagating and evanescent plane waves [14].

4.1.1.3 Regularization

A well established technique to improve the stability of the solutions of inverse methods is the use of regularization in the inversion of the propagation matrix $\mathbf{H}(\omega)$ [139] [143] [144]. Tikhonov regularization is one approach used for this purpose, it is based on the concept of changing the cost function $J(\omega)$ by the inclusion of an additional term [139], that is

$$J(\omega) = \mathbf{e}^H(\omega)\mathbf{e}(\omega) + \beta(\omega)\mathbf{q}^H(\omega)\mathbf{q}(\omega), \tag{4.14}$$

where $\beta(\omega)$ is the regularization parameter, whose value is usually taken between $\|\mathbf{H}(\omega)\|^2/1000$ and $\|\mathbf{H}(\omega)\|^2/5000$ [13]. $\|\mathbf{H}(\omega)\|$ corresponds to the norm (the largest singular value) of the propagation matrix $\mathbf{H}(\omega)$. The minimization of the cost function $J(\omega)$ established in equation (4.14) is given by [139]

$$\mathbf{q}(\omega) = \left[\mathbf{H}^H(\omega)\mathbf{H}(\omega) + \beta(\omega)\mathbf{I}\right]^{-1}\mathbf{H}^H(\omega)\mathbf{p}(\omega). \tag{4.15}$$

Equation (4.14) indicates that the minimization of the cost function takes into account the sum of the squared errors between the reconstructed and target acoustic pressure and, in addition, the sum of the squared norm of plane wave amplitude vector. Figures 4.25 to 4.38 illustrate the results of simulations according to the conditions previously described but with Tikhnonov regularization. The value of β was selected based on the range recommended by Støfringsdal and Svensson [13] where a similar problem was addressed and by the evaluation of the complex amplitude of

the plane waves required to synthesize the sound field. The value of β is reported for each specific case and it was calculated as

$$\beta = \|\mathbf{H}(\omega)\|^2 \epsilon \tag{4.16}$$

in which $\|\cdot\|$ is the norm (the largest singular value) of the propagation matrix $\mathbf{H}(\omega)$ and ϵ is an arbitrary constant.

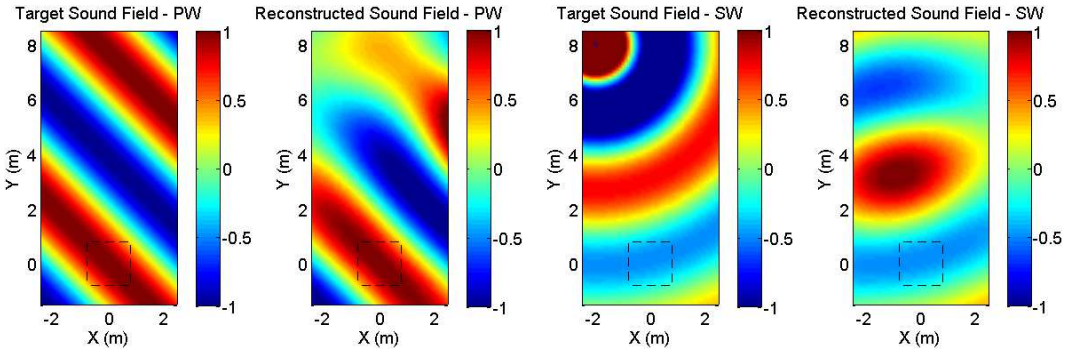


FIGURE 4.25: Target and reconstructed field at 63 Hz, regularized case $\beta = 0.033$, array position 1, PWE

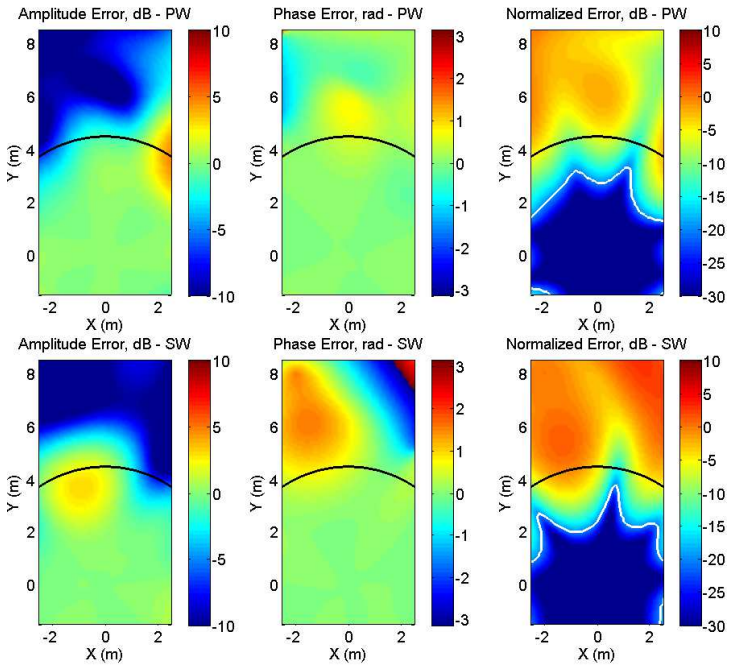


FIGURE 4.26: Errors at 63 Hz, regularized case $\beta = 0.033$, array position 1, PWE

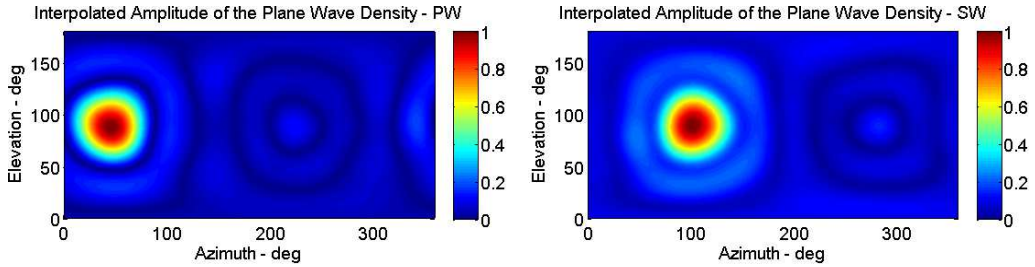


FIGURE 4.27: Normalized gain PWE at 63 Hz, regularized case $\beta = 0.033$, array position 1, q_{total}(PW)=0.34, q_{total}(SW)=0.08

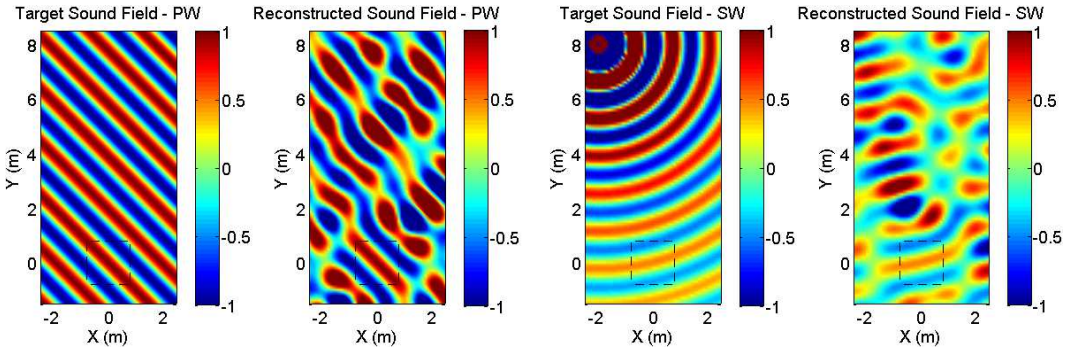


FIGURE 4.28: Target and reconstructed field at 250 Hz, regularized case $\beta = 0.003$, array position 1, PWE

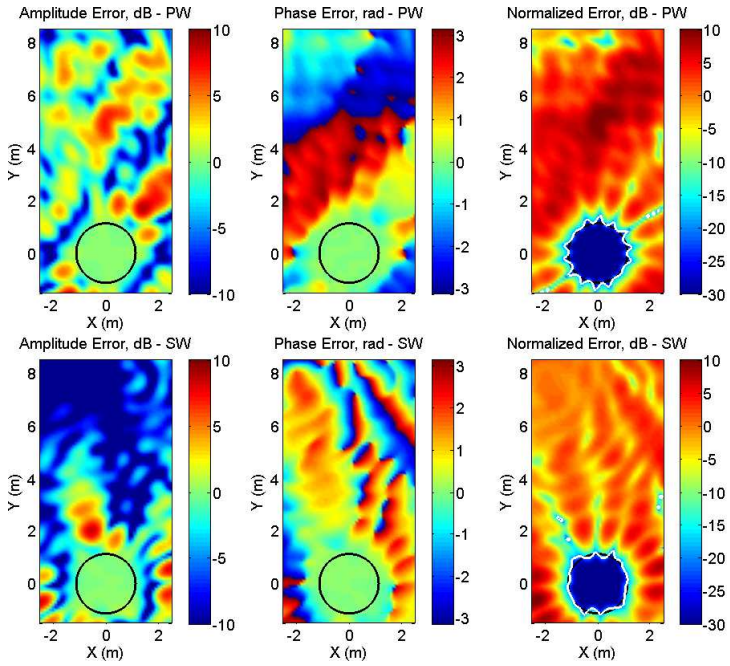


FIGURE 4.29: Errors at 250 Hz, regularized case $\beta = 0.003$, array position 1, PWE

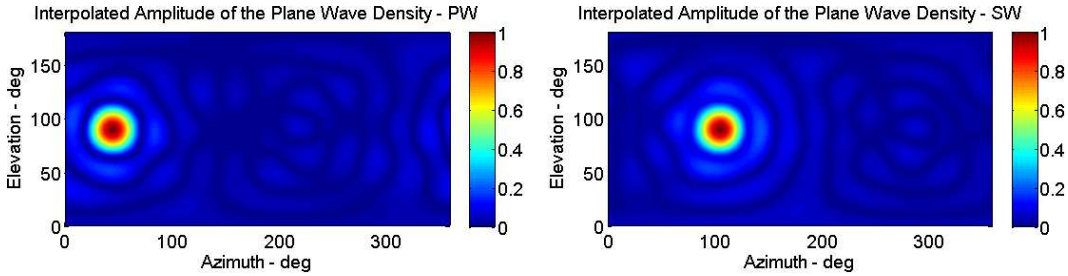


FIGURE 4.30: Normalized gain PWE at 250 Hz, regularized case $\beta = 0.003$, array position 1, q_{total}(PW)=1.04, q_{total}(SW)=0.27

An analysis of the normalized error (see Figure 4.26) reveals that the implementation of regularization reduces the area of accurate reconstruction at 63 Hz compared to the non-regularized case (Figure 4.8). In contrast, the regularization has not a significant effect in the sound field reconstruction at 250 Hz. This can be explained because the condition number at this frequency is very low (45.9), which indicates that the regularization has no relevance in the inversion of the matrix $\mathbf{H}(\omega)$. These findings are consistent for both types of acoustic field, the plane and spherical propagation, respectively. Nevertheless, according to Figures 4.25 and 4.26 the amplitude of the acoustic fields outside the radius predicted by equation (4.12) tends to be smaller for the regularized case at 63 Hz. This particular result is discussed later under the perspective of a translated acoustic field. Finally, Figure 4.27 shows that the energy distribution of the plane wave density is more directionally concentrated at 63 Hz compared to the results obtained for the non-regularized case (Figure 4.9). This is expected as the area of accurate reconstruction has been reduced, which leads to the curvature of the spherical acoustic field being less prevalent in that area. In other words, the plane and spherical acoustic fields are more similar if the area where they are compared is smaller (this relation being dependent on the distance to the point source and the frequency of the acoustic fields).

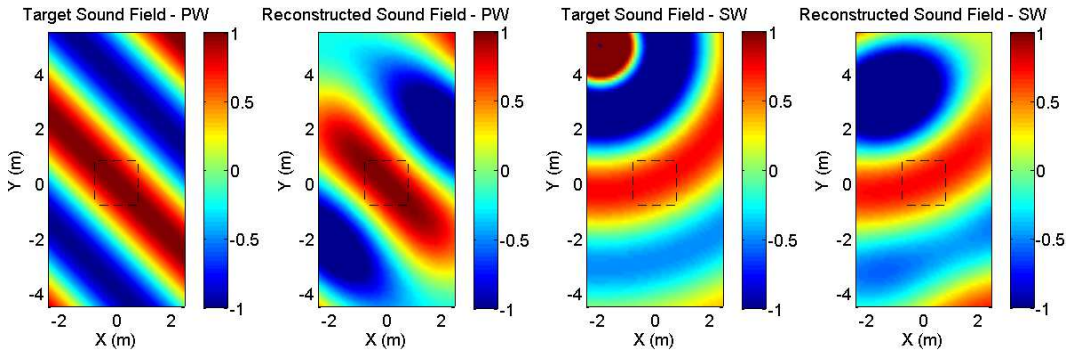


FIGURE 4.31: Target and reconstructed field at 63 Hz, regularized case $\beta = 0.33$, array position 2, PWE

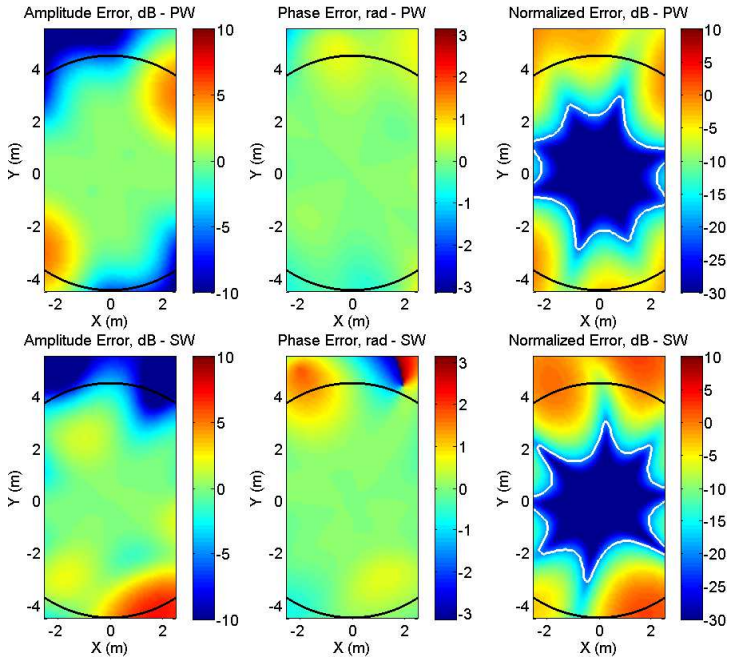


FIGURE 4.32: Errors at 63 Hz, regularized case $\beta = 0.33$, array position 2, PWE

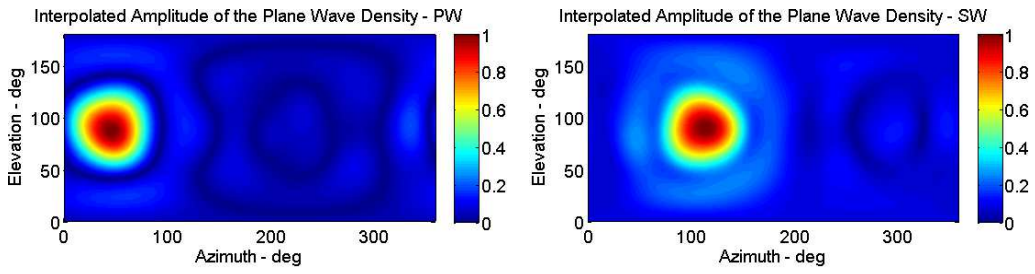


FIGURE 4.33: Normalized gain PWE at 63 Hz, regularized case $\beta = 0.33$, array position 2, q_{total}(PW)=0.25, q_{total}(SW)=0.15

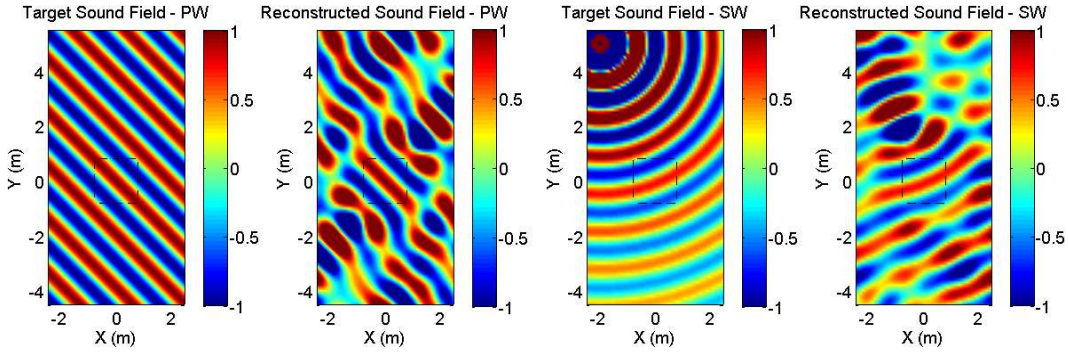


FIGURE 4.34: Target and reconstructed field at 250 Hz, regularized case $\beta = 0.03$, array position 2, PWE

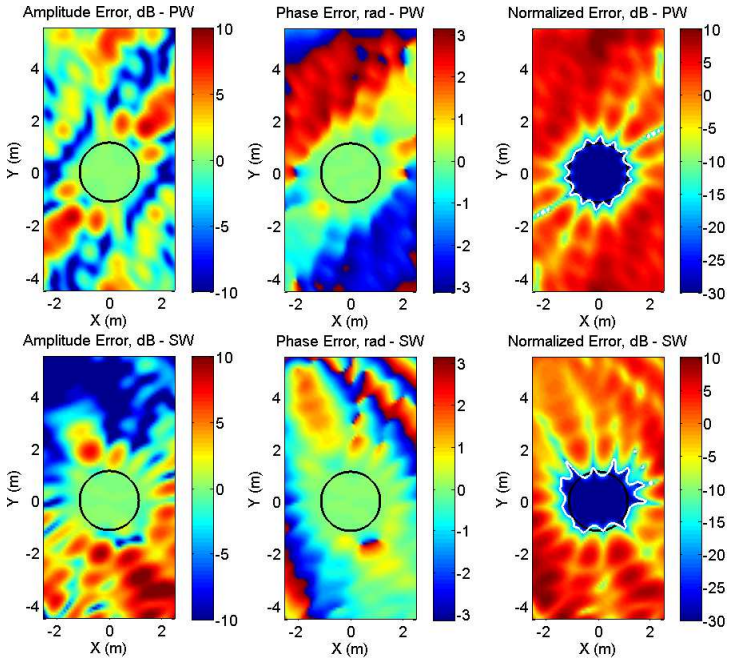


FIGURE 4.35: Errors at 250 Hz, regularized case $\beta = 0.03$, array position 3, PWE

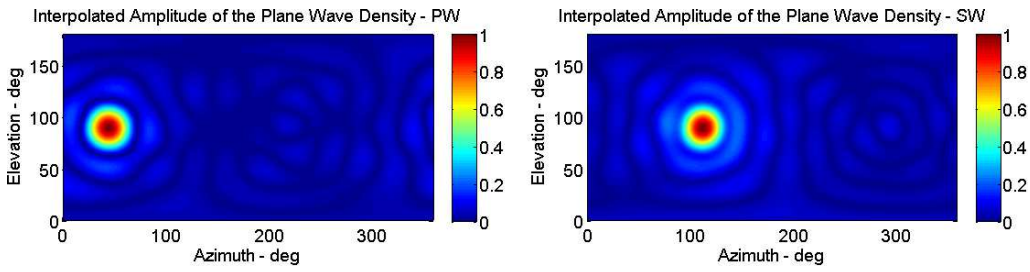


FIGURE 4.36: Normalized gain PWE at 250 Hz, regularized case $\beta = 0.03$, array position 2, q_{total}(PW)=1.04, q_{total}(SW)=0.61

The results are consistent with the outcomes found for the regularized case at the array position 1. The area within which the normalized error is smaller than -20 dB is reduced and the energy distribution of the plane wave density is more directionally concentrated compared to the non-regularized case. Although the array is closer to the source, there is no effect of the regularization at 250 Hz.

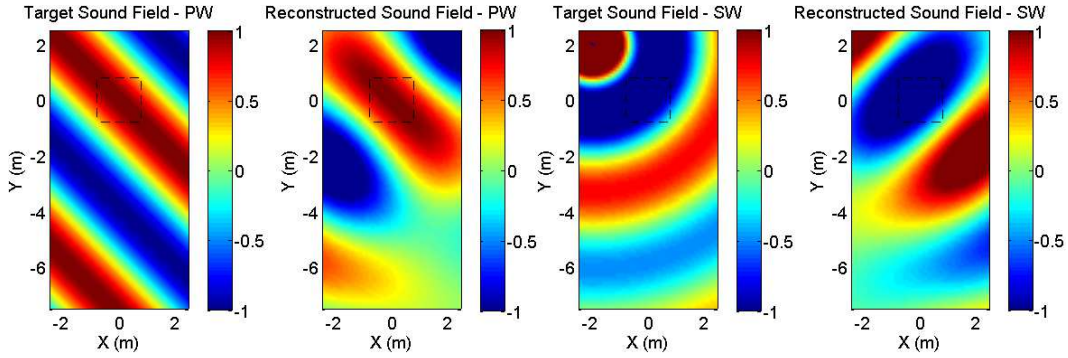


FIGURE 4.37: Target and reconstructed field at 63 Hz, regularized case $\beta = 32.74$, array position 3, PWE

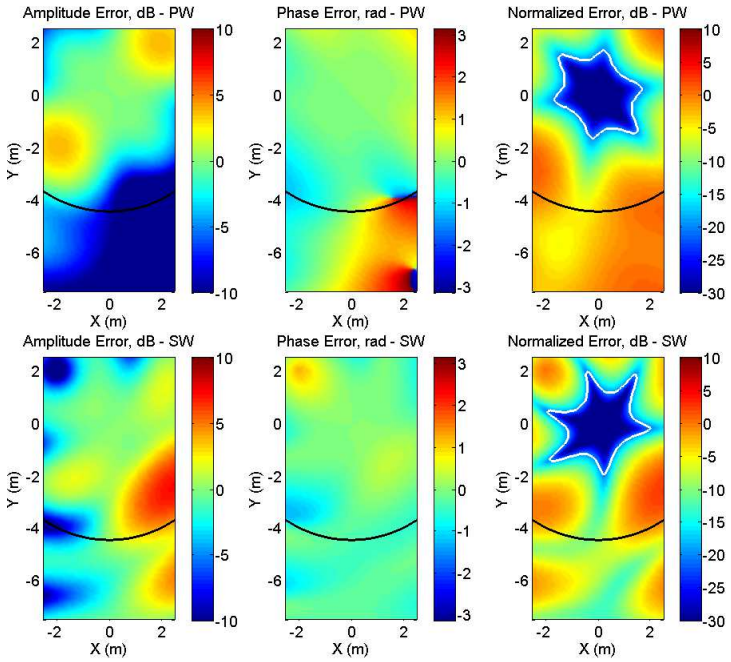


FIGURE 4.38: Errors at 63 Hz, regularized case $\beta = 32.74$, array position 3, PWE

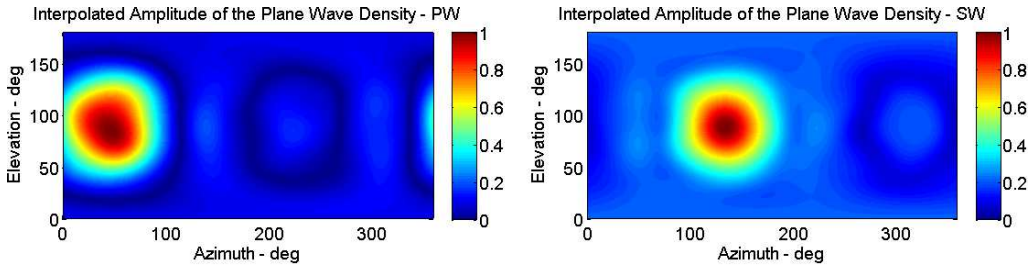


FIGURE 4.39: Normalized gain PWE at 63 Hz, regularized case $\beta = 32.74$, array position 3, q_{total}(PW)=0.12, q_{total}(SW)=0.31

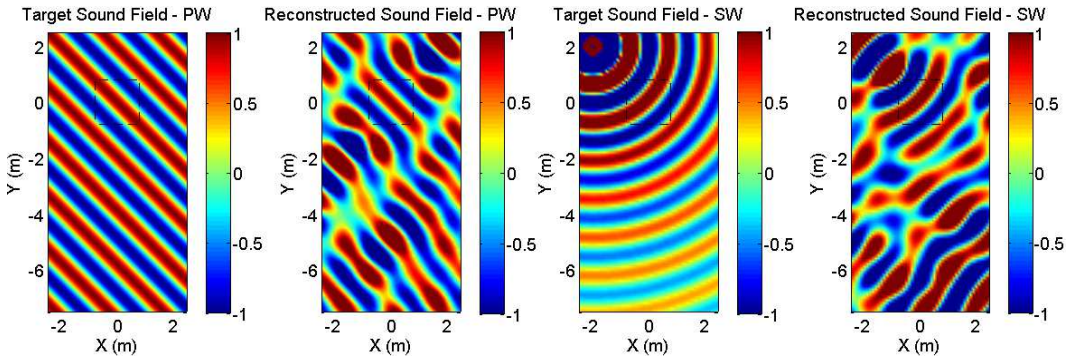


FIGURE 4.40: Target and reconstructed field at 250 Hz, regularized case $\beta = 2.81$, array position 3, PWE

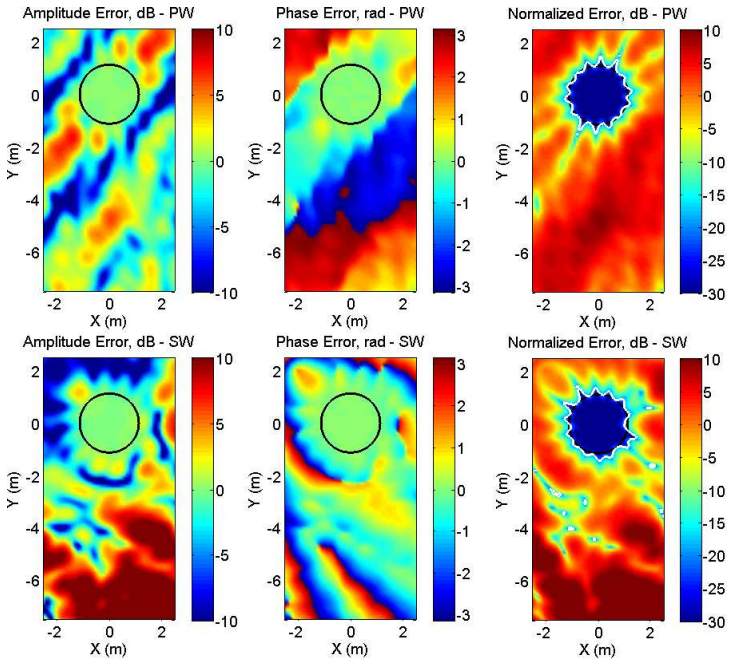


FIGURE 4.41: Errors at 250 Hz, regularized case $\beta = 2.81$, array position 3, PWE

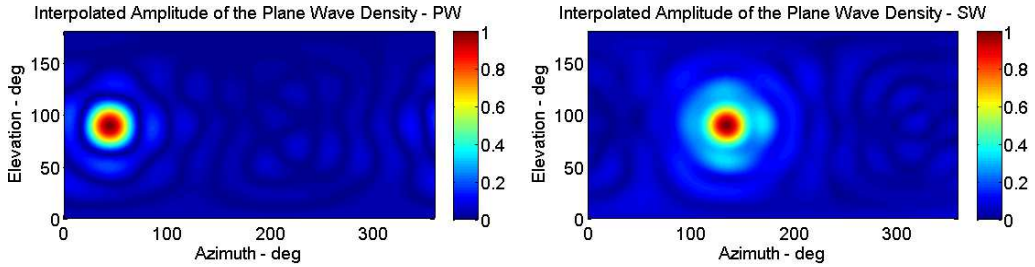


FIGURE 4.42: Normalized gain PWE at 250 Hz, regularized case $\beta = 2.81$, array position 3, q_{total}(PW)=0.84, q_{total}(SW)=1.75

The comparison of the acoustic fields for the non-regularized and regularized case (Figures 4.19 and 4.37) indicates important differences in the sound field reconstruction for the array position 3 at 63 Hz. For the spherical field, this array is located at the closest distance to the point source compared to the other 2 array positions. This condition implies a more relevant influence of the near field produced by the acoustic source, particularly at 63 Hz. At this array position, while the synthesized acoustic field at 63 Hz without regularization tends to preserve the curvature of the spherical wave propagation for the spherical field, the opposite happens with regularization in which the wavefronts are more similar to a plane wave propagation. However, outside the radius predicted by equation (4.12) a significant amount of energy is required for the non-regularized approach compared to the regularized one. This is confirmed by the analysis of the acoustic errors for both cases (Figures 4.20 and 4.38). An interesting result was found for the plane wave propagation at 63 Hz. According to Figures 4.21 and 4.39, the energy distribution of the plane wave density is less concentrated with regularization than the non-regularized case because the value of β is much higher compared to the other two array positions. This clearly shows the significance of β in the inversion of the matrix $\mathbf{H}(\omega)$.

The results for the 3 array positions indicate that Tikhonov regularization leads the energy distribution of the plane wave density to be more directionally concentrated for the spherical field. However, this is done at the expense of reducing the area where the normalized error is smaller than -20 dB. Regarding the acoustic fields, the inclusion of regularization affects both the plane and spherical fields, but mainly at low frequencies. The reason for this outcome is associated with the small value of the condition number (45.9) at 250 Hz, which suggests that the regularization does not have a significant influence in the inversion of the propagation matrix. A further analysis has been carried out to evaluate the effect of including regularization in the

inverse problem. Figures 4.43 to 4.45 illustrate the absolute value squared of the elements of vector $\mathbf{q}(\omega)$. The non-regularized and regularized cases are compared for the three different positions previously considered based on the spherical acoustic field. Only 63 Hz is illustrated because this is the frequency range where the regularization has a significant effect.

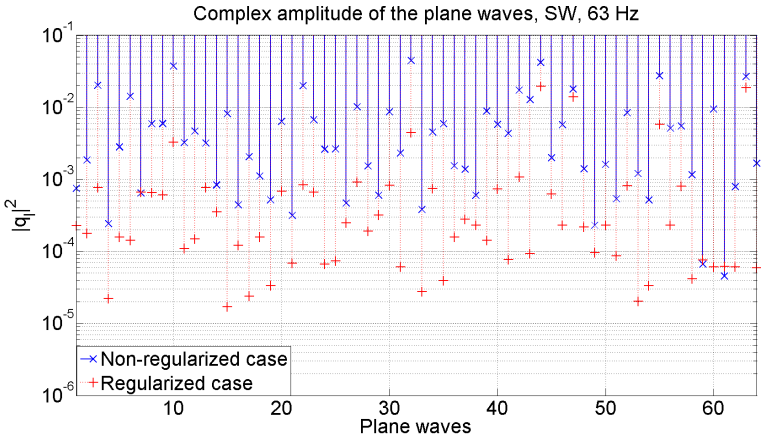


FIGURE 4.43: Complex amplitude of the PWE, SW, 63 Hz, position 1

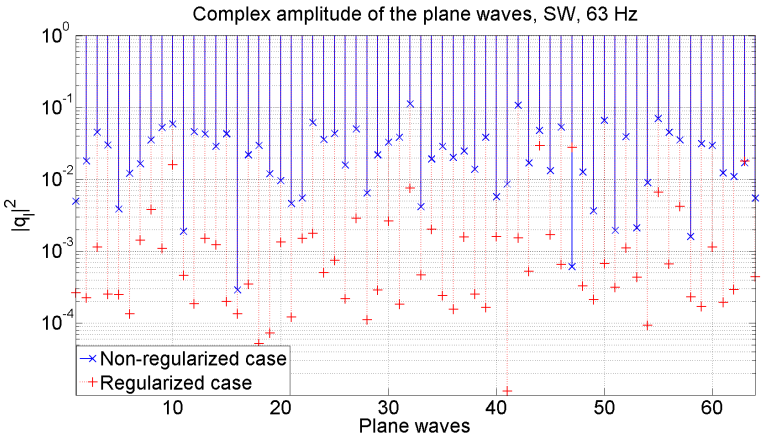


FIGURE 4.44: Complex amplitude of the PWE, SW, 63 Hz, position 2

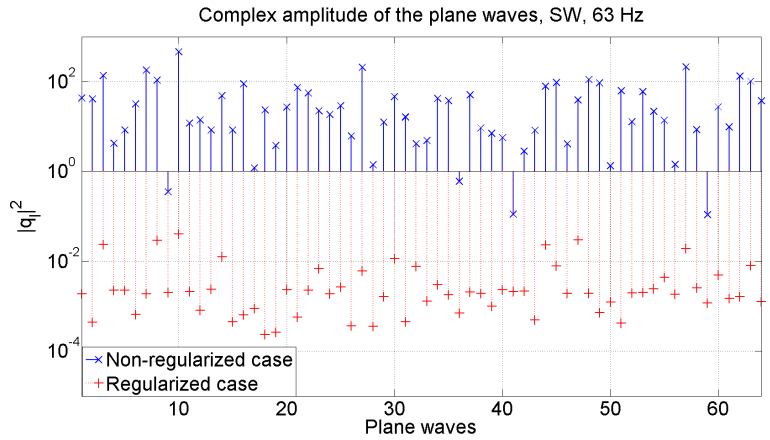


FIGURE 4.45: Complex amplitude of the PWE, SW, 63 Hz, position 3

Figures 4.43, 4.44 and 4.45 indicate that the application of regularization allows for a reduction of the energy required to reconstruct the sound field. The decrease of the energy depends on the amount of regularization, which in turn, can be determined by the distance of the virtual microphone array to the point source. In general, the regularization produces a synthesis of the sound field that uses less energy. Nevertheless, as consequence of that, the area of accurate reconstruction is reduced. The implementation of regularization in the inversion is convenient for two main reasons: firstly, it reduces the energy outside the radius given by equation (4.12). This is important because if a translation of a sound field is applied, it is not desirable to have zones with higher energy. Secondly, the regularization reduces the effect of numerical errors given by the FE simulations.

An additional analysis in terms of spherical harmonics was performed to evaluate the influence of regularization. The expansion of equation (4.1) in terms of spherical harmonics using the Jacobi-Anger relation leads to

$$p(\mathbf{x},\omega) = \int_{\Omega} \sum_{n=0}^{\infty} \sum_{m=-n}^{n} j_n(kr_x) Y_n^m(\theta_x,\phi_x,\omega) j^n 4\pi Y_n^m(\theta_y,\phi_y,\omega)^* q(\theta_y,\phi_y,\omega) d\Omega \tag{4.17}$$

which in turn, after the discretization of the domain in a finite number of plane waves is transformed into

$$p(\mathbf{x},\omega) = \sum_{l=1}^{L}\sum_{n=0}^{\infty}\sum_{m=-n}^{n} j_n(kr_x)Y_n^m(\theta_x,\phi_x,\omega)j^n 4\pi Y_n^m(\theta_l,\phi_l,\omega)^* q(\theta_l,\phi_l,\omega)\Delta\Omega_l \tag{4.18}$$

where $\Delta\Omega_l$ is the area attributed to each direction $\widehat{\mathbf{y}}_l$ as result of the discretization of the integral representation. The complex spherical harmonic coefficients $\widetilde{A}_{nm}(\omega)$ are defined based on the general expansion of the sound field by spherical harmonics (equation (2.55)) as

$$\widetilde{A}_{nm}(\omega) = \sum_{l=1}^{L} j^n 4\pi Y_n^m(\theta_l,\phi_l,\omega)^* q(\theta_l,\phi_l,\omega)\Delta\Omega_l \tag{4.19}$$

Figures 4.46 to 4.48 show the sum of the absolute value of the complex spherical harmonic coefficients $\tilde{A}_{nm}$ squared for each order n up to 7^{th} order. The non-regularized and regularized cases are compared for the three different array positions for the spherical acoustic field at 63 Hz.

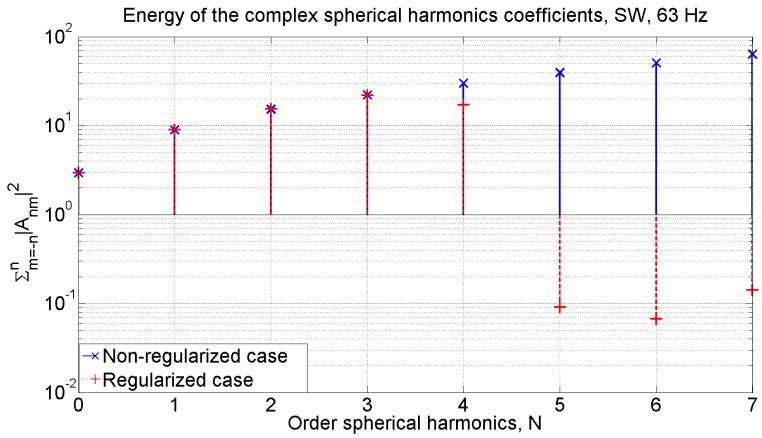


FIGURE 4.46: Energetic sum of complex spherical harmonic coefficients for each order n, SW, 63 Hz, position 1

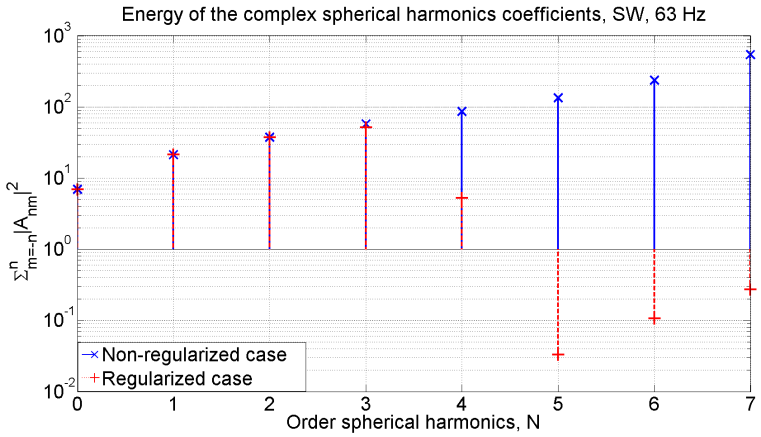


FIGURE 4.47: Energetic sum of complex spherical harmonic coefficients for each order n, SW, 63 Hz, position 2

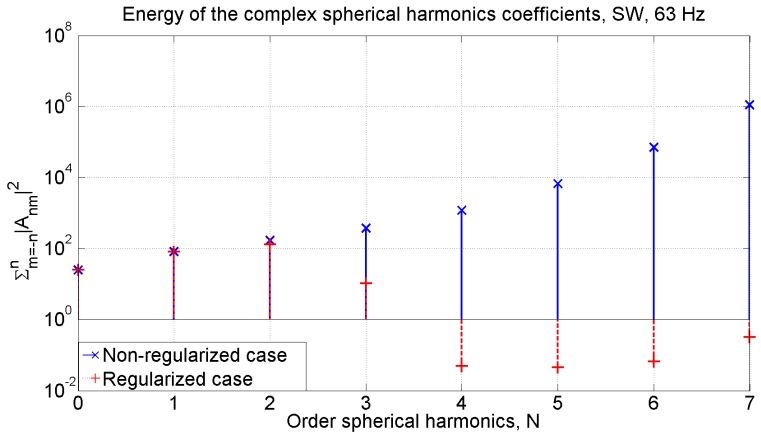


FIGURE 4.48: Energetic sum of complex spherical harmonic coefficients for each order n, SW, 63 Hz, position 3

The implementation of regularization has a significant effect on the high spherical harmonics. In this case, the energy of orders higher than 3^{rd} (2^{nd} for the closest position to the source) is drastically reduced. This condition may be explained by the fact that for a specific kr value, just a finite order of spherical harmonics is required to achieve an accurate sound field reconstruction ($N = kr$) [77]. If the area is reduced, which happens with the regularization, a lower order is necessary. In addition, the calculation of high order coefficients is the cause of the instability in the inverse problem, so by reducing the energy at high orders, regularization achieves a more stable solution.

Influence of β in the regularization of the inverse problem

Previous section demonstrated the suitability of Tikhonov regularization to optimize the results obtained from the inverse problem. In this section, an investigation about the influence of β as regularization parameter is addressed. This is performed by modifying the value of ϵ in equation (4.16). A sound field corresponding to a spherical wave propagation was analytically calculated in a domain of (10m x 10m x 3m). The frequency of the wave was 125 Hz. The acoustic source was located at (5, 9.5, 1.2) origin at the left-bottom corner. The sound field was sampled using a cubic virtual microphone array with linear dimensions of 1.6 m and a spatial resolution of 0.2 m (729 microphone positions). The central point of the microphone array was located at (5, 7, 1.2). Four values of ϵ, a = 10^{-6}, b = 10^{-5}, c = 10^{-4} and d = 10^{-3}, were used to calculate β and reconstruct the acoustic field. Figure 4.49 shows the synthesized sound field according to the values of ϵ. The non-regularized case is also illustrated for comparison purposes.

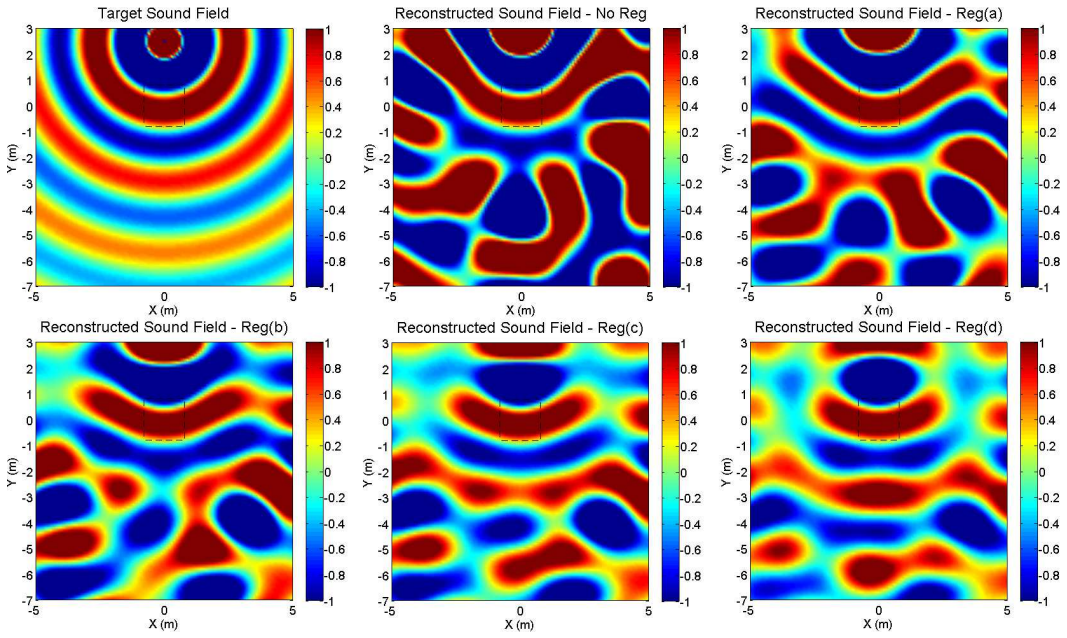


FIGURE 4.49: Target and reconstructed fields for different ϵ values

The results indicate that the influence of β is significant outside of the region of the microphone array. In this case, the amplitude of the acoustic pressure decreases as ϵ increases. An analysis based on acoustic errors is presented in Figures 4.50 to 4.52.

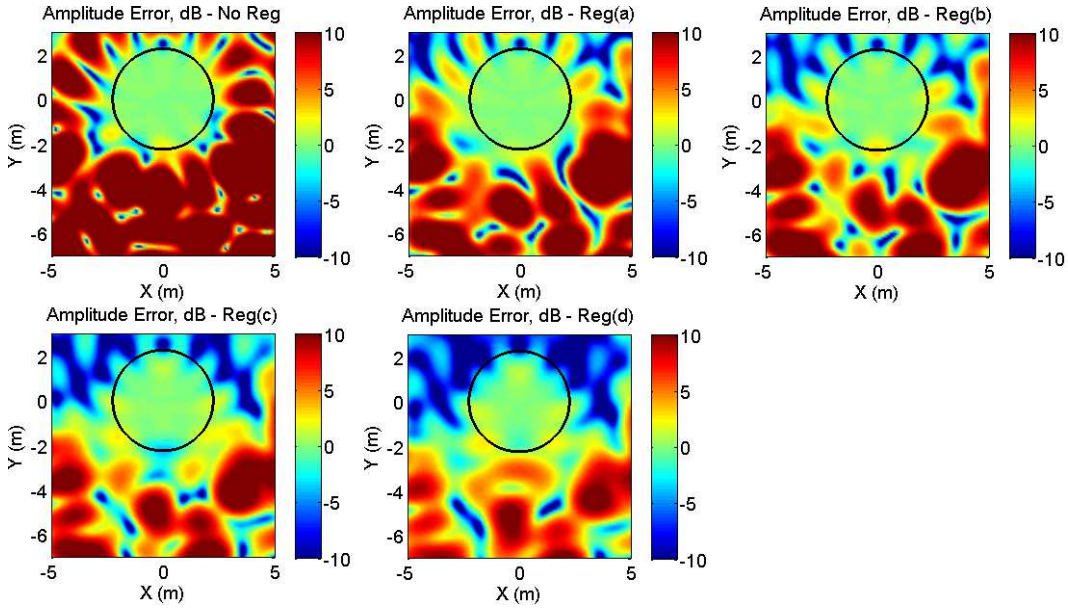


FIGURE 4.50: Amplitude error for different ϵ values

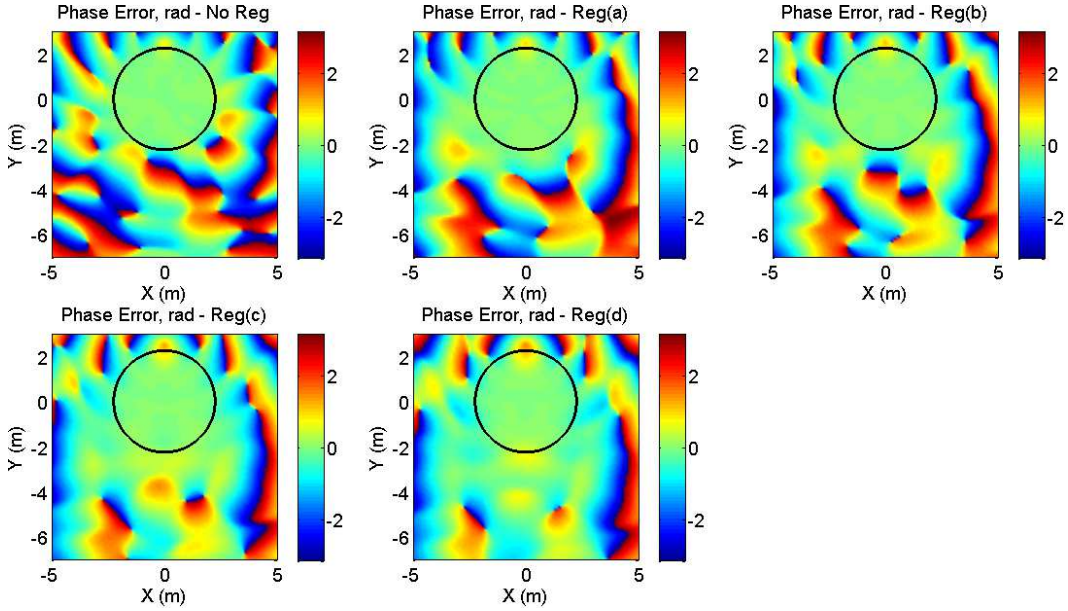


FIGURE 4.51: Phase error for different ϵ values

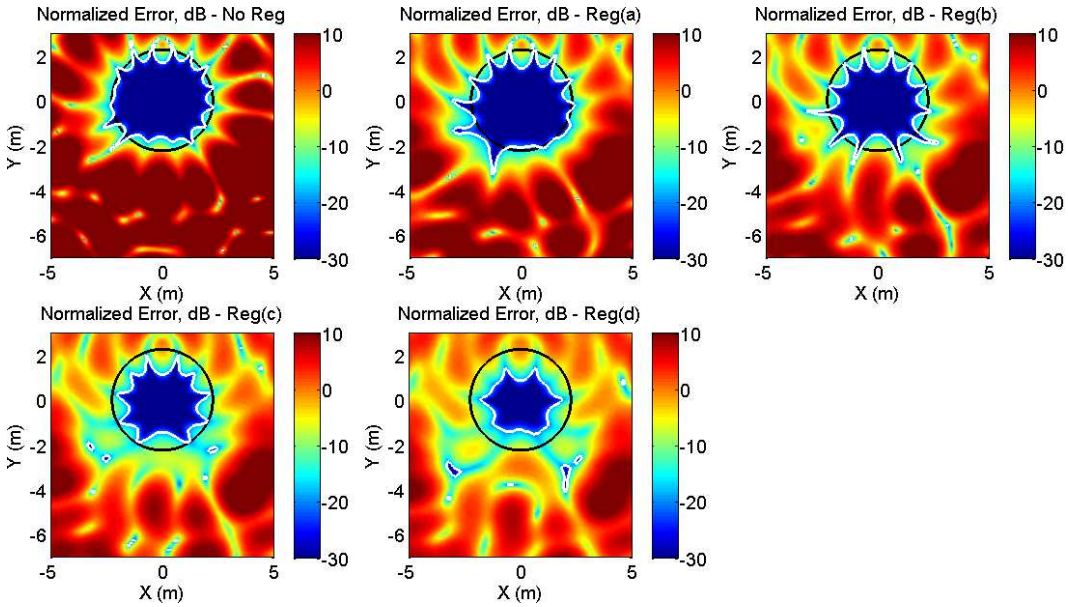


FIGURE 4.52: Normalized error for different ϵ values

The amplitude errors indicate that the reconstructed sound field is quieter when the value of ϵ is increased. This result is appropriate, as zones with high acoustic pressure are not desired from the translation point of view. Nevertheless, not significant changes were found between the cases (c) and (d), which suggest that in (d) more regularization than required is being applied. These outcomes are also consistent for the phase errors. Finally, the normalized errors reveal that the region in which the reconstruction of the sound field is accurate does not longer match with equation (4.12), being reduced as ϵ increases. However, the opposite happens outside of the radius predicted by equation (4.12) where the reconstruction is more accurate. A further analysis in terms of the total energy of the plane wave expansion is presented in Figure 4.53. Additional values of ϵ have been calculated to evaluate its impact in the plane wave density.

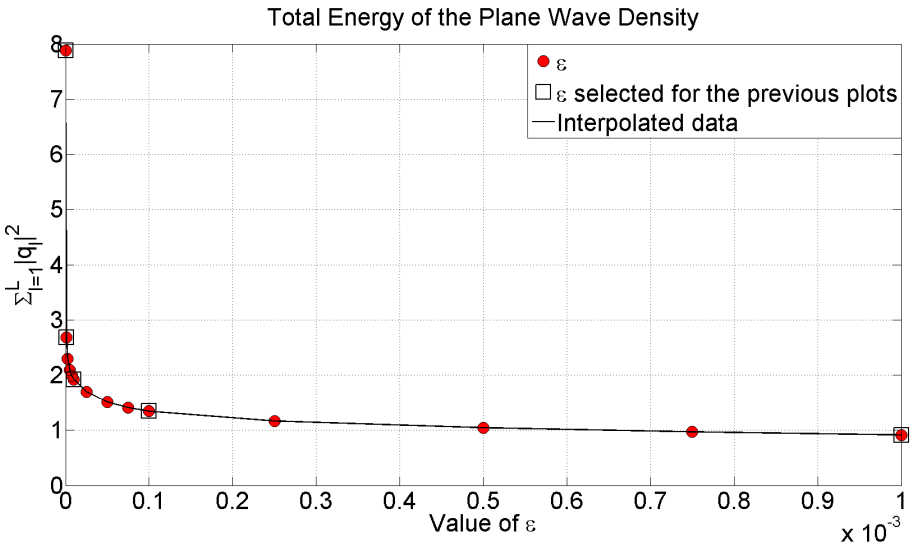


FIGURE 4.53: Total energy of the plane wave density for different ϵ values

Figure 4.53 shows that ϵ has a significant impact in the energy used for the plane wave expansion to reconstruct the acoustic field. The total energy is reduced following a logarithmic decay as higher values of ϵ are selected. This result is expected as ϵ determines the amount of regularization, which in turn, aims to minimize a cost function where the complex amplitude of the plane waves are considered. As corollary, the selection of ϵ mainly influences:

- The accuracy of the reconstructed acoustic field inside and outside of the region determined by (4.12).
- The total energy of the plane wave density.

In this thesis, ϵ is selected based on an iterative process in which the total energy of the plane wave density and the accuracy of the synthesized acoustic field are considered together. This generates a trade-off between the efficiency and accuracy of the plane wave expansion. Furthermore, the selection of ϵ is carried out aiming to reduce zones with high acoustic pressure outside of the region determined by equation (4.12), which is convenient for a translation operator.

4.1.2 Spherical harmonic expansion

Analogous to the PWE, the decomposition of the sound field using spherical harmonics (SHE) is another suitable representation which allows to describe the acoustic

field in an unbounded or bounded space. In this subsection, the same concept of an inverse approach is implemented to estimate a finite set of spherical harmonic coefficients $A_{nm}(\omega)$, which provide a local description of the sound field. The inverse problem is formulated based on Equation (2.55), which establishes the general representation of the sound pressure using spherical harmonics and spherical Bessel functions. The complex acoustic pressure at B virtual microphone positions are denoted using vector notation as

$$\mathbf{p} = [p_1, p_b, \ldots, p_B]^T, \tag{4.20}$$

where p_b is the acoustic pressure at the b-th virtual microphone. The complex coefficients A_{nm} used to decompose the sound field are represented by the vector

$$\mathbf{a} = \left[A_{\nu(1)\mu(1)}, A_{\nu(s)\mu(s)}, \ldots, A_{\nu(S)\mu(S)}\right]^T. \tag{4.21}$$

in which, $\nu(s) = \lceil\sqrt{s} - 1\rceil$ and $\mu(s) = s - 1 - \nu - \nu^2$. $\lceil\cdot\rceil$ corresponds to ceiling rounding. Finally, the transfer function $\mathbf{Y}$ that describes the relation between each complex coefficient A_{nm} and each virtual microphone can be arranged in matrix notation as:

$$\mathbf{Y} = \begin{vmatrix} y_{11} & \cdots & y_{1S} \\ \vdots & y_{bs}(\omega) & \vdots \\ y_{B1} & \cdots & y_{BS} \end{vmatrix}$$

in which $y_{bs} = j_\nu(kr_b)Y_{\nu\mu}(\theta_b, \phi_b)$. Consequently, the relationship between the complex coefficients and the virtual microphone signals is

$$\mathbf{p} = \mathbf{Y}\mathbf{a}. \tag{4.22}$$

The solution of equation (4.22) for $\mathbf{a}$ is given by [139]:

$$\mathbf{a}(\omega) = \mathbf{Y}^\dagger(\omega)\mathbf{p}(\omega), \tag{4.23}$$

in which $\mathbf{Y}^\dagger(\omega)$ is the Moore-Penrose pseudo-inverse of the propagation matrix $\mathbf{Y}(\omega)$ [139]. In the following, an analysis is considered of the relation between the size of the virtual array and the maximum order of spherical harmonic coefficients that can

be computed with a good level of accuracy. The condition number of the matrix to be inverted and the effects of the Tikhonov regularization used to stabilize the solution are also discussed. Firstly, a simple incoming plane wave of 63 Hz ($\theta = 90°$,$\phi = 45°$) in free field was selected as a target to evaluate the relation between the size of the virtual microphone array and the number of spherical harmonic coefficients. The sound field was analytically calculated in a rectangular domain of dimensions (5m, 10m, 3m) and captured by 4 different virtual cube arrays with linear dimensions of 0.8 m, 1.2m, 1.6m and 2 m, respectively. The spatial resolution between microphones corresponded to 0.1 m. Table 4.2 shows the condition number for different sizes of array and maximum order of spherical harmonics estimated.

Array	**N = 5**	**N = 6**	**N = 7**	**N = 8**	**N = 9**	**N = 10**
0.8 m (729 mics)	2.54e+05	6.97e+06	1.49e+08	4.74e+09	1.49e+11	7.41e+12
1.2 m (2197 mics)	3.63e+04	6.54e+05	9.56e+06	1.96e+08	4.17e+09	1.23e+11
1.6 m (4913 mics)	8.82e+03	1.19e+05	1.32e+06	2.03e+07	3.27e+08	7.04e+09
2 m (9261 mics)	2.86e+03	3.09e+04	2.77e+05	5.07e+08	4.40e+07	7.51e+08

TABLE 4.2: Condition number of matrix $\mathbf{Y}$ for different maximum order of spherical harmonics N and size of the microphone array, 63 Hz

Consistently with the results of the plane wave expansion, the condition number decreases as the size of the array increases. Nevertheless, regarding the order of spherical harmonics, the condition number increases with the maximum order N. Based on equation (4.12), a low order expansion is required to reconstruct the sound field in the array area. Spherical harmonic coefficients that are not significant to synthesize the sound field at 63 Hz within the radius predicted by (4.12) are being calculated which affects the stability of the solution. A configuration of order $N = 7$ and a cube-shape microphone array with linear dimensions of 1.6 m with a spatial resolution of 0.2 m (729 microphone positions), which provides a radius of accurate reconstruction of approximately 0.6 m when the inversion is carried out up to 447 Hz, was selected for the simulations. Figures 4.54 to 4.61 illustrate the synthesized sound field and the acoustical errors at the array positions 1 and 3. The black square represents the position of the array, the white contour indicates the area where the normalized error is smaller than -20 dB and the circle corresponds to the radius of validity computed using equation (4.12).

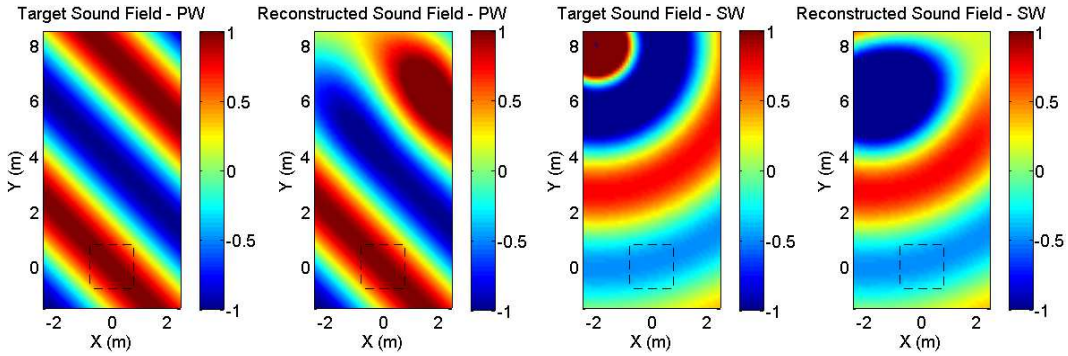


FIGURE 4.54: Target and reconstructed field at 63 Hz, array position 1, SHE

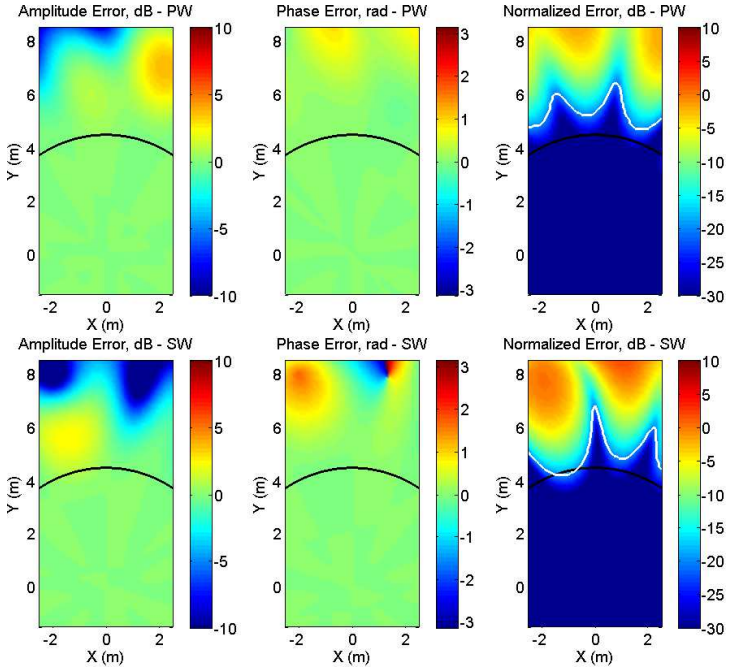


FIGURE 4.55: Errors at 63 Hz, array position 1, SHE

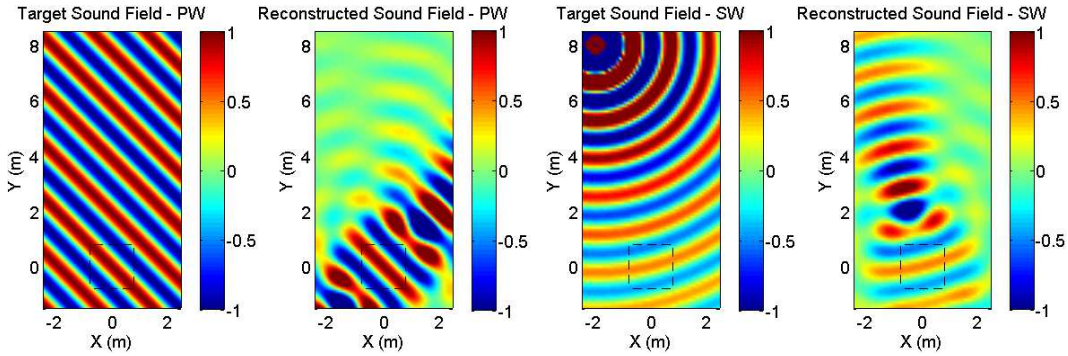


FIGURE 4.56: Target and reconstructed field at 250 Hz, array position 1, SHE

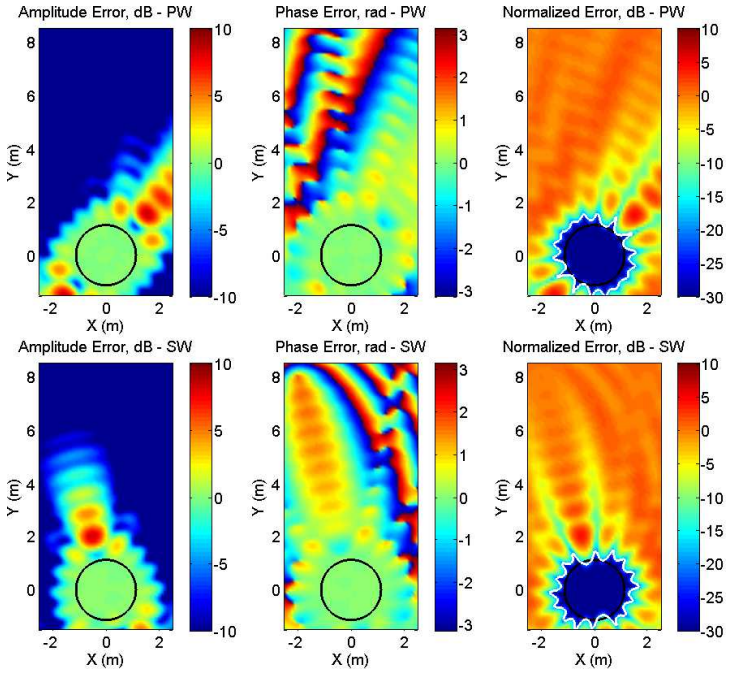


FIGURE 4.57: Errors at 250 Hz, array position 1, SHE

The results indicate that the inverse method is a suitable approach to estimate a set of complex coefficients A_{nm}, which provide a local description of the target acoustic field. Regarding the sound field reconstruction, Figures 4.54 and 4.56 show that the energy outside the radius predicted by equation (4.12) is highly reduced. This finding is a consequence of the truncation of the spherical harmonic expansion. Higher orders have a more significant contribution far from the centre of the expansion due to behaviour of the spherical Bessel functions. The lower energy in the reconstructed sound field beyond the radius predicted by equation (4.12) is confirmed by the analysis of the amplitude errors (Figures 4.55 and 4.57). A good agreement between the radius of validity (equation (4.12)) and the area where the normalized error was smaller than -20 dB was also found for the spherical harmonic expansion.

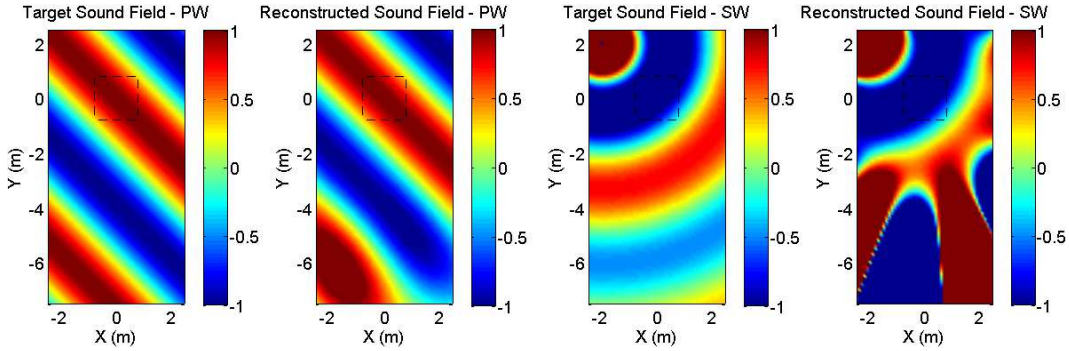


FIGURE 4.58: Target and reconstructed field at 63 Hz, array position 3, SHE

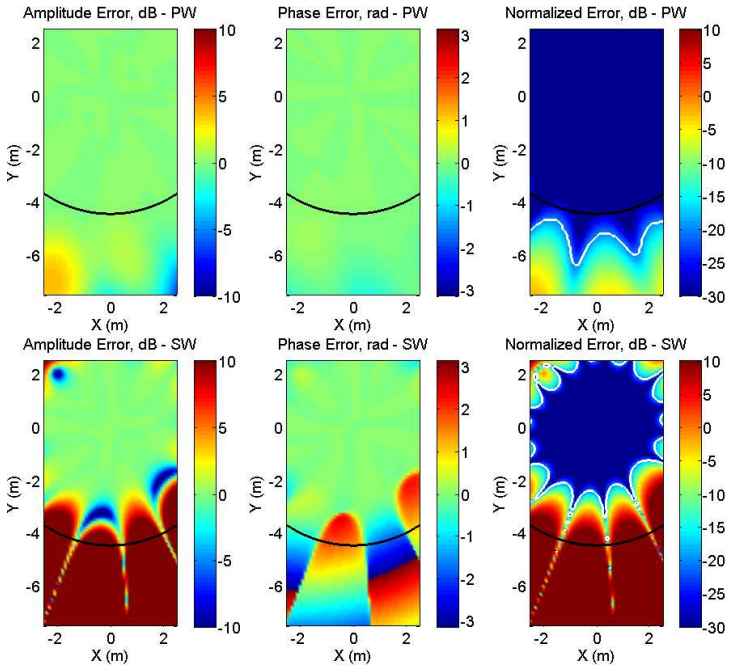


FIGURE 4.59: Errors at 63 Hz, array position 3, SHE

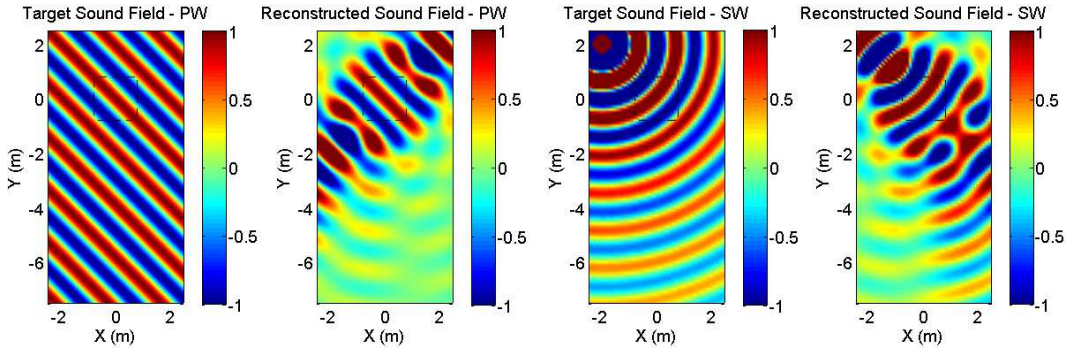


FIGURE 4.60: Target and reconstructed field at 250 Hz, array position 3, SHE

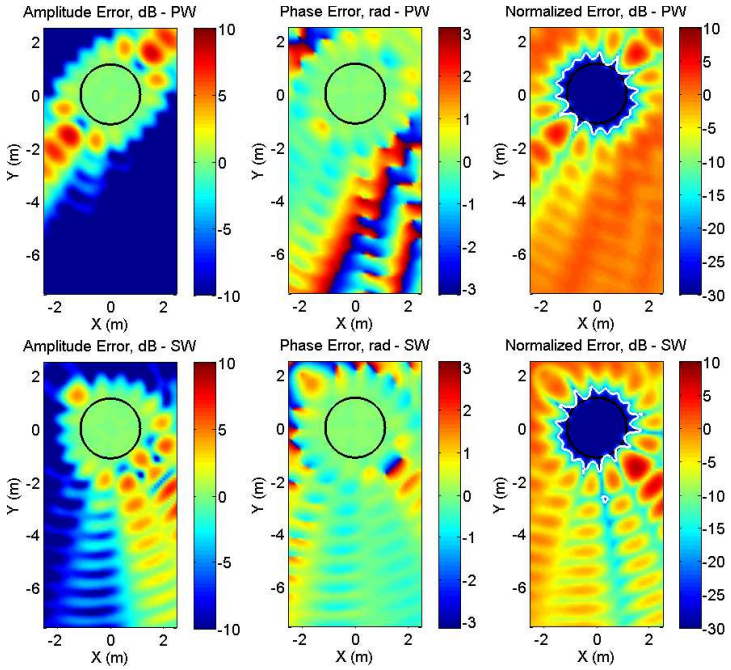


FIGURE 4.61: Errors at 250 Hz, array position 3, SHE

The figures indicate that, as expected, the area over which the reconstruction is accurate depends on frequency and is wider at lower frequencies. A comparison of the normalized errors shows good agreement between the area where the error is less than -20 dB and the radius predicted by equation (4.12) regardless of the position of the array. However, this relation remains in the case of the spherical field as long as the predicted radius is smaller than the distance between the centre of the array and the position of the point source. If that happens, the area of accurate reconstruction shrinks to an extent that depends on the distance of the source. This can be explained by the fact that the spherical harmonic formulation (interior case) relies on the assumption that no acoustic source is present within the area of reconstruction. Similar results were found for the amplitude and phase errors. A comparison between the sum of the squared absolute value of the A_{nm} coefficients for each order n predicted by the inverse method and those calculated analytically is presented in figure 4.62 for the three array positions. The analytical coefficients are calculated from equation (2.57).

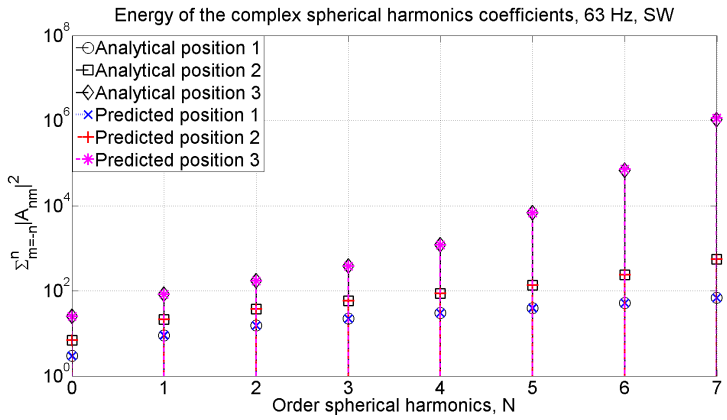


FIGURE 4.62: Comparison of the complex spherical harmonic coefficients, SW, 63 Hz

Figure 4.62 indicates that the inverse method is able to estimate with high accuracy the total energy of the complex spherical harmonic coefficients according to the order. A small mismatch was found for the highest order in the array position 3, which can be attributed to numerical errors. Nevertheless, Figure 4.58 shows zones with high acoustic pressure values outside of the radius predicted by equation (4.12) for the spherical field when the array is located closer to the source. This suggests that a large amount of energy is used at high orders to reconstruct the sound field outside the radius of reconstruction. This outcome is related to the contribution of the highest orders in the sound field reconstruction, which is higher further from the central point of the expansion due to the shape of the spherical Bessel functions.

4.1.2.1 Regularization

The implementation of Tikhonov regularization in the inversion of spherical harmonic matrix $\mathbf{Y}$ is addressed below. Figures 4.63 and 4.64 illustrate the reconstructed sound field and the acoustic errors at 63 Hz based on the array located at the position 3, but with Tikhnonov regularization. The case of 250 Hz is not illustrated because the regularization does not have a significant effect at this frequency. A comparison in terms of the energy of the complex spherical harmonic coefficients is shown in Figure 4.65.

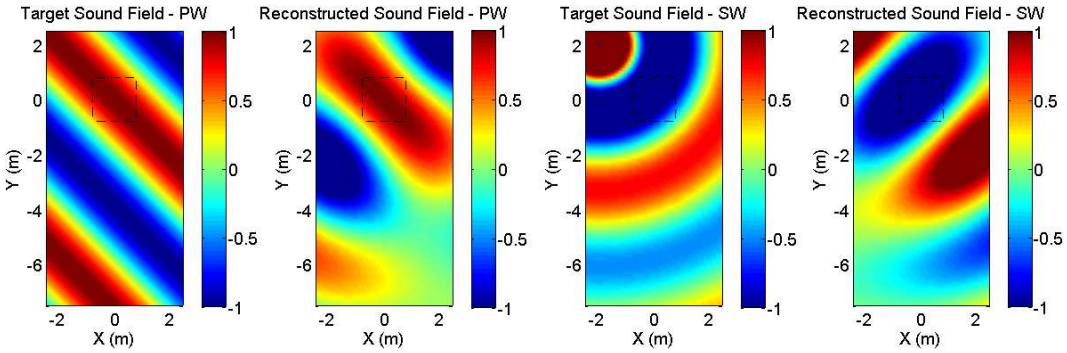


FIGURE 4.63: Target and reconstructed field at 63 Hz, regularized case $\beta = 0.04$, array position 3, SHE

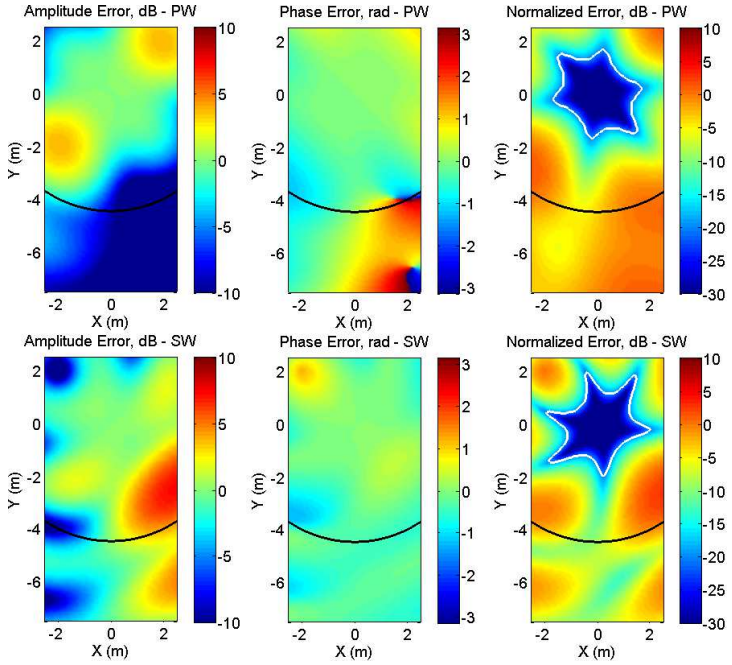


FIGURE 4.64: Errors at 63 Hz, regularized case $\beta = 0.04$, SHE

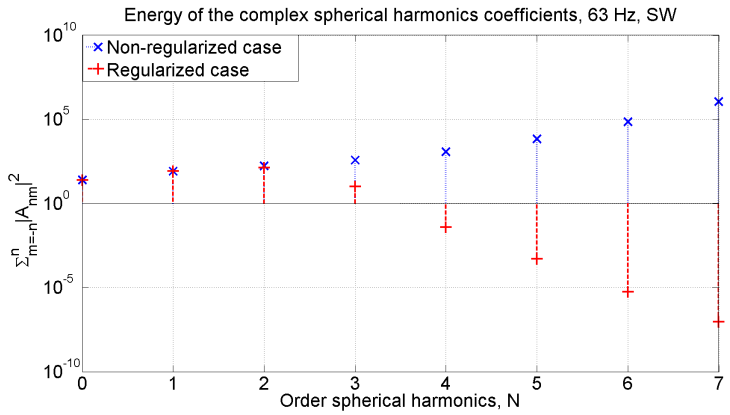


FIGURE 4.65: Comparison of the complex spherical harmonic coefficients, regularized, array position 3, SHE

The outcomes suggest that the implementation of Tikhnonov regularization drastically reduces the energy of the highest orders, but at the expense of reducing the radius of accurate reconstruction. However, the spatial average of the normalized error over the whole domain where the calculations were performed (see Figure 4.3) indicates that the sound field reconstruction is performed more accurately with regularization. These results show that the solution with regularization is more efficient in terms of the energy required to synthesize the sound field, but it is also more accurate in terms of the spatially-averaged normalized error when the whole domain is considered. Table 4.3 shows the spatially-averaged normalized acoustic error for the two cases.

Inversion	**Spatially-averaged normalized error (dB)**
Non-regularized	28.6
Regularized	-5.8

TABLE 4.3: Normalized acoustic error for the non-regularized and regularized case averaged over the domain of the calculation

4.1.3 Comparison between the plane wave and spherical harmonic expansion

A comparison of the plane wave and spherical harmonic representation has been carried out in terms of the acoustic errors. The frequency corresponds to 250 Hz to reduce the constraint imposed by the near field at 63 Hz. The number of spherical harmonic coefficients considered for the analysis is 64, which is the same as the

number of plane waves used to synthesize the sound fields. Figures 4.66 and 4.67 shows the reconstructed sound field based on a plane wave and spherical harmonic expansion, respectively. A plane and spherical acoustic wave propagation have been selected as a reference. The acoustic errors are illustrated in Figures 4.68 to 4.70. Finally, Table 4.4 indicates the acoustic errors spatially-averaged over the whole domain of the calculation (see Figure 4.3). For the case of amplitude and phase errors, the spatial average has been performed using the absolute value of these quantities.

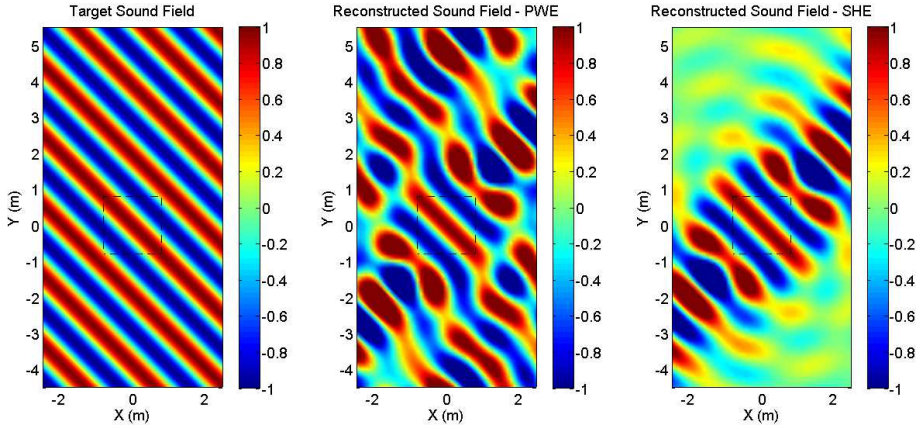


FIGURE 4.66: Comparison of reconstructed sound fields, plane wave, 250 Hz, PWE and SHE

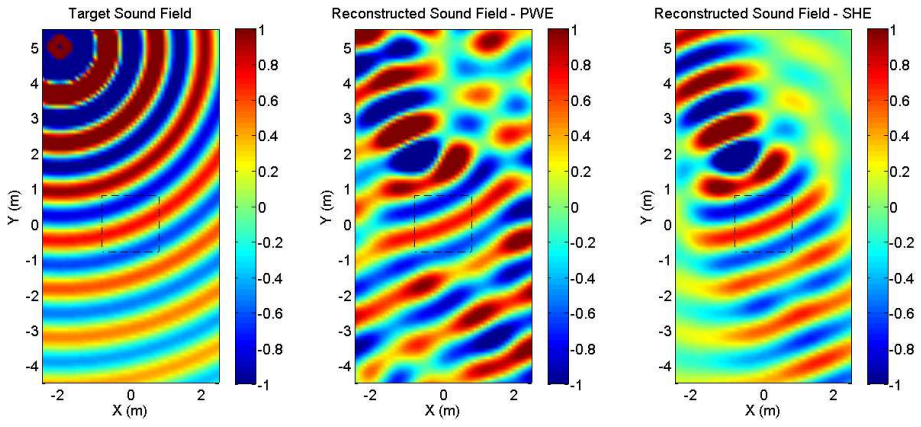


FIGURE 4.67: Comparison of reconstructed sound fields, spherical wave, 250 Hz, PWE and SHE

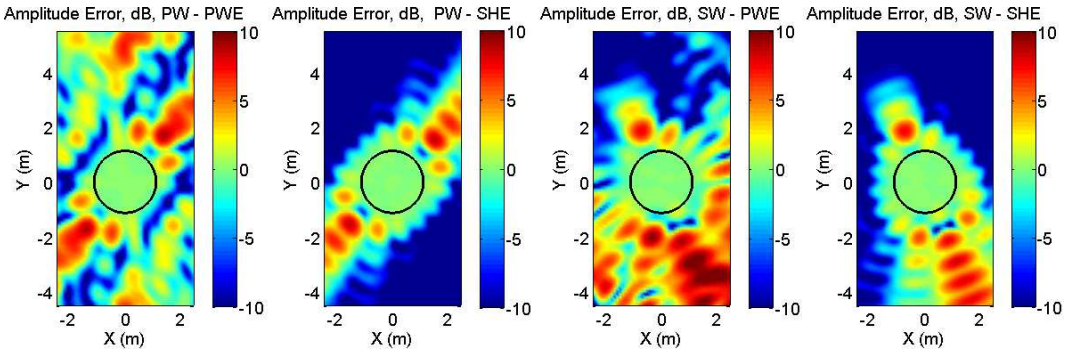


FIGURE 4.68: Amplitude error, 250 Hz, PW and SW

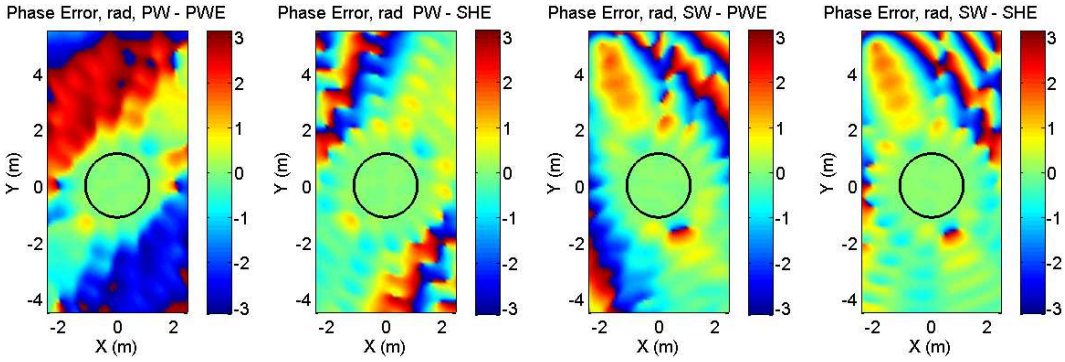


FIGURE 4.69: Phase error, 250 Hz, PW and SW

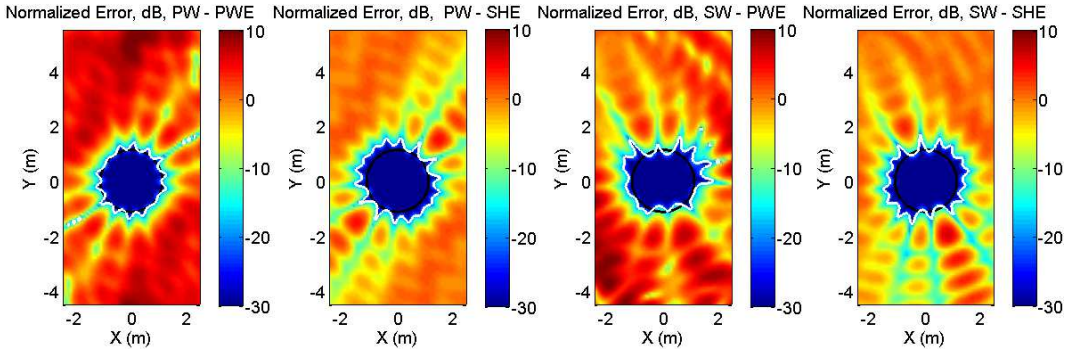


FIGURE 4.70: Normalized error - 250 Hz - PW and SW

Error	**PW-PWE**	**PW-SHE**	**SW-PWE**	**SW-SHE**
Amplitude (dB)	1.74	5.05	2.95	4.46
Phase (rad)	1.56	0.84	1.14	0.78
Normalized (dB)	-1.24	-5.64	-2.43	-6.24

TABLE 4.4: Spatially-averaged acoustic errors, 250 Hz, PWE and SHE

The results indicate that there is a good agreement between the area of accurate reconstruction and the region predicted by equation (4.12) for both approaches. However, the synthesized sound fields outside the region of validity are different. In the case of the spherical harmonic representation, the sound field is attenuated due to the absence of higher orders above order N, which mainly contribute to the reconstruction away from the origin and which were excluded from the representation. In contrast, the plane wave representation contains the information of higher orders, but they are incorrect. An analysis of the acoustical errors shows that the plane wave representation provides better accuracy in terms of amplitude but this is not the case for the phase and normalized error.

If the spherical harmonic representation is used to reconstruct the sound field using a multichannel technique such as Ambisonics, the decoding stage will produce the higher orders leading to results similar to the plane wave expansion. This means that the higher orders will be generated but not correctly. This outcome is related to the fact that the PWE and the SHE are just the spatial transformation of each other. A well established technique to decode Ambisonic signals is called the mode-matching approach [78]. The reconstruction of an acoustic field described by the complex coefficients A_{nm} using a set of L plane waves, each of them with different complex amplitude $\tilde{q}_l$ and direction (θ_l, ϕ_l), can be expressed using the Jacobi-Anger expansion as:

$$\sum_{n=0}^{N}\sum_{m=-n}^{n} A_{nm}(\omega) j_n(kr_x) Y_n^m(\theta,\phi) = 4\pi \sum_{l=1}^{L} \tilde{q}_l(\omega) \sum_{n=0}^{N} j^n j_n(kr) \sum_{m=-n}^{n} Y_n^m(\theta,\phi) Y_n^m(\theta_l,\phi_l)^*. \quad (4.24)$$

The simplification of the above equation using the orthogonality relation of the spherical harmonics yields the following mode-matching equation for each n and m

$$A_{nm}(\omega) = 4\pi \sum_{l=1}^{L}\sum_{n=0}^{N} j^n q_l(\omega) Y_n^m(\theta_l,\phi_l)^*, \quad (4.25)$$

for $n = 0...N$ and $|m| \leq n$. This is a finite set of linear equations that can be solved in terms of the least squares solution using an inverse method. In order to have

at least one solution, the number of spherical harmonics $(N+1)^2$ is required to be lower than, or equal to, the number of plane waves, namely $L \geq (N+1)^2$. Figures 4.71 and 4.72 show the comparison of reconstructed sound fields that have been predicted using 3 different approaches. The first method corresponds to a plane wave expansion whose complex amplitudes $\mathbf{q}(\omega)$ have been calculated according to section 4.1.1. The second one is a spherical harmonic representation in which the complex coefficients $\mathbf{a}(\omega)$ have been estimated according to section 4.1.2. Finally, the last approach is a plane wave expansion whose complex amplitudes $\mathbf{q}(\omega)$ are predicted by implementing a decoding stage to the complex spherical harmonic coefficients $\mathbf{a}(\omega)$. This is performed by solving equation (4.25) for $\tilde{\mathbf{q}}$, namely

$$\tilde{\mathbf{q}} = \mathbf{Y}^{\dagger}\mathbf{a}. \tag{4.26}$$

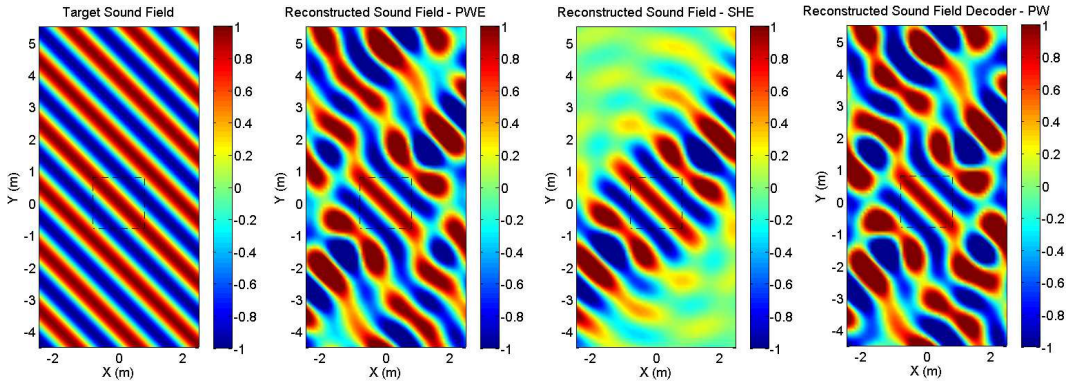


FIGURE 4.71: Comparison of reconstructed sound fields, plane wave, 250 Hz, PWE, SHE and Decoder

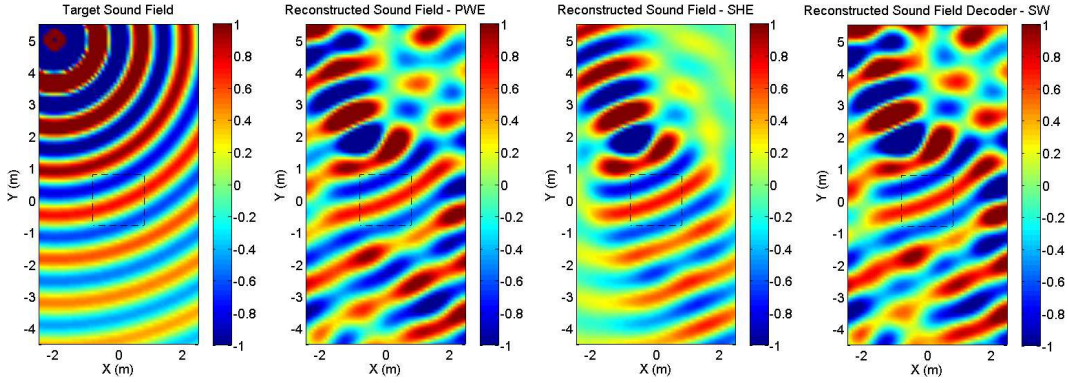


FIGURE 4.72: Comparison of reconstructed sound fields, spherical wave, 250 Hz, PWE, SHE and Decoder

Figures 4.71 and 4.72 indicate that the decoding of spherical harmonic coefficients into a set of L plane waves leads to a similar results as a plane wave expansion of L plane waves whose amplitudes $\mathbf{q}(\omega)$ have been estimated by the inverse problem formulated in section 4.1.1. The reason for the discrepancy may be associated to the implementation of regularization to predict $\mathbf{q}(\omega)$. Although the plane wave and spherical harmonic expansions are equivalent, the discretization of the domain and the truncation of the series generate different artifacts for each representation. An analysis of both methods is conducted by the transformation of the PWE in terms of spherical harmonics using equation 4.19. Then, the resulting coefficients $\tilde{A}_{nm}$ are compared with the A_{nm} coefficients calculated from the analytical expressions given by equations (2.56) or (2.57) according to the type of wave propagation. Figure 4.73 illustrates the sum of the absolute value of the coefficients $\tilde{A}_{nm}$ and A_{nm} squared for each order n up to 10th order. Results indicates that the coefficients are the same up to 7th, which suggests that the equivalence between the plane wave and spherical harmonic representation holds only up to the order given by the number of plane waves ($N = \sqrt{64} - 1$), namely 7th.

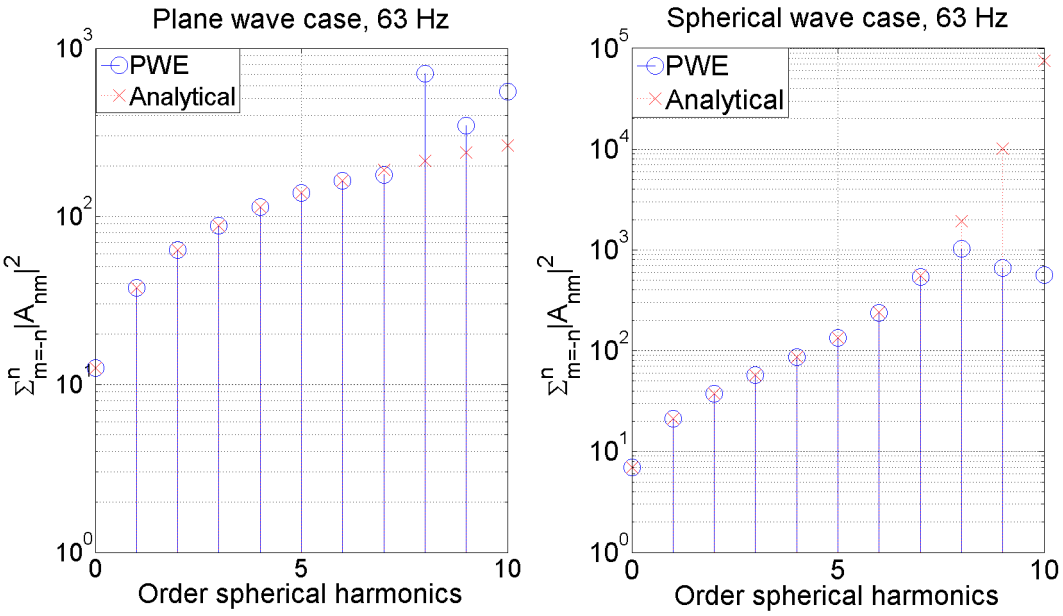


FIGURE 4.73: Comparison of A_{nm} coefficients computed from the PWE and the analytical expression

4.2 Encoding directional information from FE simulations

The previous sections support the concept of encoding spatial information from acoustic pressure meshed data by means of an inverse method. In this section, results obtained from FE simulations are processed using a plane wave expansion for the same purpose. Evaluations of the reconstructed sound fields are presented for several reference cases. The use of regularization in the inversion of the propagation matrix $\mathbf{H}(\omega)$ is analyzed based on the acoustic errors and the energy distribution of the plane wave density.

The complex acoustic pressures obtained from the finite element method simulations are used to generate a cubic virtual microphone array with linear dimensions of 1.6 m. The spacing between microphones was chosen to be 0.2 m based on the results obtained in section 4.1.1.2. Because the position of the virtual microphones does not necessarily match with the nodes of the mesh, the complex acoustic pressures at those locations are determined by an interpolation process.

4.2.1 Single reflecting wall

This scenario corresponds to a rectangular domain with dimensions of (5m, 10m, 3m) whose boundaries provide an approximate 100 % of sound absorption except for one of them, which is rigid. This is done by coupling perfect matched layers at the boundaries where no reflections are intended to be generated. The simulations were conducted using a mesh resolution of 6 elements per wavelength. The elements used for the discretization of the domain correspond to regular tetrahedrons. Two virtual microphone arrays with central locations at (4, 1.5, 1.2) and (4, 5.5, 1.2) were considered. The acoustic source was located at (4, 9.5, 1.2). Figure 4.74 illustrates a sketch of the domain with the virtual microphone array located at position 1 and the cross-section used ($z = 1.2$ m) to display the results.

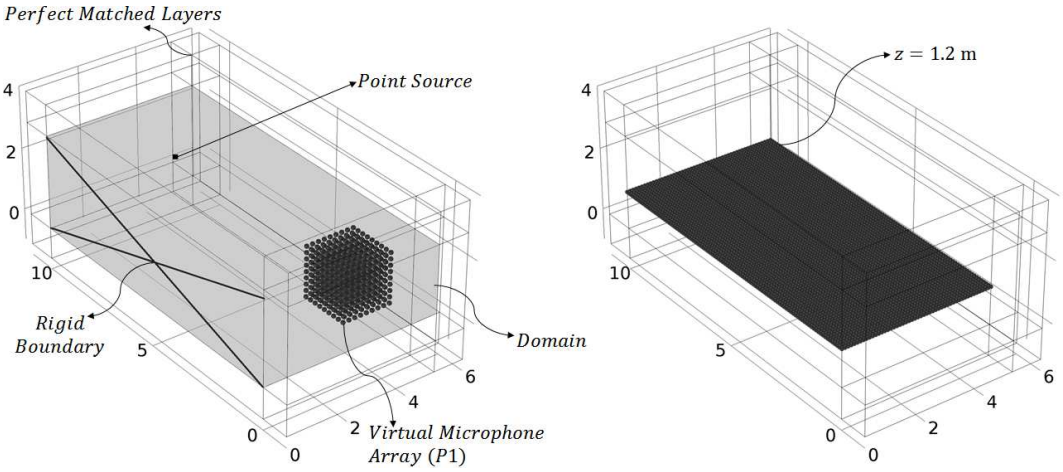


FIGURE 4.74: Sketch of the single-wall model, FEM

Figures 4.75 and 4.76 illustrate the reconstruction of the sound field at position 1 by means of the inverse method with and without regularization, respectively. The frequencies of 63 Hz and 250 Hz have been considered for the analysis. The results indicate that the inverse method is a suitable approach to generate a plane wave representation of the acoustic fields predicted by FEM. A comparison in terms of the energy distribution of the plane wave density is presented in Figures 4.77 and 4.78. The white circles represent the direction of the acoustic source ($\theta = 90°$, $\phi = 90°$) and its image ($\theta = 90°$, $\phi = 135°$), which is a good representation of the reflection produced by the hard boundary in this specific case.

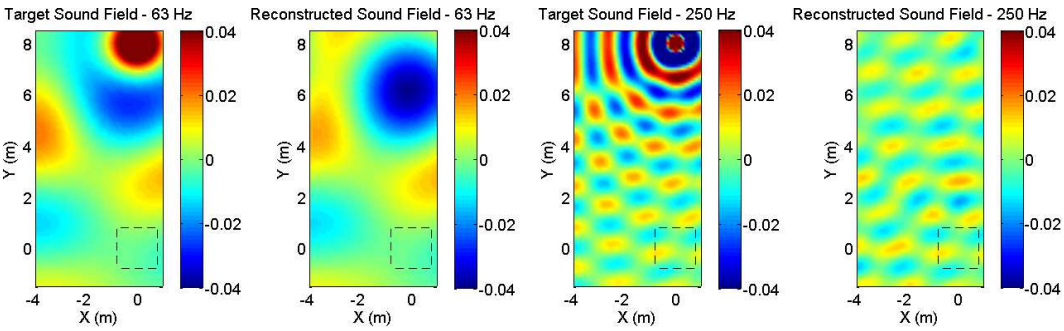


FIGURE 4.75: Target and reconstructed field at 63 Hz and 250 Hz, single wall, position 1

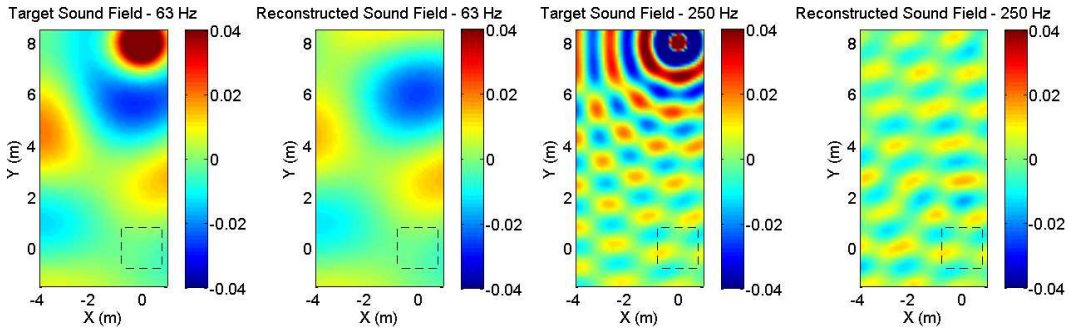


FIGURE 4.76: Target and reconstructed field at 63 Hz and 250 Hz, single wall, regularized $\beta = 0.33 \times 10^{-3}$, position 1

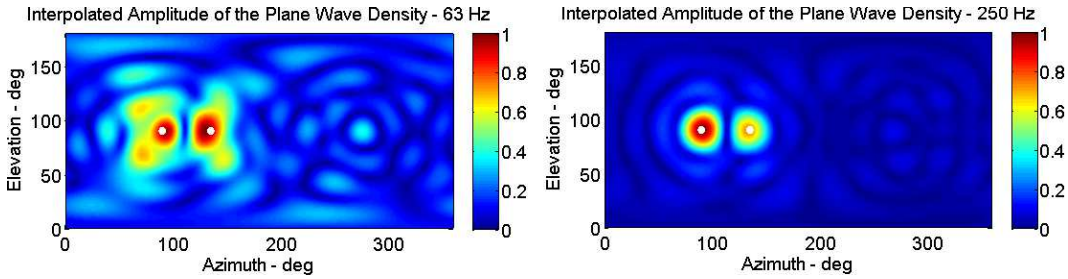


FIGURE 4.77: Normalized amplitude PWE at 63 Hz and 250 Hz, single wall, position 1, $q_{total}(63)=1.56 \times 10^{-4}$, $q_{total}(250)=7.1 \times 10^{-5}$

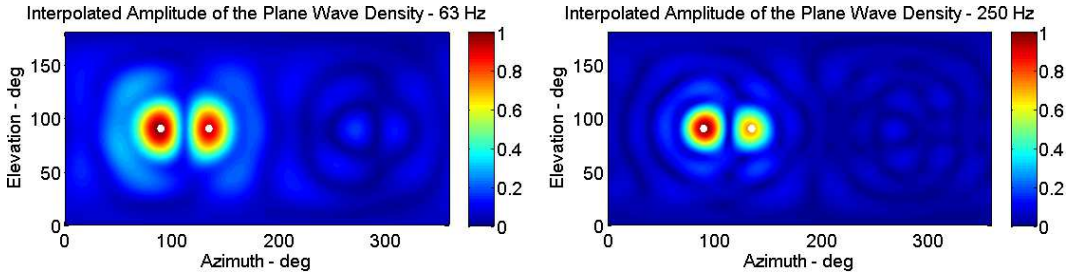


FIGURE 4.78: Normalized amplitude PWE at 63 Hz and 250 Hz, regularized $\beta = 0.33 \times 10^{-3}$, single wall, position 1, $q_{total}(63)=4.17 \times 10^{-5}$, $q_{total}(250)=7.1 \times 10^{-5}$

The figures reveal that the inclusion of regularization in the inverse problem leads to similar outcomes to those obtained in section 4.1.1.3. A more directionally concentrated distribution in the energy of the plane wave density was found for the regularized case at 63 Hz. Furthermore, this concentration is focused on the direction given by the image source model. Figure 4.79 also indicates that the application of regularization yields a more efficient approach in terms of the amplitude of the

plane waves required to synthesize the sound fields. The acoustic errors are presented in Figures 4.80 and 4.81.

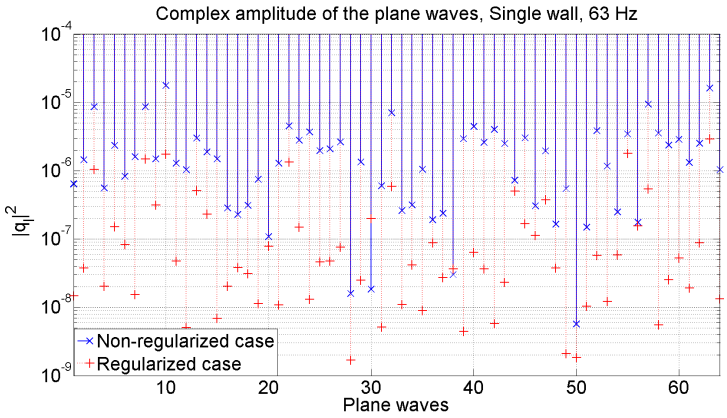


FIGURE 4.79: Complex amplitude of the PWE, single wall, 63 Hz, position 1

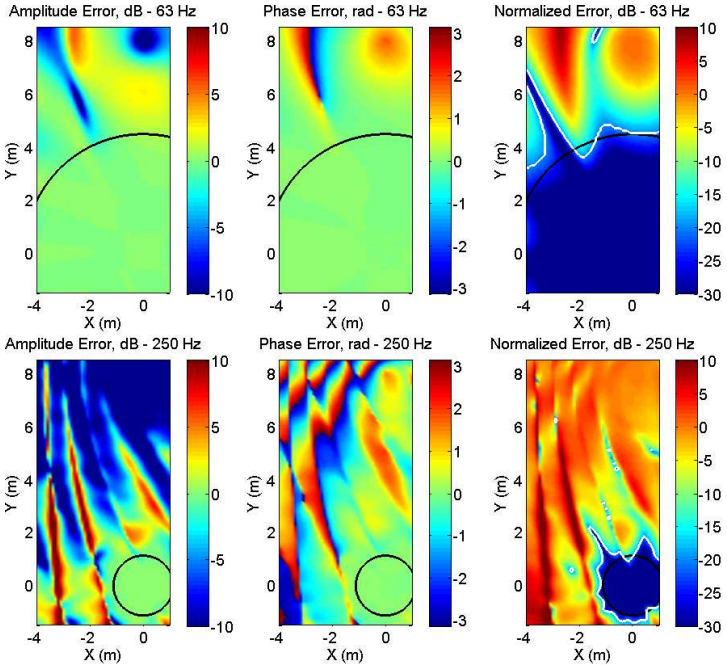


FIGURE 4.80: Errors at 63 Hz and 250 Hz, single wall, position 1

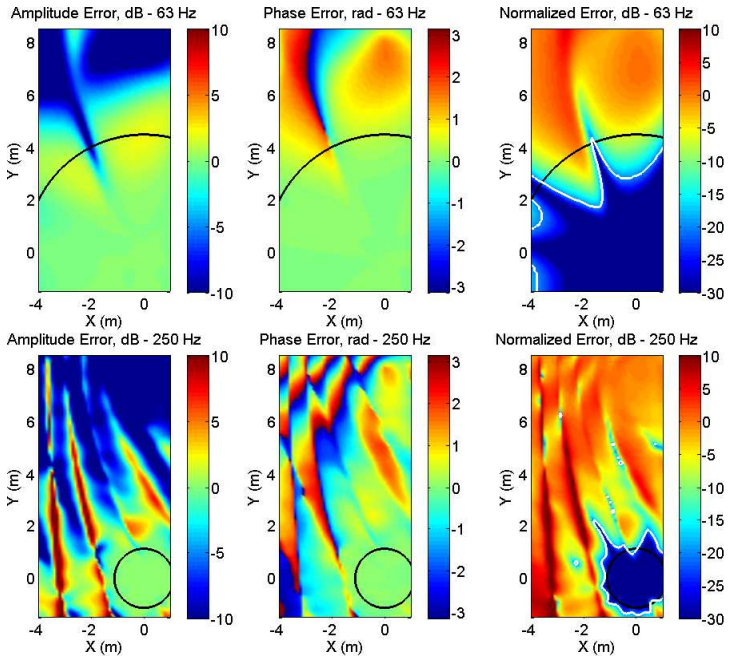


FIGURE 4.81: Errors at 63 Hz and 250 Hz, regularized $\beta = 0.33 \times 10^{-3}$, single wall, position 1

As expected, the application of regularization reduces the area where the normalized error is smaller than -20 dB at the frequencies were the regularization has a relevant effect in the inversion of the problem. Figures 4.82 and 4.83 illustrate the sound field reconstruction for the array position 2 with and without regularization, respectively.

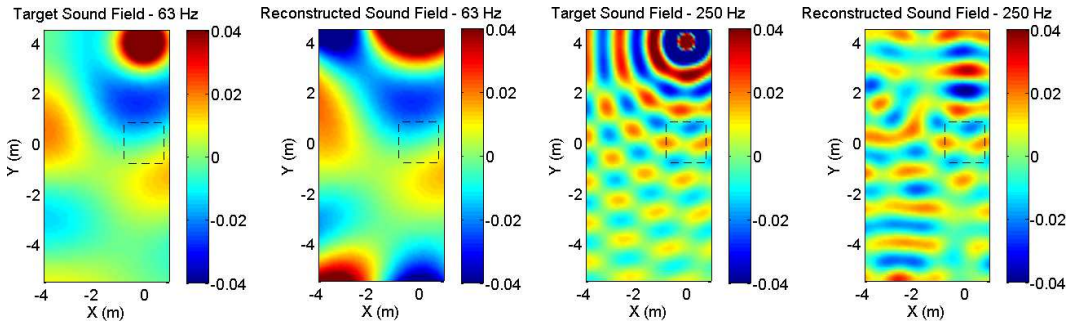


FIGURE 4.82: Target and reconstructed field at 63 Hz and 250 Hz, single wall, position 2

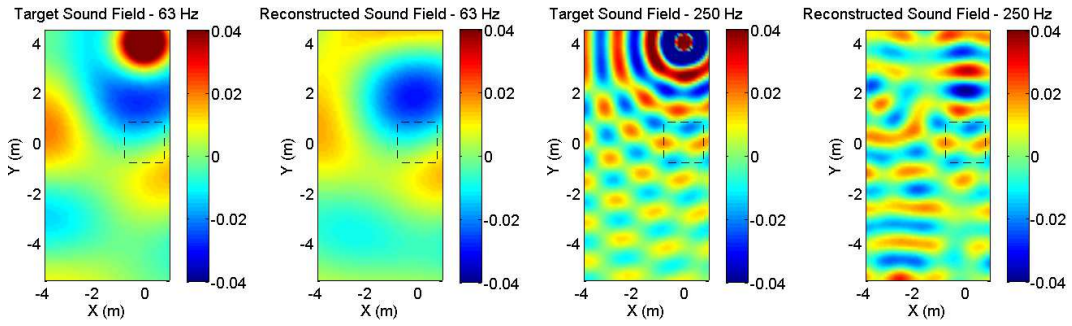


FIGURE 4.83: Target and reconstructed field at 63 Hz and 250 Hz, single wall, regularized $\beta = 0.33$, position 2

The outcomes are consistent with previous simulations showing that for the non-regularized case the amplitude of the reconstructed field is bigger than the target field in some areas when the array is close to the acoustic source. Figures 4.84 and 4.88 show the spatial distribution of the energy in the plane wave density. The incoming direction of the corresponding image source is ($\theta = 90°$, $\phi = 153.4°$) for this configuration.

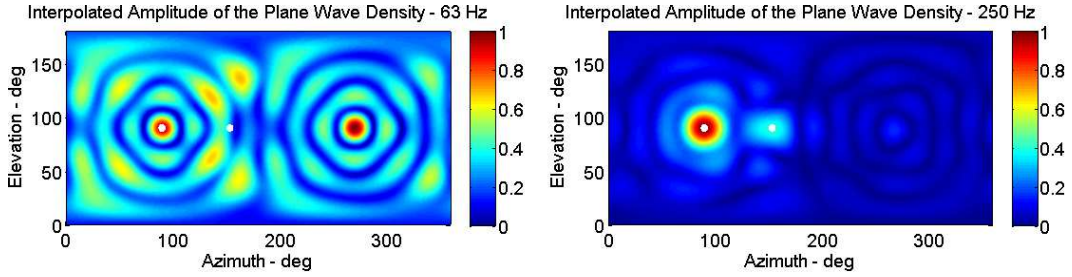


FIGURE 4.84: Normalized amplitude PWE at 63 Hz and 250 Hz, single wall, position 2, $q_{total}(63)=2.8 \times 10^{-3}$, $q_{total}(250)=2.0 \times 10^{-4}$

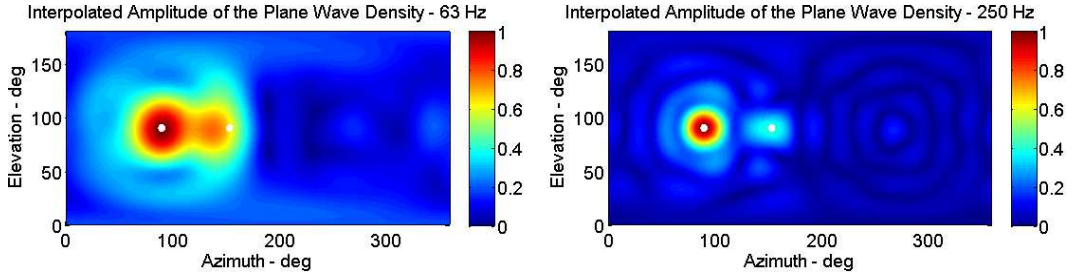


FIGURE 4.85: Normalized amplitude PWE at 63 Hz and 250 Hz, regularized $\beta = 0.33$, single wall, position 2, $q_{total}(63)=6.73 \times 10^{-5}$, $q_{total}(250)=2.0 \times 10^{-4}$

The figures reveal that regularization has an important effect in the spatial distribution of the energy of the plane wave density at 63 Hz. For this frequency, a higher directionally concentrated representation was found when regularization is implemented. A good agreement between the incoming direction of the image source and the energy distribution across the plane wave density was found at 250 Hz. At 63 Hz, a higher agreement using regularization was found compared to the non-regularized case. Figure 4.86 also shows that the complex amplitude of the plane waves is much lower when the regularization is applied. The acoustic errors are presented in Figures 4.87 and 4.88 for the non-regularized and regularized cases, respectively.

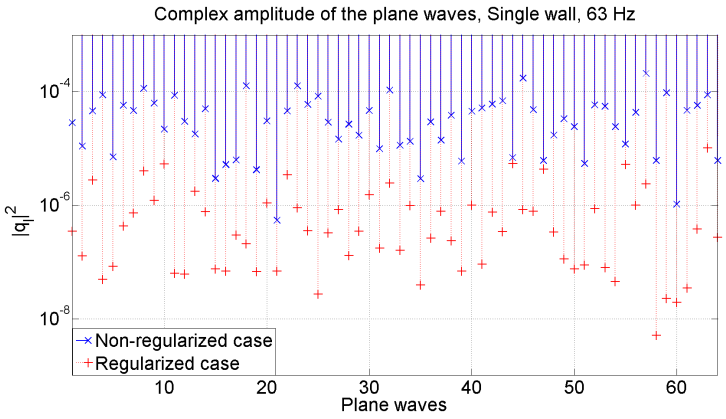


FIGURE 4.86: Complex amplitude of the PWE, single wall, 63 Hz, position 2

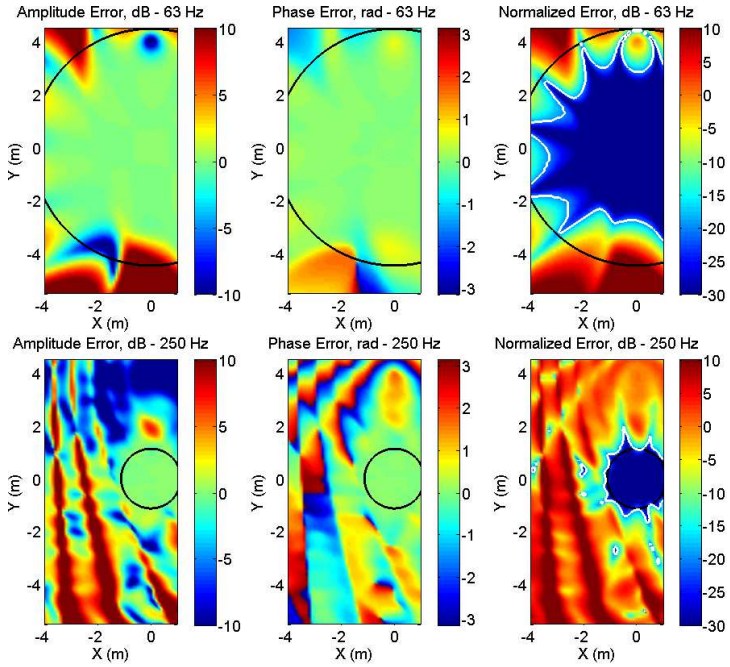


FIGURE 4.87: Errors at 63 Hz and 250 Hz, single wall, position 2

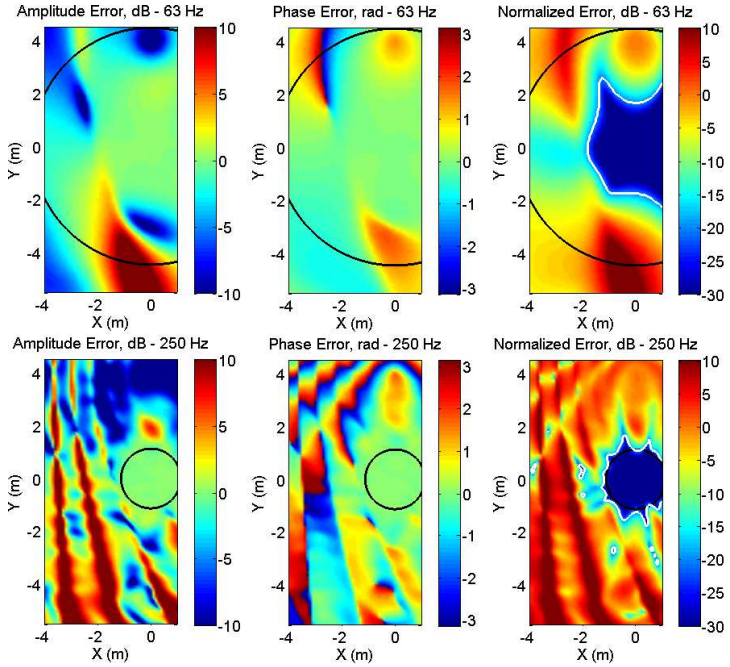


FIGURE 4.88: Errors at 63 Hz and 250 Hz, regularized $\beta = 0.33$, single wall, position 2

The analysis of the normalized errors indicates that the area where the error is smaller than -20 dB shrinks to approximately 3 m^2 when regularization is applied. However,

the amplitude errors suggest that the reconstructed sound field is quieter than the target field in some areas for the regularized case. The opposite happens for the non-regularized case where the reconstructed field is louder. This result may have a significant effect when a translation operator for the sound field is applied because zones with much higher acoustic energy are not desired.

4.2.2 Meeting room

A more complex environment was selected to carry out the simulations and evaluate the robustness of the inverse method. The meeting room 4079 is a small room where the distance between source(s)-receiver(s) is relatively short, thus undermining the assumption of plane wave propagation. This condition provides a good scenario to test the applicability of regularization, the accuracy of the sound field reconstruction and the distribution of the energy across the plane wave density. A sketch of the setup is presented in Figure 4.89. The source position corresponded to B1 according to the configuration described in Figure 3.2.

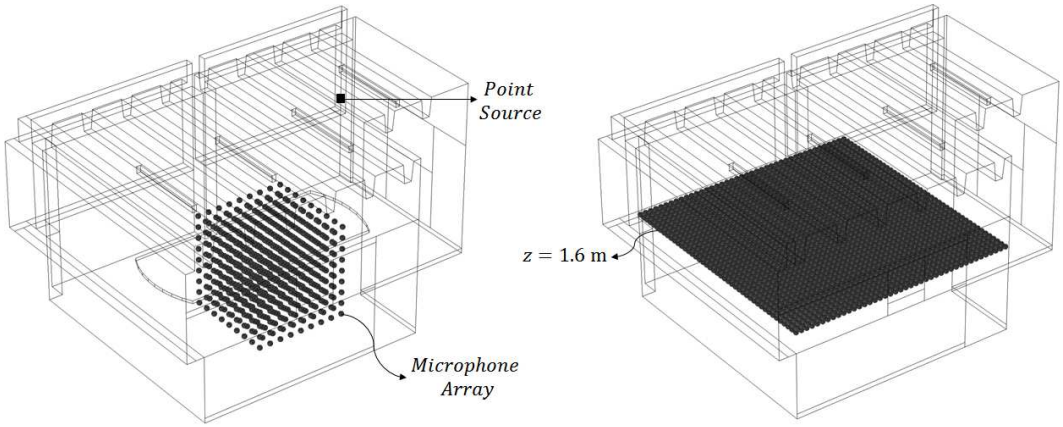


FIGURE 4.89: Sketch of the meeting room model, FEM

Figures 4.90 to 4.93 show the reconstruction of the sound field on the cross-section illustrated in Figure 4.89. The regularized and non-regularized cases are compared at 63 Hz and 250 Hz. The results indicate that the inverse method is able to predict a plane wave expansion whose complex amplitudes synthesize the sound field even at low frequencies where the plane wave propagation assumption is not completely fulfilled.

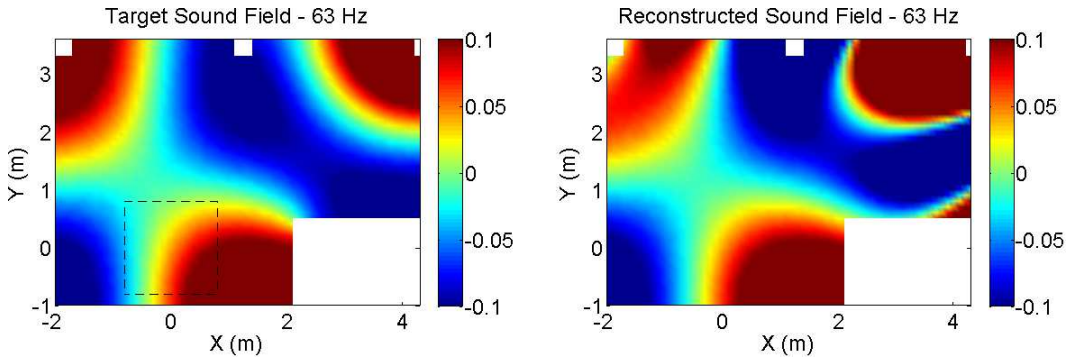


FIGURE 4.90: Target and reconstructed field, meeting room, 63 Hz

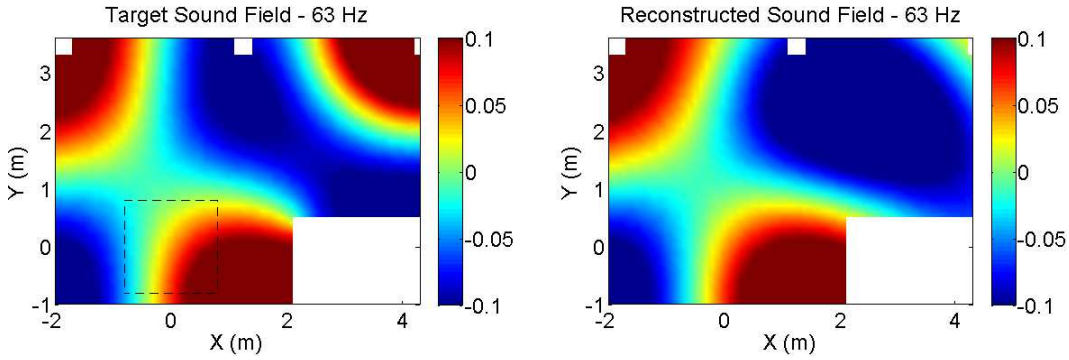


FIGURE 4.91: Target and reconstructed field, regularized $\beta = 0.33$, meeting room, 63 Hz

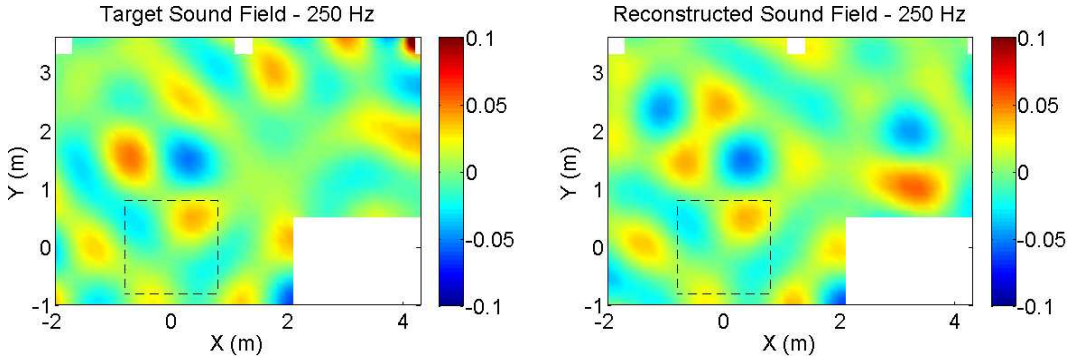


FIGURE 4.92: Target and reconstructed field, meeting room, 250 Hz

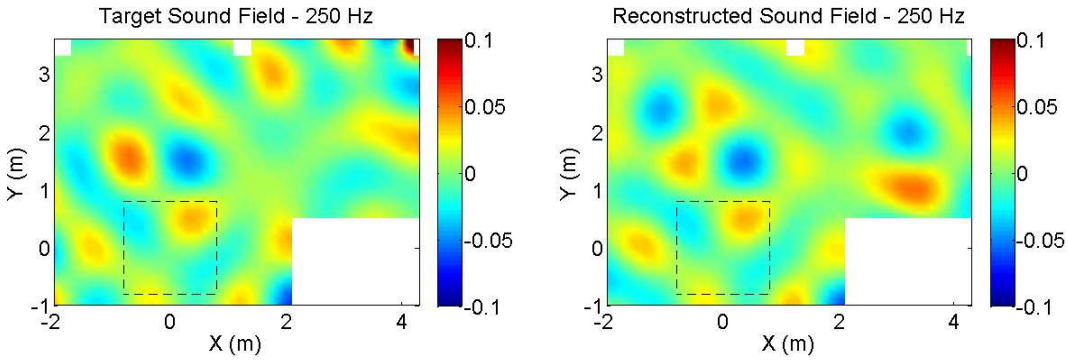


FIGURE 4.93: Target and reconstructed field, regularized $\beta = 0.33$, meeting room, 250 Hz

The spatial energy distribution of the plane wave density is presented in Figures 4.94 and 4.95 for the non-regularized and regularized cases. The results show that a more directionally concentrated energy distribution of the plane wave density is obtained using regularization at 63 Hz. An interesting result was also found by the implementation of regularization at this frequency, the reconstructed sound field seems to be dominated by the spherical harmonic of order 2 and mode 2 [14] (see Figure 4.96).

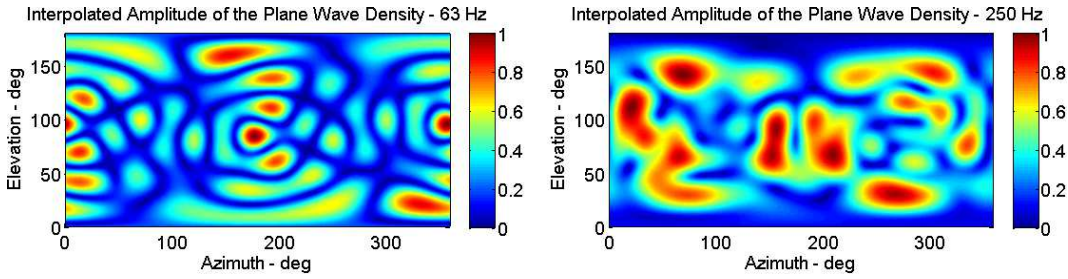


FIGURE 4.94: Normalized amplitude PWE, meeting room, $q_{total}(63)$=26.11, $q_{total}(250)=7 \times 10^{-4}$

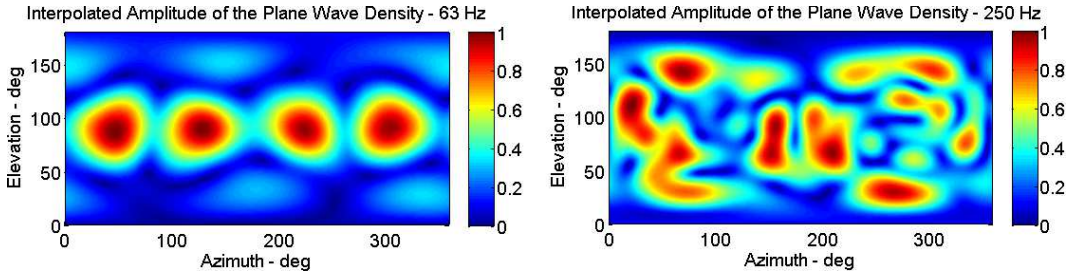


FIGURE 4.95: Normalized amplitude PWE, regularized $\beta = 0.33$, meeting room, $q_{total}(63)=2.6 \times 10^{-3}$, $q_{total}(250)=7 \times 10^{-4}$

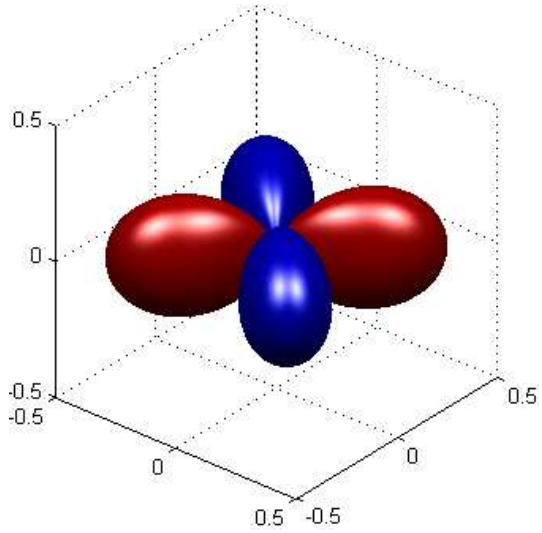


FIGURE 4.96: Imaginary part of the spherical harmonic of order n = 2 and mode m = 2

Figure 4.97 shows a significant reduction in the energy of the plane waves when regularization is applied. A difference between 1 and 2 orders of magnitude was found for the non-regularized and regularized cases. The acoustic errors are presented in Figures 4.98 to 4.103.

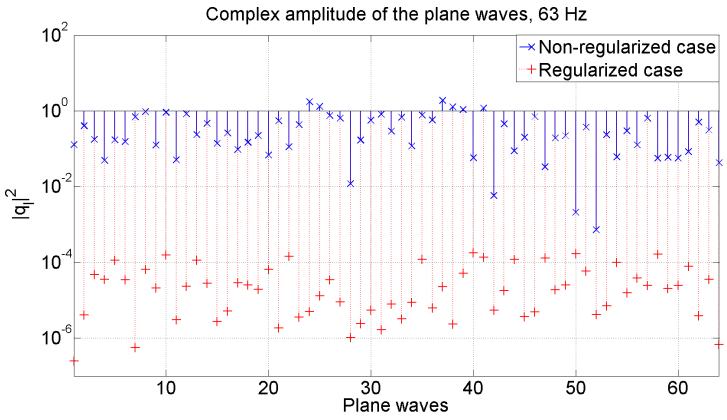


FIGURE 4.97: Complex amplitude of the PWE, meeting room

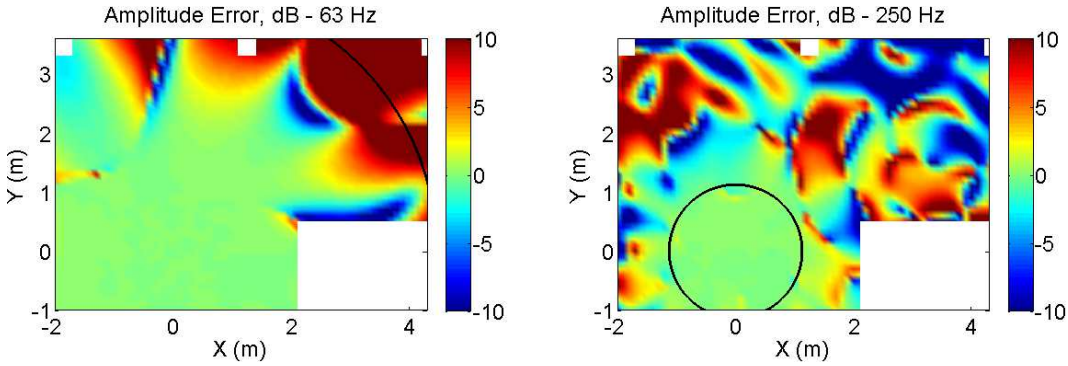


FIGURE 4.98: Amplitude error, meeting room

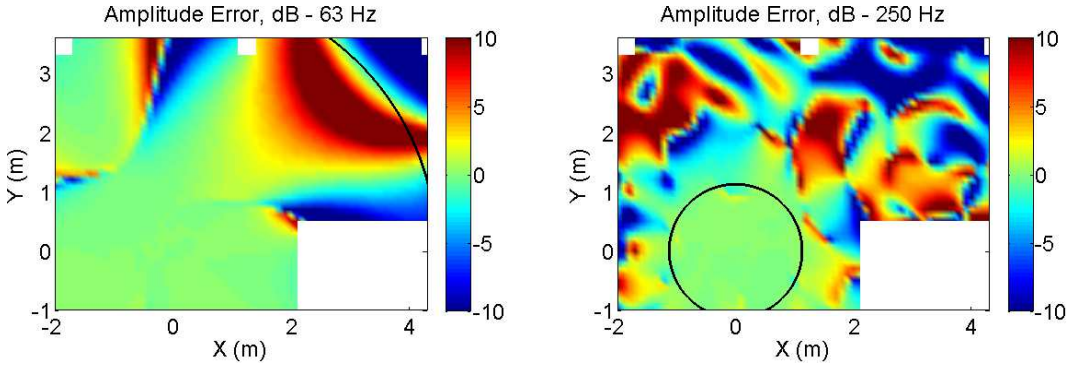


FIGURE 4.99: Amplitude error, regularized $\beta = 0.33$, meeting room

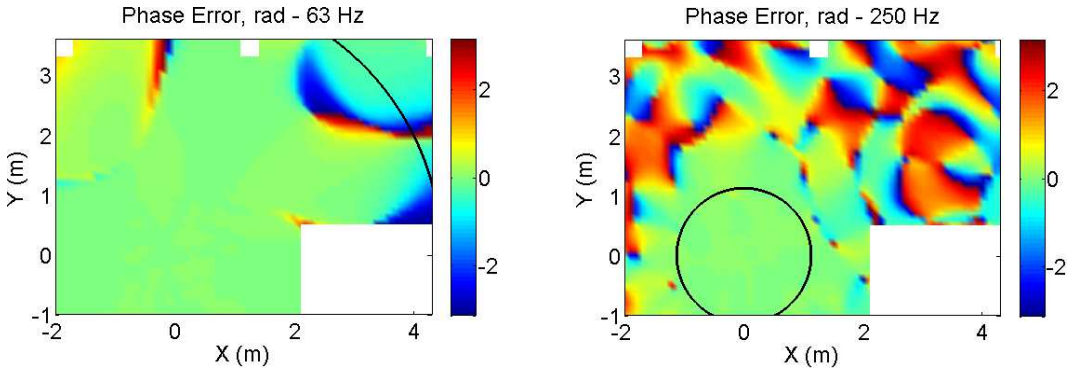


FIGURE 4.100: Phase error, meeting room

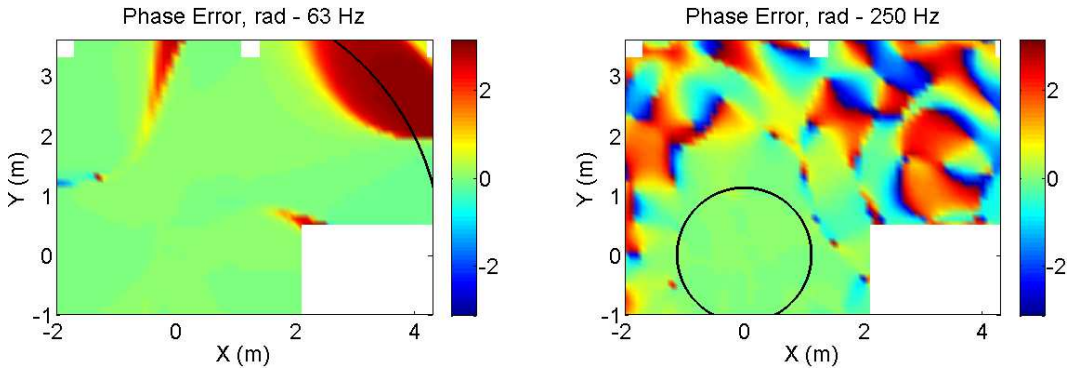


FIGURE 4.101: Phase error, regularized $\beta = 0.33$, meeting room

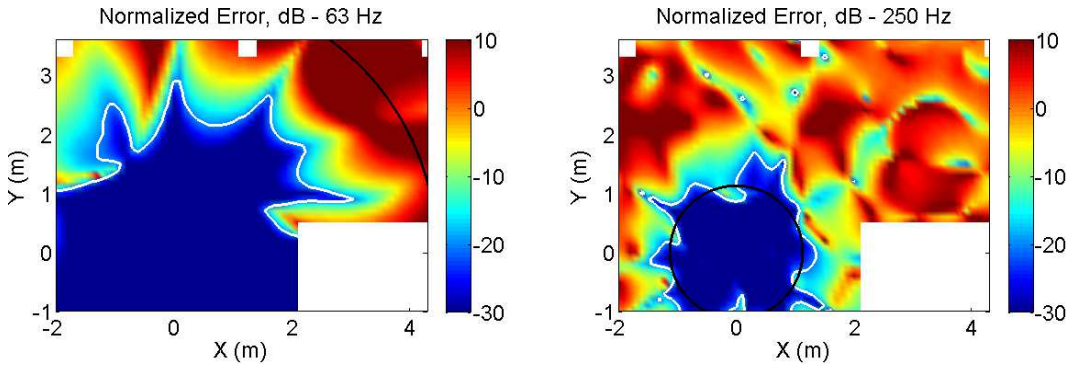


FIGURE 4.102: Normalized error, meeting room

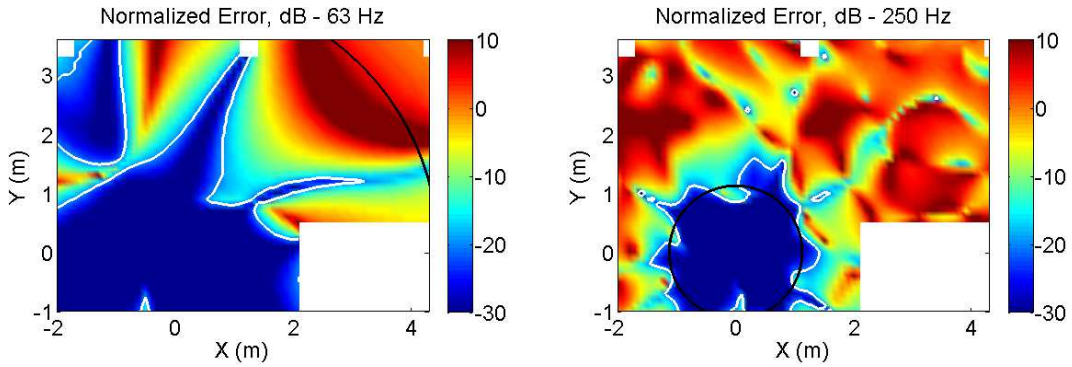


FIGURE 4.103: Normalized error, regularized $\beta = 0.33$, meeting room

Although the reduction in amplitude of the plane waves is significant when regularization is applied, the analysis of the acoustic errors reveals that the shrink of the area where the normalized error is smaller than -20 dB is not as significant as the previous results presented for the other reference cases. This outcome highlights the

importance of evaluating the regularization parameter for each specific problem because there is not a direct relation between the amount of regularization and the area where the normalized error is smaller than -20 dB. Regarding the radius where the reconstruction is accurate, the normalized errors show good agreement with equation (4.12) only in the case of 250 Hz; for 63 Hz, the area is overestimated. This condition is related to the presence of the walls. In this case, the homogeneous Helmholtz equation is not satisfied in the whole interior region due to the presence of reflective boundaries, which play a similar role as acoustic sources. Finally, in accordance with the results obtained for the single wall case, the amplitude errors show that the reconstructed sound field is quieter than the target field at 63 Hz when regularization is applied.

4.2.3 Ightham Mote

A plane wave expansion derived from an inverse method was generated on the source-receiver path B0R3 (see Figure 3.7) because this was the configuration that achieved the best agreement between the FE model and the measurements in terms of time-domain parameters. Figure 4.104 illustrates the location of the virtual microphone array and the area of sampled complex acoustic pressure data used for the analysis. The reconstruction of the acoustic fields at 63 Hz and 250 Hz is presented in Figures 4.105 and 4.106 for the non-regularized and regularized cases, respectively.

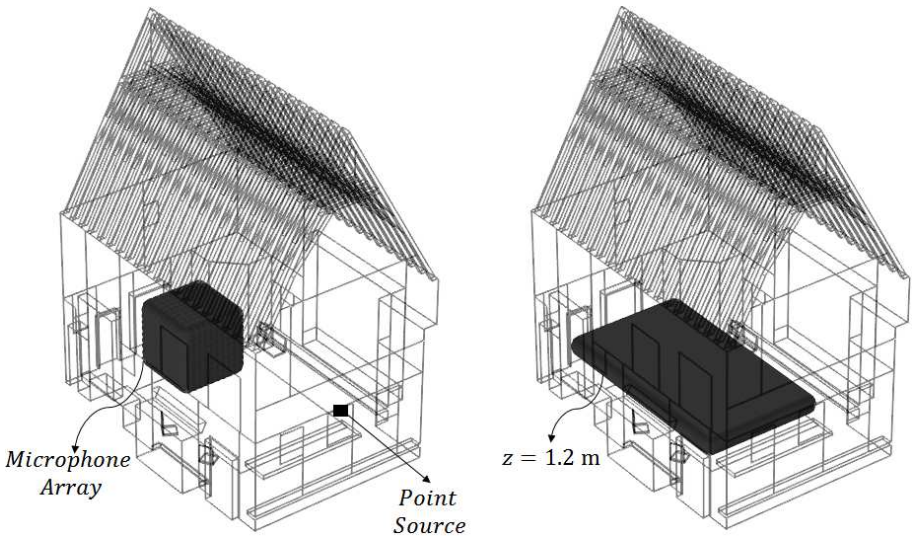


FIGURE 4.104: Sketch of the Ightham Mote model, FEM

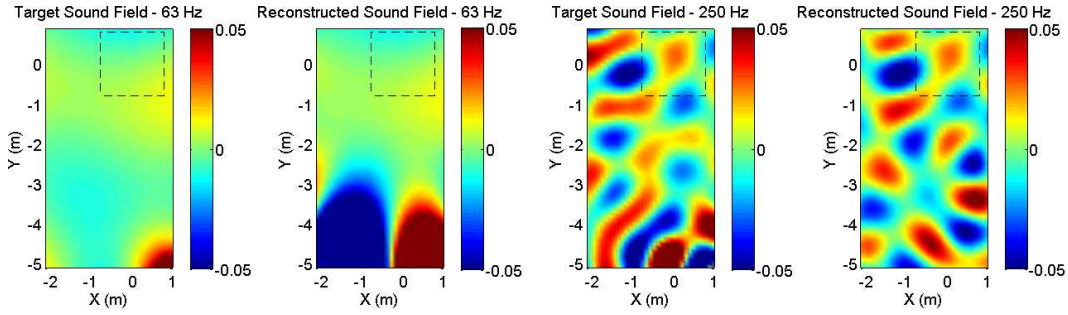


FIGURE 4.105: Target and reconstructed field, Ightham Mote

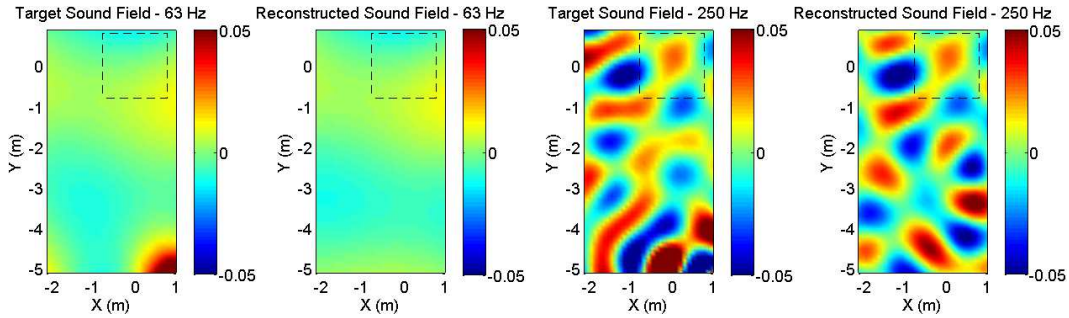


FIGURE 4.106: Target and reconstructed field, regularized $\beta = 0.05$, Ightham Mote

The figures indicate that, as expected, the sound field reconstruction is correctly performed around the microphone array. However, zones of high acoustic pressure close to the location of the acoustic source were found for the non-regularized case. Figures 4.107 and 4.108 show the energy distribution across the plane wave density.

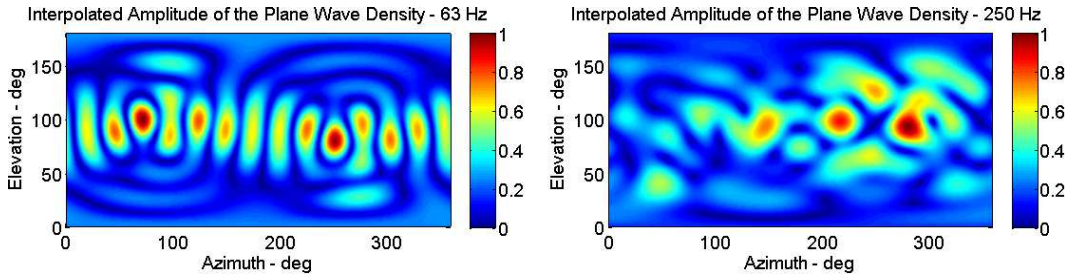


FIGURE 4.107: Normalized gain PWE, Ightham Mote, $q_{total}(63)$=0.89, $q_{total}(250)$=9.53×10^{-4}

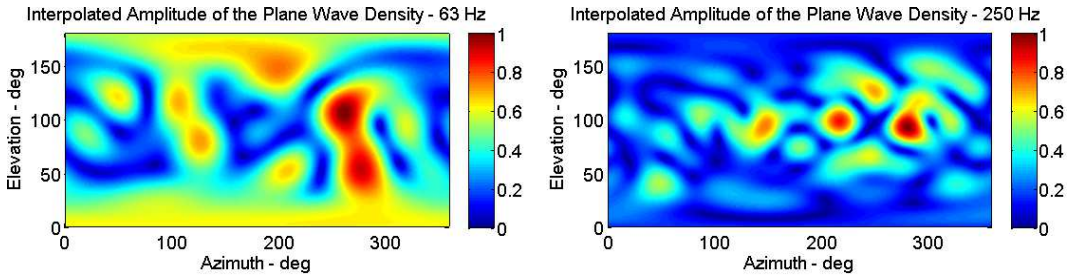


FIGURE 4.108: Normalized amplitude PWE, regularized $\beta = 0.05$, Ightham Mote, $q_{total}(63)=9.14 \times 10^{-5}$, $q_{total}(250)=9.47 \times 10^{-4}$

The results indicate that the regularization yields a more compact representation. However, the diffuse behaviour of the sound field in this enclosure is evident in comparison with the meeting room. The complex amplitude of the plane waves and the acoustic errors are presented in Figures 4.109 to 4.111.

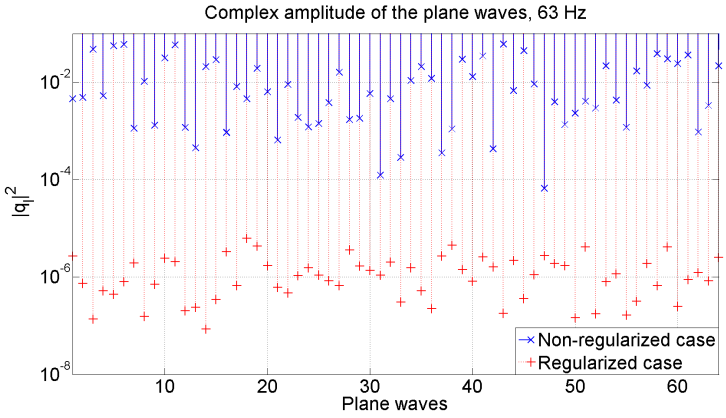


FIGURE 4.109: Complex amplitude of the PWE, Ightham Mote

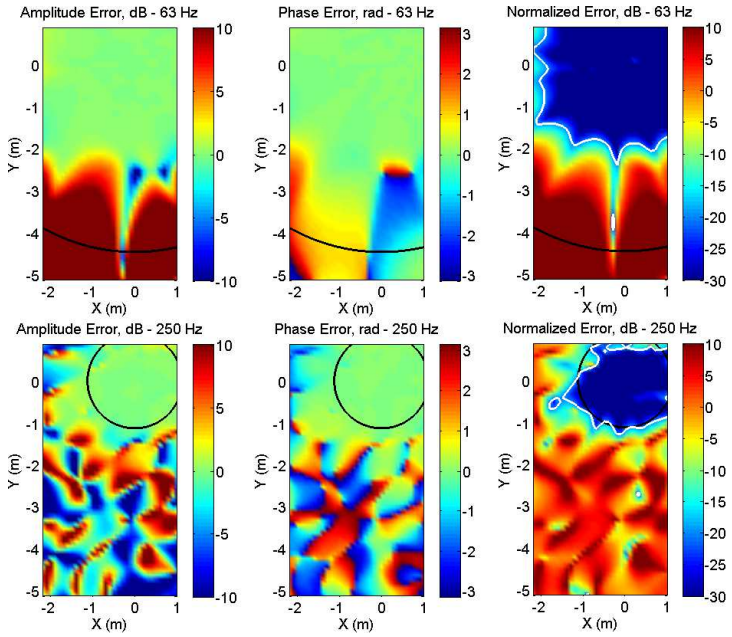


FIGURE 4.110: Acoustic errors, Ightham Mote

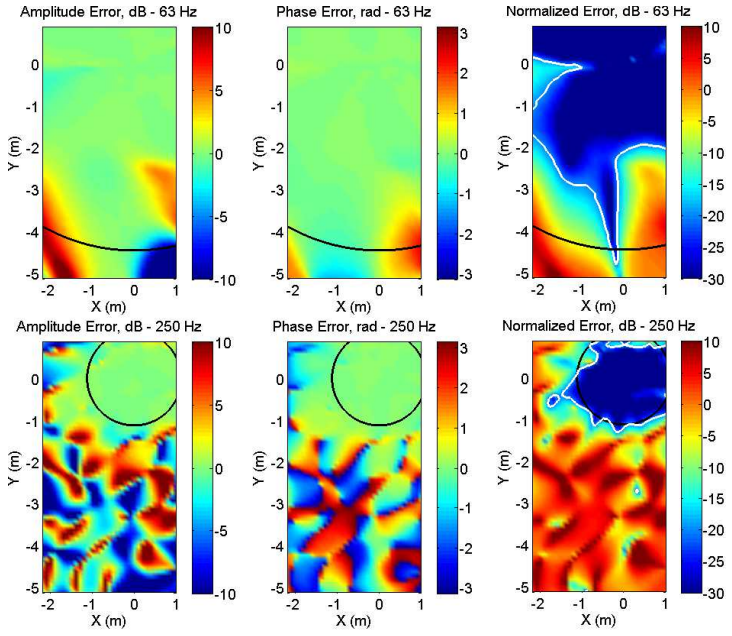


FIGURE 4.111: Acoustic errors, Ightham Mote, regularized $\beta = 0.05$, Ightham Mote

Figures 4.109, 4.110 and 4.111 reveal that for this specific enclosure, the application of regularization not only reduces the energy required to synthesize the sound field, but also leads to a more accurate sound field reconstruction in comparison with the

non-regularized pseudoinverse of the propagation matrix $\mathbf{H}(\omega)$. Table 4.5 shows the acoustic errors spatially-averaged over the cross-section illustrated in Figure 4.104.

Error	**Regularized**	**Non-regularized**
Amplitude (dB)	0.83	3.38
Phase (rad)	0.22	0.52
Normalized (dB)	-22.91	-14.41

TABLE 4.5: Spatially-averaged acoustic errors, Ightham Mote, 63 Hz

4.2.4 Office room

The office room is the smallest enclosure of the four spaces considered. The distance between the source and the receiver is approximately 3.8 m, which imposes a limitation on the assumption of plane wave propagation. Figure 4.112 illustrates the setup implemented for the prediction of the plane wave expansion. The central point of the PWE corresponds to the receiver used to calibrate the simulations (see Figure 3.12).

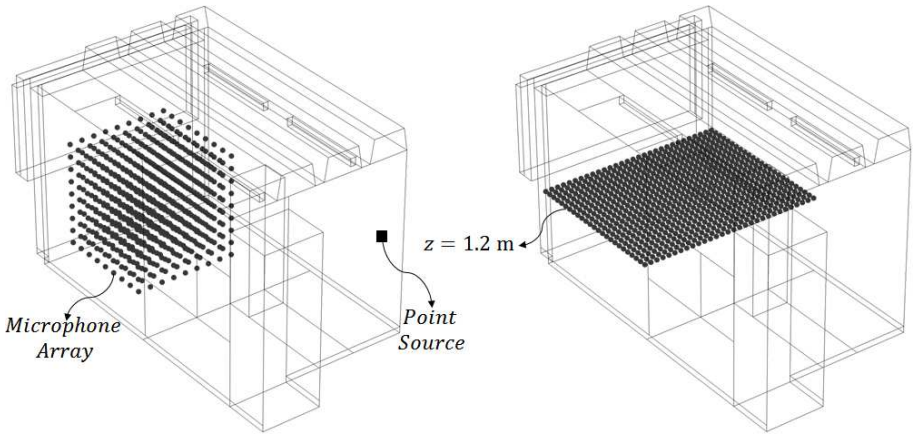


FIGURE 4.112: Sketch of the office room model, FEM

Figures 4.113 and 4.114 show the reconstruction of the sound field for the non-regularized and regularized cases at 63 Hz and 250 Hz, respectively.

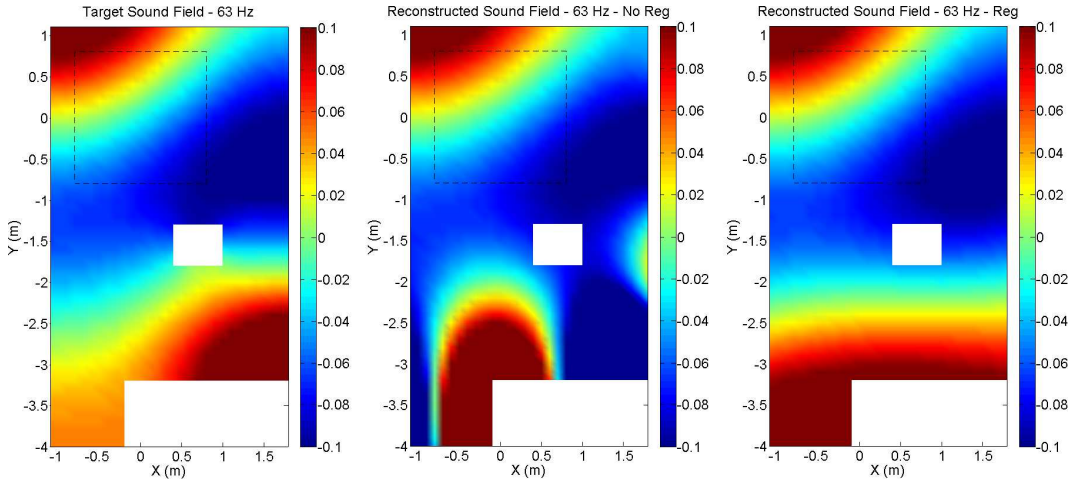


FIGURE 4.113: Target and reconstructed fields, office room, $\beta = 0.03$, 63 Hz

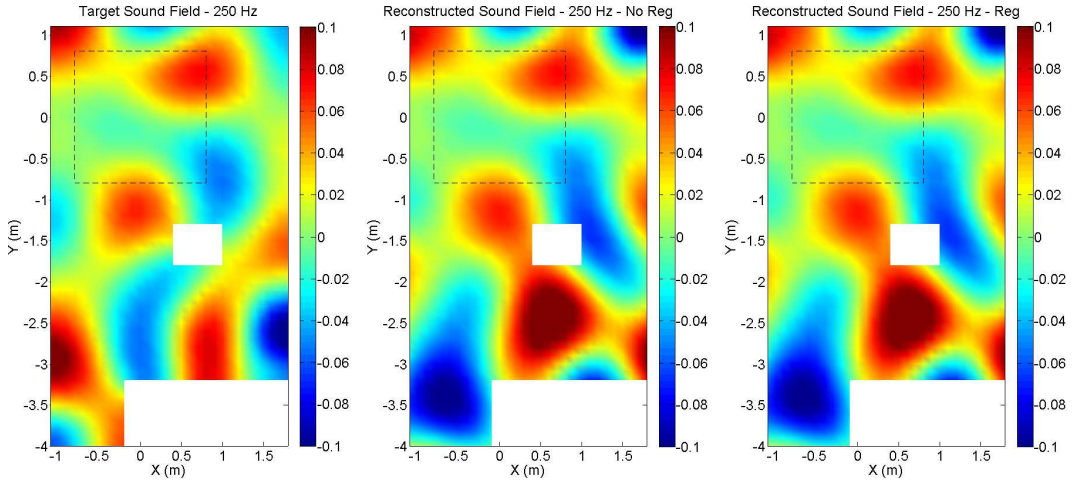


FIGURE 4.114: Target and reconstructed fields, office room, $\beta = 0.03$, 250 Hz

The outcomes indicate that the inverse method is a suitable approach to estimate a set of plane waves whose complex amplitudes reconstruct the sound field, even in small enclosures. At 63 Hz, the differences between the inversion with and without regularization are more significant further from the microphone array. As expected, there are not significant differences at 250 Hz between the two approaches. Figures 4.115, 4.116 and 4.117 illustrate the spatial distribution of the energy across the plane wave density and the comparison between the complex amplitudes of the plane waves for the non-regularized and regularized cases.

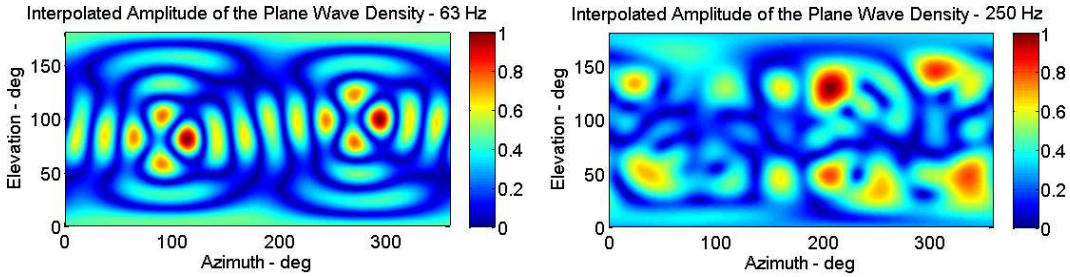


FIGURE 4.115: Normalized amplitude PWE, office room, $q_{total}(63)$=165.70, $q_{total}(250)$=8.0×10^{-3}

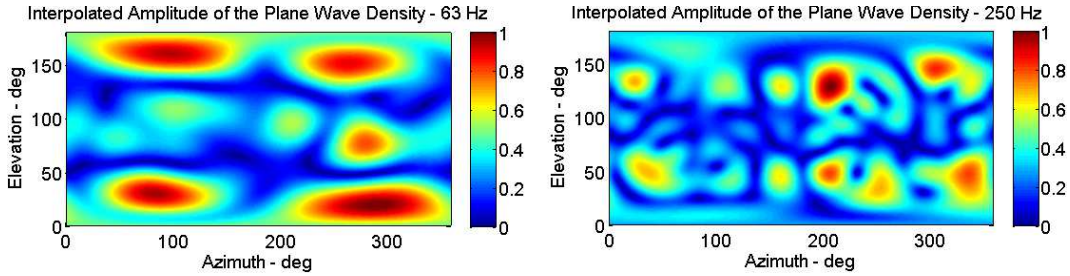


FIGURE 4.116: Normalized amplitude PWE, regularized $\beta = 0.03$, office room, $q_{total}(63)$=9.1×10^{-3}, $q_{total}(250)$=8.0×10^{-3}

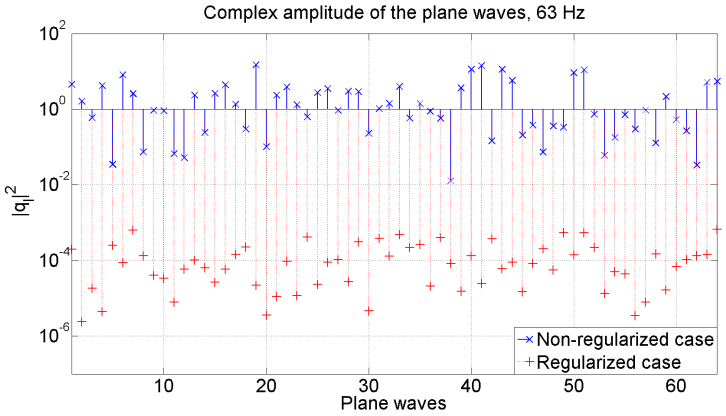


FIGURE 4.117: Complex amplitude of the PWE, office room

Figures 4.115 and 4.116 shows that there is a significant change in the spatial distribution of the energy of the plane wave density at 63 Hz. It seems that the inclusion of regularization in the inversion leads to a different spherical harmonic becoming predominant in the sound field reconstruction. Moreover, according to Figure 4.117, a reduction of up to 3 orders of magnitude in the complex amplitude of the plane waves was found when regularization is considered in the inverse problem. Figures

4.118, 4.119 and 4.120 illustrate the acoustic errors. Only the non-regularized case is displayed at 250 Hz due to the insignificant effect of the regularization at this frequency.

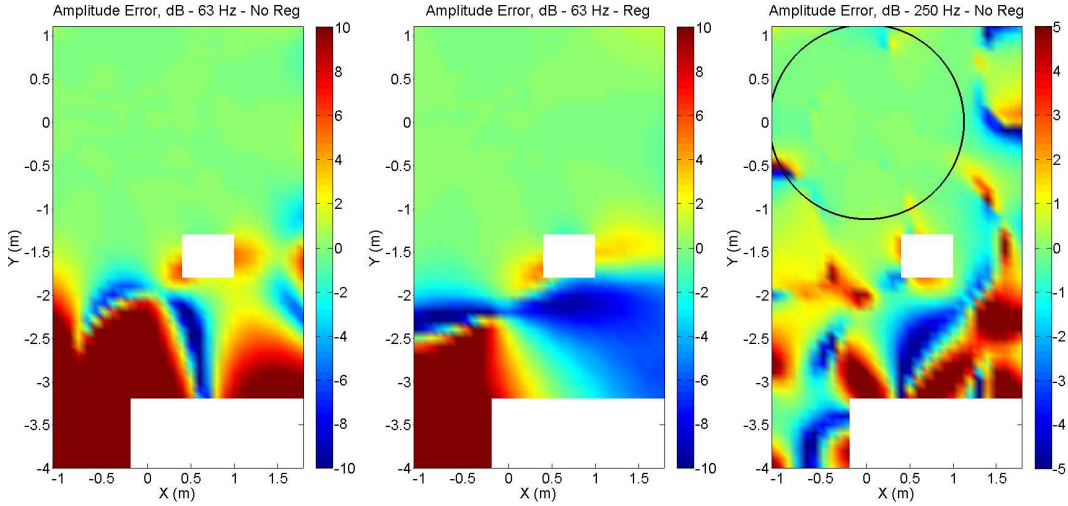


FIGURE 4.118: Amplitude error, $\beta = 0.03$, office room

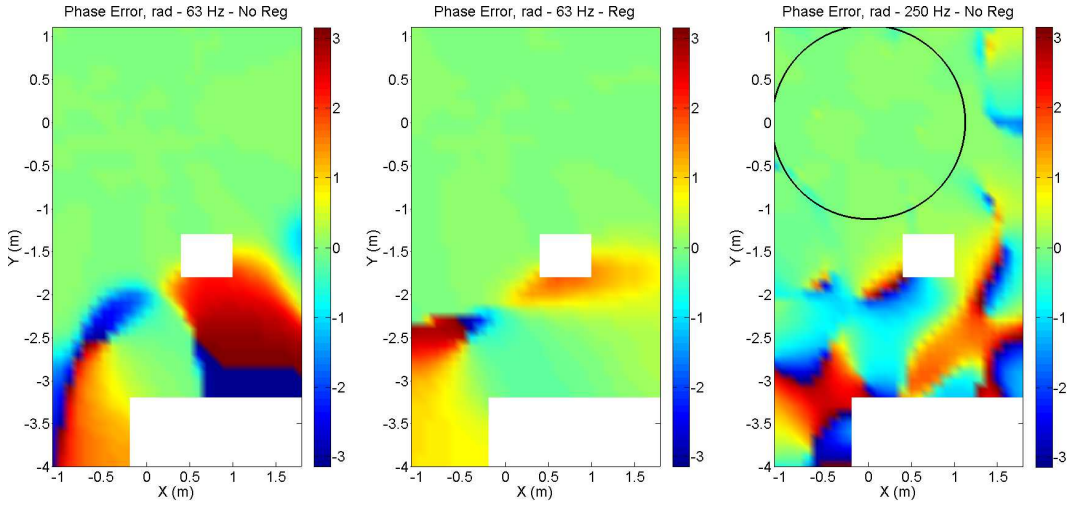


FIGURE 4.119: Phase error, $\beta = 0.03$, office room

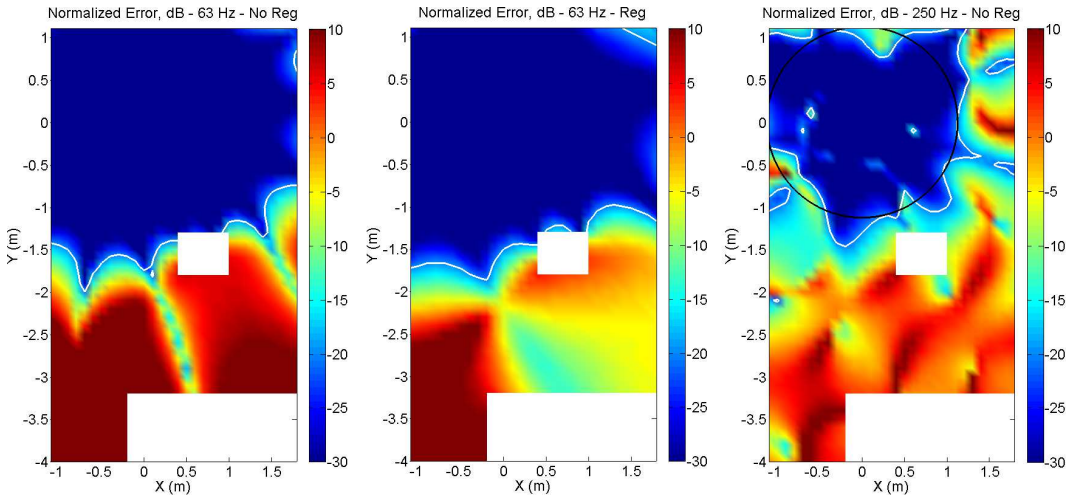


FIGURE 4.120: Normalized error, $\beta = 0.03$, office room

The figures are consistent with the results obtained for the meeting room. In this case, the use of regularization yields a reduction of the energy required to synthesize the sound fields, but the shrink in the area where the normalized error is smaller than -20 dB is not significant compared to the non-regularized case. The same trend is found for the amplitude and phase errors. These results imply that the use of regularization leads to a more efficient solution in terms of energy without highly compromising the accuracy of the sound field reconstruction for this specific enclosure.

The outcomes for the four reference cases indicate that the implementation of an inverse method provides a practical methodology to extract directional information from FE simulations. The results also show that the plane wave representation is a suitable approach even in small enclosures. The use of regularization leads to solutions that are more efficient in terms of the energy required to synthesize the sound fields, but its implementation may involve a reduction of the area where the reconstruction is accurate. Nevertheless, the shrink in this area was not significant in 3 of the 4 reference cases considered.

4.3 Encoding directional information from GA simulations

The aim of this section is to generate a plane wave expansion from the geometrical acoustics data that is compatible with the plane wave representation obtained from the FE results. This is done by assuming that the reflections produced by the GA model corresponds to plane waves propagating within the room. Different methods to extract the reflection patterns from the GA simulations are presented as follows. Although a histogram of reflections can be straightforwardly calculated when a GA simulation is performed, the use of a commercial package imposes limitations about the amount of information that can be accessed. Catt-Acoustics provides three alternatives regarding spatial data. The first one is an image source model able to estimate up to 9th order of reflections. The second one, corresponds to 3rd order Ambisonic impulse responses, and finally, the last one is the use of an external tool developed by Catt-Acoustics called ReflPhinder.

An image source model is expected to be different from a full impulse response calculation due to the assumption of only specular elements. The importance of diffuse reflections to generate the late components of the impulse response has been pointed out in the literature review. In addition, the outcome will be an echogram, whose phase has to be synthesized creating more uncertainty in the data. A standard simulation from Catt-Acoustics only estimates the first order of reflections using ISM, beyond that, the prediction is performed with a cone tracing algorithm. Figure 4.121 shows the comparison between an impulse response of an enclosure and its equivalent image source model up to 9th order. The phase from the ISM has been recreated using minimum phase filters. It is clear from the picture that an additional approach to recreate the tail of the impulse response is required.

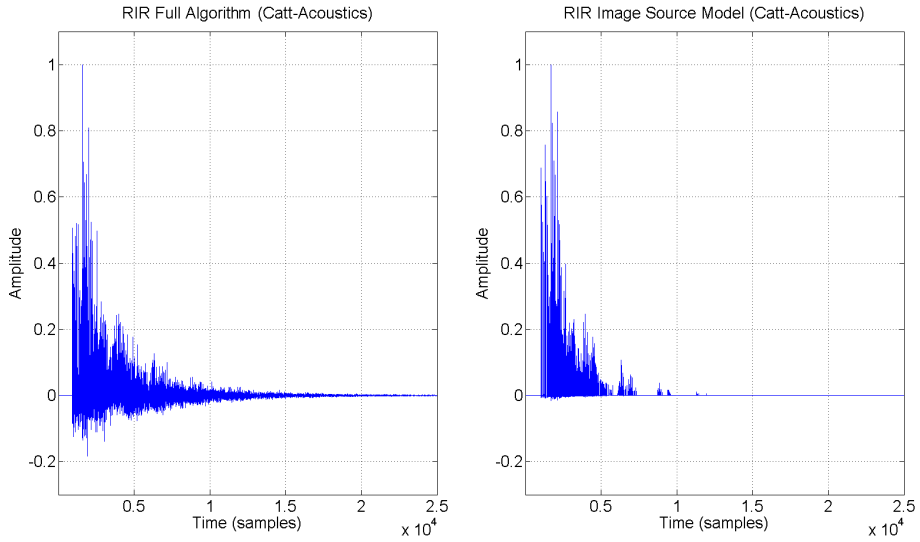


FIGURE 4.121: Comparison of RIR, full calcuation algorithm and ISM (Catt-Acoustics)

In contrast, the Ambisonic signals are computed using the full calculation algorithm but the outcomes are restricted by the use of only 16 coefficients. Compared to a PWE of 64 plane waves, the spatial resolution given by this approach is much lower, thus reducing the accuracy at high frequencies. Figure 4.122 illustrates the reconstruction of the acoustic field for a plane wave of 500 Hz using a 7th and 3rd order in the spherical harmonic series. The black circle represents the radius of validity estimated from equation (4.12).

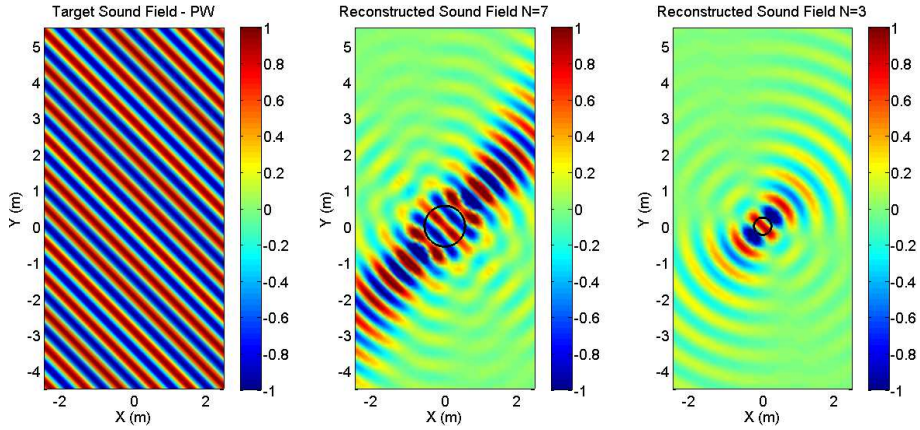


FIGURE 4.122: Comparison of the sound field reconstruction, 7th and 3rd order of spherical harmonics

Alternatively, ReflPhinder v.1.0c is a software developed by Catt-Acoustics that allows directional impulse responses to be created using a patented high-resolution virtual microphone (*sector mic*) [145]. The spatial analysis is based on measured or predicted b-format signals and on the implementation of the *sector mic* to sample a unit sphere. The algorithm creates a virtual microphone with a given directivity pattern covering a square window whose aperture can be selected according to different resolutions (See Figure 4.123). The output of the *sector mic* corresponds to a directional impulse response composed only of the contributions of the reflections whose direction agree with the directivity pattern for a specific azimuth and elevation angles. Nevertheless, the tessellation of the unit sphere is performed using a square window whose aperture does not change with the elevation angle. This generates redundant information due to the overlapping between windows at different angles (see Figure 4.124).

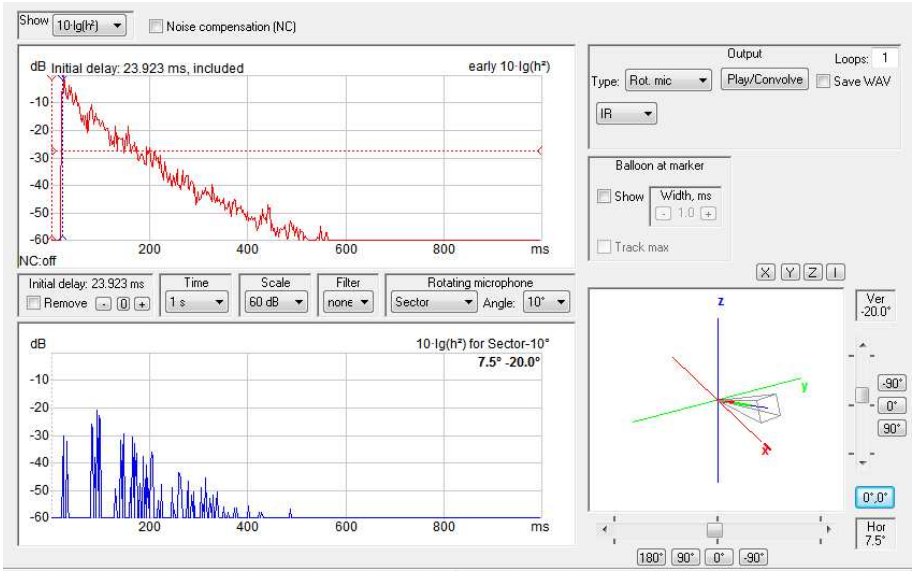


FIGURE 4.123: ReflPhinder commercial package, the tessellation of the sphere is done with a square window

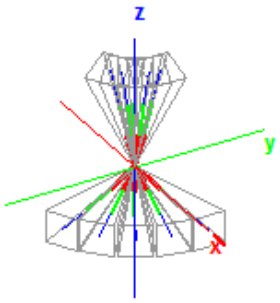


FIGURE 4.124: Overlapping produced by the use of the *sector mic*

The use of ReflPhinder seems to be the most suitable approach from the three possibilities discussed above. The results from the *sector mic* are based on the full calculation algorithm of Catt-Acoustics. In addition, the possibility of using a high spatial resolution window to sample the unit sphere benefits the sound field reconstruction at high frequencies. However, the errors produced by the tessellation of the sphere have to be taken into consideration. Therefore, ReflPhinder was implemented to extract directional information from Catt-Acoustics simulations. An angle of 10 degrees was used to sample the unit sphere yielding 614 directional impulse responses for each enclosure.

A rectangular enclosure with dimensions of (5m, 10m, 3m) was simulated to evaluate the overlapping produced by the tessellation of the sphere using the *sector mic*. The boundaries were characterized using absorption and scattering coefficients of 0.15 and 0.1, respectively. Figure 4.125 illustrates the location of the acoustic source A0 (4, 9.5, 1.5) and the receiver 01 (1, 2.5, 1.5) used for the simulations.

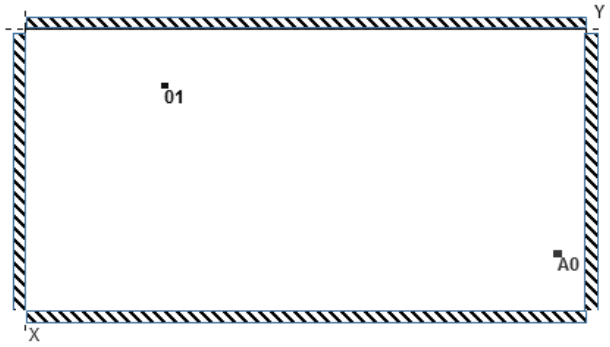


FIGURE 4.125: Sketch of the rectangular room, GA simualtion

Figure 4.126 shows the omnidirectional impulse response at the receiver 01 (a), the overlapped plots of the 614 estimated directional impulse responses (b) and their sum (c), which ideally should lead to the same omnidirectional information.

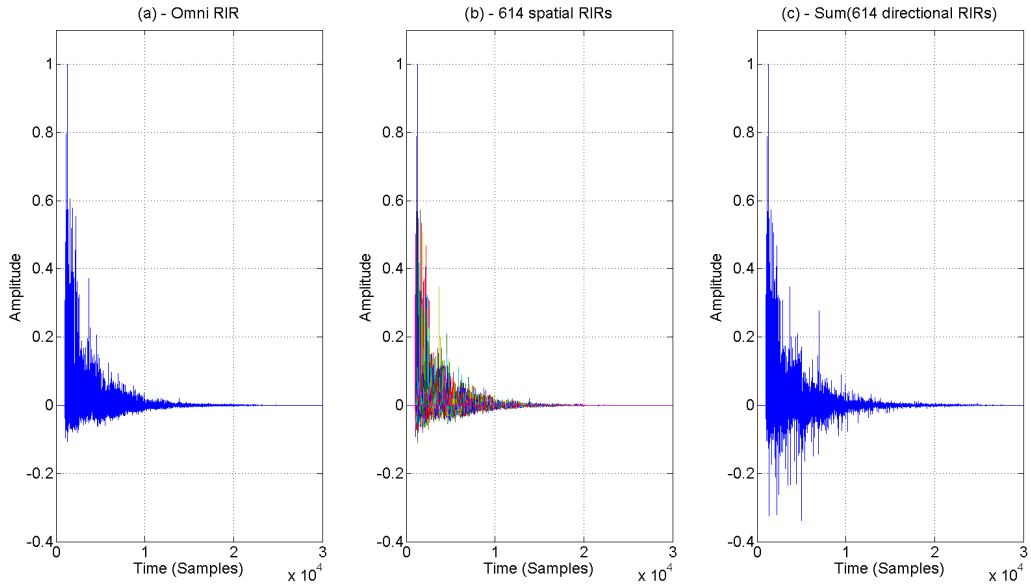


FIGURE 4.126: Comparison between omnidirectional and 614 directional impulse responses

Figure 4.126 shows that the envelope of the omnidirectional RIR (a) is similar to the envelope obtained from the overlapping of the 614 directional impulse responses (b), but their sum (c) does not resemble the omnidirectional RIR (a). The reason for this discrepancy is associated with the redundant information obtained from the *sector mic* when the unit sphere is tessellated.

An algorithm was developed to reduce the error produced by the use of the *sector mic*. For each time sample, the maximum of the 614 directional impulse responses was selected. Then, the scalar product between the unitary vector identifying the direction of the maximum and the unitary vectors of the remaining directional impulse responses was calculated. The directional impulse responses whose unitary vector leads to a scalar product less than a target value are fixed to zero at that sample. The target value is different for each elevation angle based on the areas where the overlapping is bigger (e.g. close to the poles of the unit sphere). The current maximum is discarded and the procedure is carried out again for the subsequent new maximum value. The loop remains until all the reflections at that sample have been

evaluated. Finally the reflections are scaled according to the ratio of the omnidirectional impulse response predicted at the origin of the plane wave expansion. This ensure that the sum of the directional impulse responses leads to the same information.

The implementation of the algorithm may reduce the low pass filter effect produced by the boundaries of the enclosures. This is because the cleaning procedure is conducted sample by sample. Therefore, the tail of a given reflection can be deleted by the cleaning procedure if a different reflection with more energy comes from a near direction. Nevertheless, the effect of this error is only relevant in the late part of the directional impulse responses where discrete reflections are not the main component and there is a possible overlapping of multiple reflections coming from closer directions. Figures 4.127 and 4.128 show the sum of the directional impulse responses (after the cleaning algorithm) and the mono impulse response in the frequency domain for narrow band and 1/3 octave bands, respectively. 0 dB corresponds to the acoustic pressure produced by a monopole source whose source strength generates 1 Pascal at 1 m of distance at the octave band of 1 kHz.

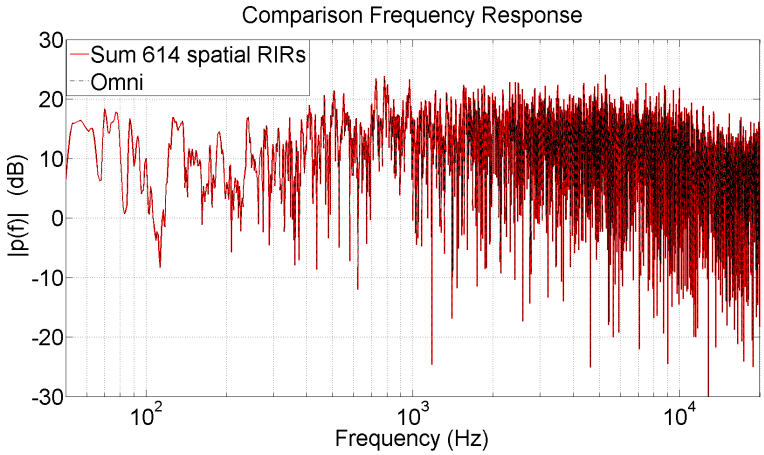


FIGURE 4.127: Comparison of room frequency responses (mono and sum of directionals)

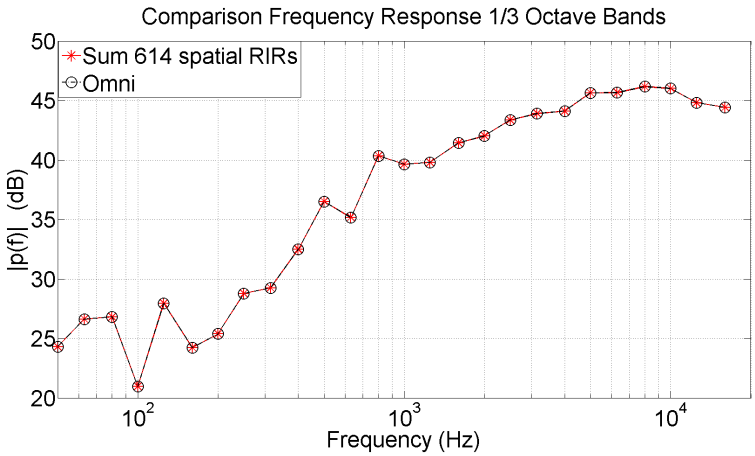


FIGURE 4.128: Comparison of room frequency responses 1/3 octave bands (mono and sum of directionals)

One important assumption made for the encoding of directional information from GA simulations is the consideration of plane waves as the only type of wave that contributes to the directional impulse responses. This assumption is reasonable for far field propagation; the relation between distance and frequency where far field can be considered is:

$$kr \gg 1, \tag{4.27}$$

where k is the wavenumber and r is the distance to the acoustic source. The direct path between the source and the receiver in the smallest enclosure (office room) is approximately 3.8 m leading to a frequency of 143 Hz considering at least one order of magnitude. It has been corroborated for all the enclosures that the GA data is used for a range of frequencies higher than the frequency calculated by equation (4.27). Furthermore, Catt-Acoustics generates scattering by changing the direction of the cone-rays according to a random number based on a directional distribution function. This implies that the distance covered by a scattered cone-ray is always larger than the direct path. However, if the plane wave propagation is not satisfied, the error will be given by the amount of translation (for an interactive auralization) and the distance to the source. The assumption of treating the reflections as plane waves is further analyzed in the next Chapter when a translation operator of the sound field is considered.

4.3.1 Resampling the plane wave expansion from 614 to 64 components

The spatial information extracted from GA simulations using ReflPhinder corresponds to 614 directional impulse responses. This information has to be transformed in order to generate a common ground with the plane wave expansion from the FE data. This means modifying the directional impulse responses according to the L possible directions given by the plane wave locations selected to recreate the sound field in the case of FEM (currently $L = 64$). Three different approaches were taken into consideration for this purpose: closest neighbour (CN), spherical harmonic interpolation (SH) and Vector Base Amplitude Panning (VBAP). It was verified that the sum of the reflections leads to the original frequency impulse response for all approaches. An example for the rectangular room is presented in Figures 4.129 and 4.130.

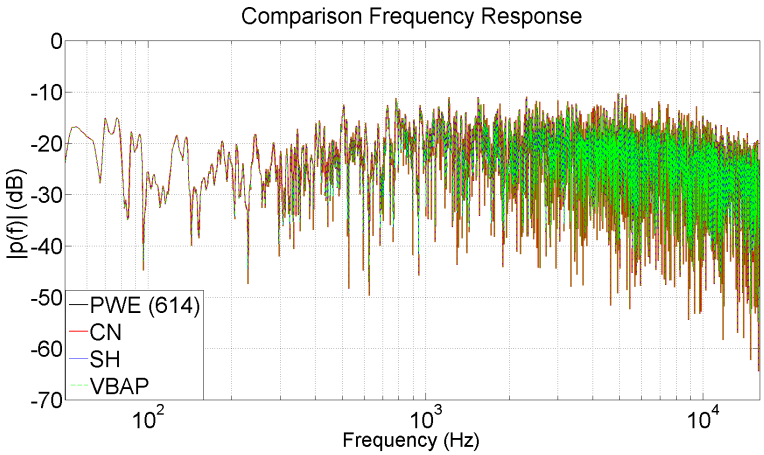


FIGURE 4.129: Comparison of room frequency responses (PWE of 614 elements and their representation into 64 plane waves using the closest neighbour, spherical harmonic interpolation and VBAP)

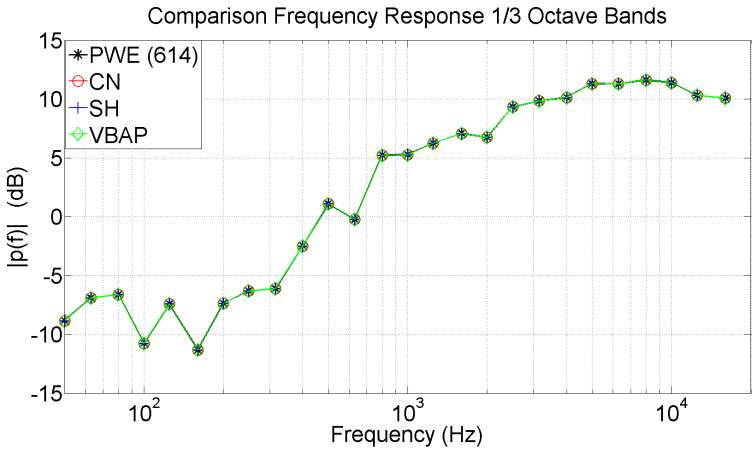


FIGURE 4.130: Comparison of room frequency responses 1/3 octave bands (PWE of 614 elements and their representation into 64 plane waves using the closest neighbour, spherical harmonic interpolation and VBAP)

4.3.1.1 Closest neighbour

This algorithm is based on a small modification in the direction of arrival of each reflection. The incoming direction of each of the 614 directional impulse responses computed with Reflphinder was changed to the closest direction among the 64 possible plane waves used for the FE data. The spatial sampling in the PWE corresponds to 64 plane waves uniformly distributed leading to an average angle of 25.2° between plane wave directions. This means that the translation of the incoming reflections is never greater than 12.6 degrees. Figure 4.131 illustrates a modified reflection path for a specific time sample (1272) using the closest neighbour approach. The circles identify the 64 possible directions given by the plane wave expansion and the colour their amplitudes. The magenta circle represents the original incoming direction of the reflection.

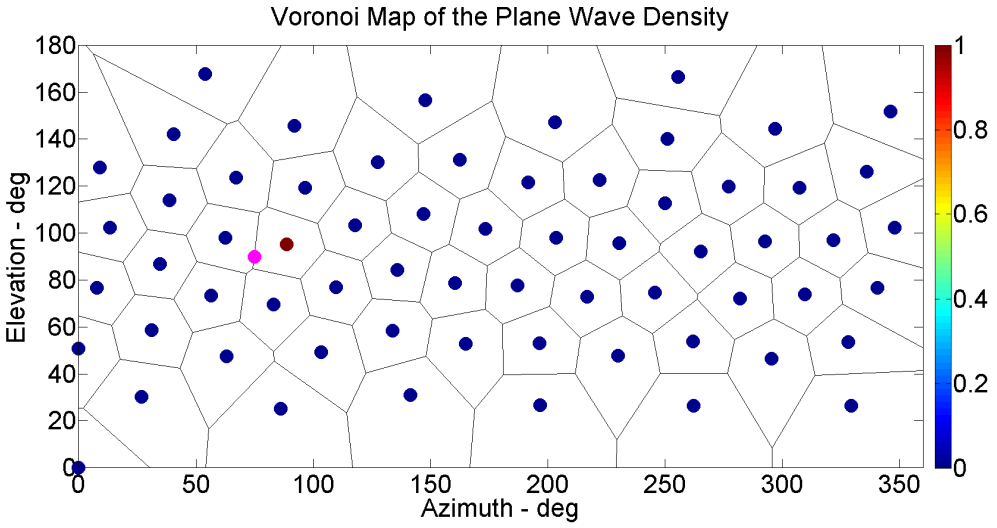


FIGURE 4.131: Reflection paths from GA, closest neighbour

4.3.1.2 Spherical harmonic interpolation

The Jacobi-Anger expansion allows a plane wave to be described in terms of spherical harmonics (equation (4.28)). Based on that, each reflection is encoded using 64 coefficients, which is equivalent to a 7th order in the expansion (equation (4.29)). Subsequently, the complex spherical harmonic coefficients are decoded into a finite set of plane waves following a mode-matching approach (equation (4.31)). The directions of the plane waves correspond to the 64 directions used to process the FE data.

$$e^{jk\mathbf{x}\cdot\hat{\mathbf{y}}} = 4\pi \sum_{n=0}^{\infty} j^n j_n(kr_x) \sum_{m=-n}^{n} Y_n^m(\theta_x, \phi_x) Y_n^m(\theta_y, \phi_y)^*, \tag{4.28}$$

$$4\pi \sum_{n=0}^{N=7} j^n j_n(kr_x) \sum_{m=-n}^{n} Y_n^m(\theta_x, \phi_x) Y_n^m(\theta_y, \phi_y)^* = 4\pi \sum_{l=1}^{L} q_l \sum_{n=0}^{N=7} j^n j_n(kr_x) \sum_{m=-n}^{n} Y_n^m(\theta_x, \phi_x) Y_n^m(\theta_l, \phi_l)^*. \tag{4.29}$$

The simplification of equation (4.29) using the orthogonality relation of the spherical harmonics yields to the following mode matching equation for each n and m

$$Y_n^m(\theta_y, \phi_y)^* = \sum_{l=1}^{L} q_l Y_n^m(\theta_l, \phi_l)^*, \tag{4.30}$$

for $n = 0 \dots N$ and $|m| \leq n$. This finite set of linear equation are solved in terms of the least squared solution by applying an inverse method. The formulation for the 614 plane waves corresponds to

$$\tilde{\mathbf{q}} = \mathbf{C}^{\dagger}\mathbf{A}\mathbf{q}, \tag{4.31}$$

in which

$$\mathbf{C} = \begin{vmatrix} Y_{00}(\theta_1, \phi_1)^* & \cdots & Y_{00}(\theta_L, \phi_L)^* \\ \vdots & Y_{nn}(\theta_l, \phi_l)^* & \vdots \\ Y_{NN}(\theta_1, \phi_1)^* & \cdots & Y_{NN}(\theta_L, \phi_L)^* \end{vmatrix}$$

,

$\mathbf{A} = [\mathbf{a}_1 \cdots \mathbf{a}_i]$, $\mathbf{a}_i = \left[Y_0^0(\theta_i, \phi_i)^* \cdots Y_N^N(\theta_i, \phi_i)^*\right]^T$, $i = 614$, $\mathbf{q} = [q_1 \cdots q_i]^T$ and $\tilde{\mathbf{q}} = [q_1 \cdots q_L]^T$. Figure 4.132 illustrates a modified reflection path for the same specific time sample (1272) using the spherical harmonic interpolation approach.

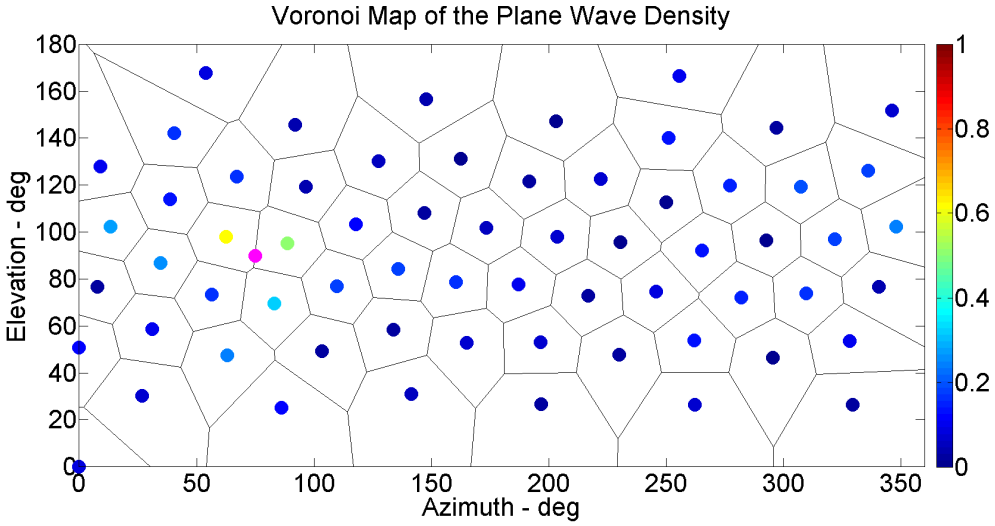


FIGURE 4.132: Reflection paths from GA, spherical harmonics

4.3.1.3 Vector base amplitude panning

Vector Base Amplitude Panning (VBAP) is a sound reproduction technique based on the formulation of amplitude panning functions as vectors and vector basis [118]. It allows the incoming direction of waves to be controlled over a unit sphere. For 3D sound reproduction, the three loudspeakers closest to the target incoming direction are selected to reconstruct the sound field. The gain of each loudspeaker is estimated by means of an inverse method. A similar concept can be implemented to synthesize the directional impulse responses extracted with ReflPhinder. A linear combination of three plane waves whose directions are restricted to 64 possible directions (the PWE used to process the FE data) are selected to reconstruct each of the 614 directional impulse responses.

For each directional impulse response generated by ReflPhinder, the three closest plane waves in terms of the direction of propagation are selected. This is done by the calculation of the "distance" between the unitary vector identifying the directional impulse response and each unitary vector of the PWE. Then, an inverse problem is formulated as follows: The direction of the three plane waves selected to synthesize the directional impulse response are established as unit vectors ($\mathbf{l}_n$) whose origin is the centre of a unit sphere. The direction of the directional impulse response is defined by the unit vector $\tilde{\mathbf{l}}$. By expressing $\tilde{\mathbf{l}}$ as a combination of the three plane waves the following relation is determined

$$\tilde{\mathbf{l}} = (g_1\mathbf{l}_1 + g_2\mathbf{l}_2 + g_1\mathbf{l}_3)\,, \tag{4.32}$$

in which g_n represents the amplitude of the plane waves. A matrix notation yields

$$\tilde{\mathbf{l}} = \mathbf{L}\mathbf{g}, \tag{4.33}$$

where $\mathbf{L} = [\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3]$ and $\mathbf{g} = [g_1, g_2, g_3]^T$. The amplitudes are calculated by solving equation (4.33) for $\mathbf{g}$, namely

$$\mathbf{g} = \mathbf{L}^{-1}\tilde{\mathbf{l}}. \tag{4.34}$$

In addition, the amplitudes are normalized based on a coherent summation where their sum leads to unity, namely

$$\mathbf{g_{normalized}} = \frac{\mathbf{g}}{g_1 + g_2 + g_3}, \tag{4.35}$$

Subsequently, the directional impulse response is multiplied by the normalized amplitudes of the three plane waves that are used to synthesize the sound field, which leads to a three new directional impulse responses. This procedure is carried out for the 614 plane waves computed with ReflPhinder. This means that the number of directional impulse responses at this stage corresponds to a 1842 (614×3), but each of them corresponds to a given direction established by the PWE. Finally, the directional impulse responses of the same direction are added together yielding 64 directional impulse responses. Figure 4.132 illustrates a modified reflection path for the same specific time sample (1272) using the VBAP approach.

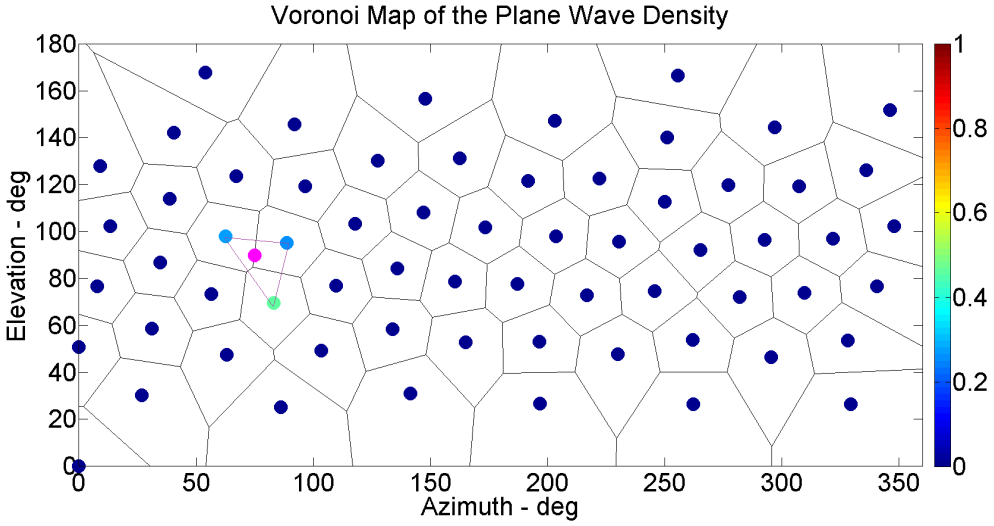


FIGURE 4.133: Reflection paths from GA, VBAP

4.3.1.4 Comparison of the closest neighbour, spherical harmonic interpolation and VBAP as approaches to resample the PWE

In the case of the spherical harmonic interpolation, the reflections are spread in all possible directions of the PWE, which is a consequence of the truncation of the series. This result can be explained using the analogy of a Dirac Delta function the complex coefficients of which have been truncated using a frequency window. By applying an inverse Fourier transform, the output corresponds to a Sinc function meaning that the energy has been distributed in time. In the same way, the energy of a plane wave is distributed in all the possible directions of the PWE when the spherical harmonic

series is truncated. This outcome may constrain the application of a translation operator because the side lobes can generate a precedence effect when the sound field is translated. Finally, VBAP preserves a more compact representation and the original direction of the reflection by using only the three closest plane waves of the PWE. However, the interference pattern generated by the interaction of the three plane waves may constrain the accuracy of a translation operator. Further experiments regarding the accuracy of these three different approaches are presented in Chapter 5 where the translation of the acoustic field is addressed.

4.4 Combination of the FE and GA data

Sections 4.2 and 4.3 provide a framework for the generation of a plane wave expansion to represent acoustic fields that have been predicted by means of the finite element method and geometrical acoustics. This is done by implementing an inverse method for the FE data and by tracing the reflections in the case of GA. The outcomes correspond to a plane wave expansion of L elements (currently, $L = 64$) for each approach. These two plane wave representations must be combined in order to have a unified PWE, which describes the acoustic field in a broader frequency band. This is performed based on the approach described in section 3.4.

The combination of the plane wave expansions is carried out in the frequency domain by applying 8th order Butterworth filters. Each element of the PWE computed from the FE data is filtered using a low-pass filter. Similarly, the directional impulse responses of the PWE from GA are transformed to the frequency domain by applying a Fourier transform. Then, the signals are filtered using a high-pass filter. The crossover frequency for the filters depends on the maximum frequency simulated in FEM for each enclosure. The central frequency of the 1/3 octave band previous to the maximum frequency simulated is selected as the crossover frequency. This ensures that the FE data is not abruptly truncated. The unified directional impulse responses are computed by adding the filtered plane wave expansions and by applying an inverse Fourier transform. The final outcome corresponds to L directional impulse responses, which can be processed to generate interactive auralizations.

Chapter 5

Sound Field Manipulation Based on the Plane Wave Expansion

The aim of this chapter is to generate a framework for the sound field manipulation that enables the listener to move in the environment in terms of listener translation and rotation. This is carried out by taking advantage of the plane wave representation presented in the previous Chapter. The translation and rotation of the listener is recreated by the relative translation and rotation of the sound field. For this, a sound field translation operator in the plane wave domain is introduced, which enables the listener to move within the enclosure. An analysis of the area in which the translation of the plane wave expansion synthesizes correctly a target field is presented. In addition, an evaluation of the accuracy of the methods proposed in section 4.3.1 to resample the plane wave expansion from GA data is presented by means of the translation operator. Then, two approaches are investigated for the rotation of the sound field. The first corresponds to a spherical harmonic transformation based on the Jacobi-Anger expansion (equation (2.56)) and the second is the use of VBAP as an interpolation method in the plane wave domain. Results on the suitability and accuracy of the two proposed approaches to rotate acoustic fields are discussed.

5.1 Translation of the sound field

The interactive feature that corresponds to the translation of the listener within the enclosure is addressed in this section. The analysis is based on the representation of an acoustic field by means of a plane wave expansion. The translation is achieved

by the implementation of a sound field translation operator that modifies the phase of the complex amplitudes of the plane waves according to the relative position of the listener. Although the translation is not produced at the listener per se, its equivalence with the translation of the sound field allows for the interactive synthesis of the acoustic field. For example, a listener translation to the right corresponds to a sound field translation to the left by the same amount. The derivation of the translation operator is conducted in the frequency domain due to its simplicity. Nevertheless, its time domain equivalent, which is implemented in the auralization system, is also considered. Figure 5.1 illustrates the vector $\mathbf{x}^{'}$, which identifies the origin of a relative coordinate system corresponding to the centre of the listener's head, such that the relative vector $\mathbf{x}_{\text{rel}}$ correponds the same point in space as that identified by the vector $\mathbf{x}$ in absolute coordinates. Therefore $\mathbf{x}_{\text{rel}}$ is given by $\mathbf{x}_{\text{rel}} = \mathbf{x} - \mathbf{x}^{'}$.

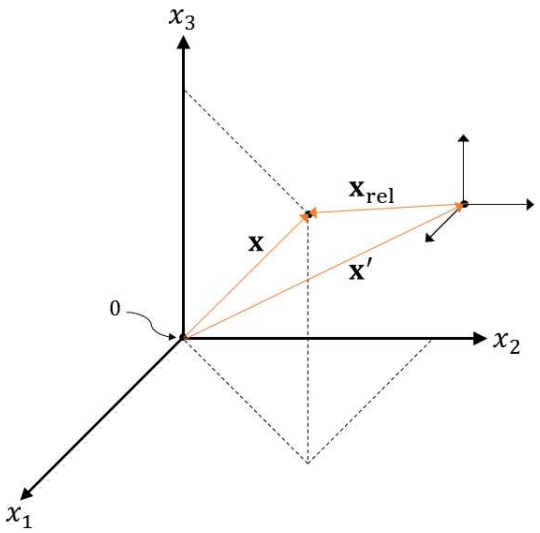


FIGURE 5.1: Vector $\mathbf{x}$ is represented as $\mathbf{x}_{\text{rel}}$ in the relative coordinate system with origin at $\mathbf{x}^{'}$

The sound field translation operator is derived from considering two plane wave expansions of the same sound field but centred at different points in space, specifically, at the origin of the absolute and relative coordinate systems. In this case, the difference between the two plane wave expansions is given by the plane wave amplitude densities $q(\hat{\mathbf{y}}, \omega)$ and $q_{\text{rel}}(\hat{\mathbf{y}}, \omega)$, respectively. The plane wave expansion of the sound field p (section 4.1.1) centred at the origin of the absolute coordinate system is defined as

$$p(\mathbf{x}) = \int_{\widehat{\mathbf{y}}\in\Omega} e^{jk\mathbf{x}\cdot\widehat{\mathbf{y}}} q(\widehat{\mathbf{y}}) d\Omega(\widehat{\mathbf{y}}), \tag{5.1}$$

where Ω is the unitary sphere. Similarly, the plane wave expansion of the same field p but centred at the origin of the relative coordinate system is expressed as

$$p(\mathbf{x}_{\text{rel}}) = \int_{\widehat{\mathbf{y}}\in\Omega} e^{jk\mathbf{x}_{\text{rel}}\cdot\widehat{\mathbf{y}}} q_{\text{rel}}(\widehat{\mathbf{y}}) d\Omega(\widehat{\mathbf{y}}). \tag{5.2}$$

It is important to draw attention at this point to the fact that the plane wave densities $q(\hat{\mathbf{y}}, \omega)$ and $q_{\text{rel}}(\hat{\mathbf{y}}, \omega)$ are different functions. Therefore, the objective is to express one density in terms of the other. This is achieved by expanding $\mathbf{x}_{\text{rel}}$ as $\mathbf{x} - \mathbf{x}^{'}$, namely

$$p(\mathbf{x}) = \int_{\widehat{\mathbf{y}}\in\Omega} e^{jk\mathbf{x}\cdot\widehat{\mathbf{y}}} e^{-jk\mathbf{x}^{'}\cdot\widehat{\mathbf{y}}} q_{\text{rel}}(\widehat{\mathbf{y}}) d\Omega(\widehat{\mathbf{y}}). \tag{5.3}$$

The two representations of $p(\mathbf{x}, \omega)$ given by equations (5.1) and (5.3) are equivalent expansions of the same field. Equating these two equations yields

$$\int_{\widehat{\mathbf{y}}\in\Omega} e^{jk\mathbf{x}\cdot\widehat{\mathbf{y}}} q(\widehat{\mathbf{y}}) d\Omega(\widehat{\mathbf{y}}) = \int_{\widehat{\mathbf{y}}\in\Omega} e^{jk\mathbf{x}\cdot\widehat{\mathbf{y}}} e^{-jk\mathbf{x}^{'}\cdot\widehat{\mathbf{y}}} q_{\text{rel}}(\widehat{\mathbf{y}}) d\Omega(\widehat{\mathbf{y}}), \tag{5.4}$$

whose solution for $q_{\text{rel}}(\hat{\mathbf{y}}, \omega)$ is given by

$$q_{\text{rel}}(\widehat{\mathbf{y}}, \omega) = q(\widehat{\mathbf{y}}, \omega) e^{jk\mathbf{x}^{'}\cdot\widehat{\mathbf{y}}}. \tag{5.5}$$

Equation (5.5) indicates that the plane wave density $q_{\text{rel}}(\widehat{\mathbf{y}}, \omega)$ can be predicted by taking the product between the plane wave density function of the plane wave expansion centred at the origin $q(\widehat{\mathbf{y}}, \omega)$ and a complex exponential whose argument depends on the vector $\mathbf{x}^{'}$. In other words, $e^{jk\mathbf{x}^{'}\cdot\widehat{\mathbf{y}}}$ is the translation operator for the plane wave expansion from the origin to $\mathbf{x}^{'}$. Its equivalence in the time domain can be easily found by using the shifting property of the Fourier transform [14]

$$\int_{-\infty}^{\infty} F(t - t^{'}) e^{-j\omega t} = f(\omega) e^{-j\omega t^{'}}. \tag{5.6}$$

Equation (5.6) indicates that the product between a complex function $f(\omega)$ and a complex exponential is equivalent to shifting the function $F(t)$ according to the

argument of the complex exponential. This means that the translation operator basically corresponds to the application of delays, namely

$$Q_{\text{rel}}(\widehat{\mathbf{y}}, t) = Q\left(\widehat{\mathbf{y}}, t - \frac{\mathbf{x}^{'} \cdot \widehat{\mathbf{y}}}{c}\right). \tag{5.7}$$

A plane wave propagation at 250 Hz coming from $\theta = 90°$ and $\phi = 45°$ was analytically synthesized in a free field region with dimensions of (5m, 10m, 3m). Then, an inverse method based on the methodology described in section 4.1.1.1 was implemented to estimate the complex amplitude of $L = 64$ plane waves, which are used to reconstruct this target field. The discretized version of the plane wave expansion is mathematically expressed as

$$p(\mathbf{x}) = \sum_{l=1}^{L} e^{jk\mathbf{x}\cdot\widehat{\mathbf{y}}_l} q(\widehat{\mathbf{y}}_l)\Delta\Omega_l, \tag{5.8}$$

where $\Delta\Omega_l$ is the area attributed to each direction $\widehat{\mathbf{y}}_l$ as result of the discretization of the integral representation. An evaluation of the sound field translation operator is presented in Figure 5.2. This is performed by the comparison between the real part of the target (analytical) and the reconstructed acoustic pressure (Pa) in a cross-section of the domain (z = 1.5m). (a) corresponds to the reference target sound field, (b) is the reconstructed sound field by means of the plane wave expansion ($L = 64$), (c) is the translated target sound field and (d) is the reconstructed and translated sound field. The initial reference point is plotted in cyan whilst the magenta point identifies the location of the desired translation. For the plane wave expansion these points represent the initial and translated positions where the plane wave expansion was centred. Finally, the amplitude, phase and normalized errors defined in section 4.1.1.2 are shown in Figure 5.3. The black circle corresponds the area of accurate reconstruction predicted by equation (4.12).

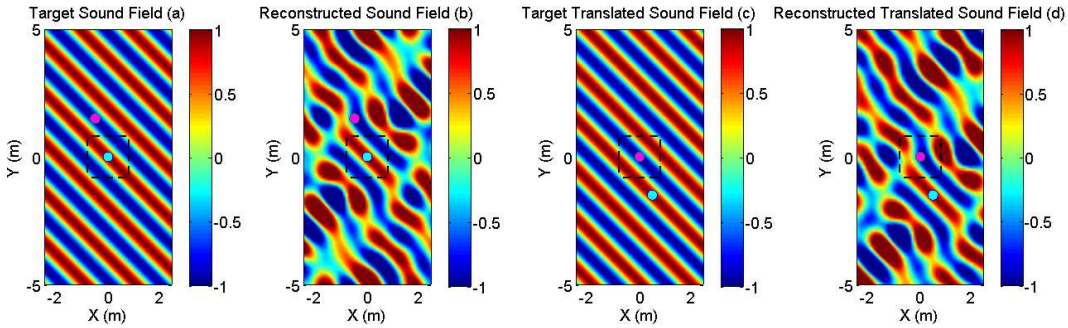


FIGURE 5.2: Translation of the sound field in the plane wave domain

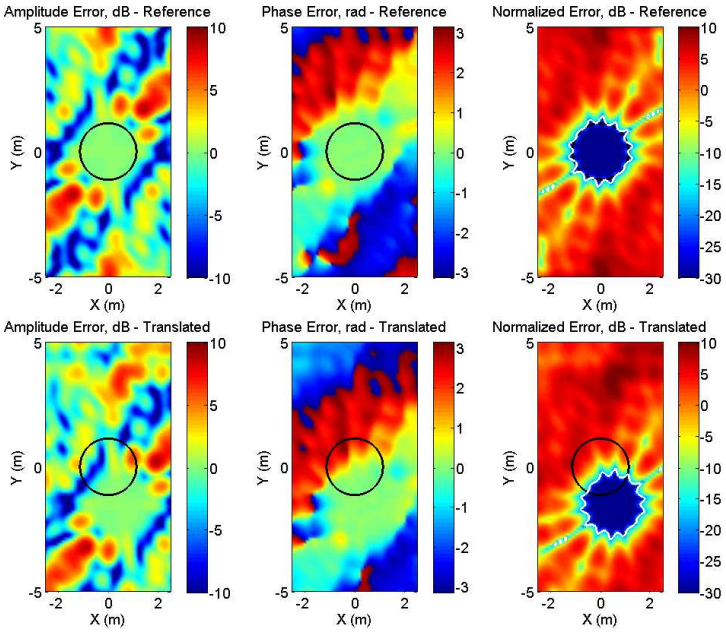


FIGURE 5.3: Acoustic errors, translation of the sound field in the plane wave domain

The reconstruction of the sound field will be exact to the target field if an integral plane wave expansion (equation (5.1)) is used for the synthesis. This outcome means that the translation operator will lead to the correct field regardless of the location where the translation is intended. However, if the plane wave expansion is approximated with a series, as in equation (5.8), the reconstructed field will contain errors and the translation operator will lead to the correct field only in the area where the discretized plane wave expansion matches the target field. In other words, the constraint in the area where the translation leads to the target field is not due to the translation operator itself, but due to the initial errors given by the discretization of the plane wave expansion. This is corroborated by Figures 5.2 and 5.3 where it

is illustrated that the translation operator affects the complete acoustic field. For auralization applications, the translation should be bounded by the area where the reconstruction is accurate subtracted by a listener area to ensure an accurate synthesis of the sound field. An insight of the amount of translation that can be correctly applied may be estimated from equation (4.12) as

$$r_t = \left(\frac{(\sqrt{L} - 1)\lambda}{e\pi} \right) - r_l. \tag{5.9}$$

where r_t is the effective translation radius, L is the number of plane waves used for the synthesis, λ is the wavelength and r_l is the radius of the listener (e.g 0.1 m) where it is desirable that the sound field is reproduced accurately. r_l (radius of the listener) has to be taken into consideration in order to preserve the binaural cues. An example of the translation radius is given in Figure 5.4, in which $R(\omega) = \frac{(\sqrt{L}-1)\lambda}{e\pi}$.

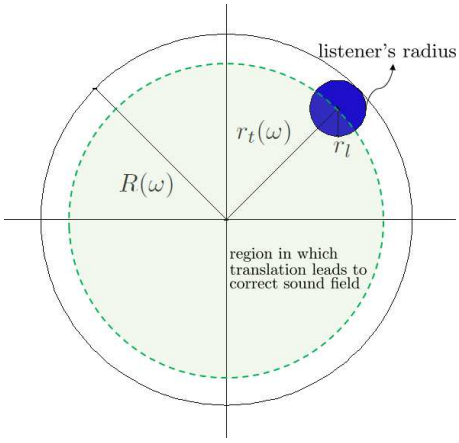


FIGURE 5.4: Region of accurate translation given by the PWE

Equation (5.9) provides an indication of the radius of the region where the translation leads to an accurate reconstruction of the target field. However, it is important to emphasize that the translation is performed correctly in the whole domain. The mismatch with respect to the target field is not due to the translation operator itself, but to the initial error produced by the discretization of the plane wave expansion. In addition, equation (5.9) assumes an interior formulation where there are no sources, scattering objects or reflective surfaces inside the reconstruction area. When representing the sound field in the interior of a room at low frequencies, the predicted radius r_t can be larger than the dimensions of the enclosure, thus limiting the area of accurate reconstruction. Furthermore, a smaller radius than that predicted by

equation (5.9) is obtained at low frequencies if regularization is implemented in the inversion of the propagation matrix $\mathbf{H}(\omega)$ (see section 4.1.1.1).

5.1.1 Evaluation of the plane wave expansion from the geometrical acoustic data using the translation operator

A method to generate directional impulse responses from geometrical acoustic simulations has been described in section 4.3. The number of directional impulse responses corresponds to 614 when the sector mic is selected to discretize the unit sphere. In this section, an evaluation of the accuracy given by the assumption of treating these directional impulse responses as plane waves (i.e. a PWE of 614 elements) is presented. This is carried out by comparing omnidirectional impulse responses simulated in Catt-Acoustics with the field predicted at the same locations by the translation of the plane wave expansion. It is relevant to emphasize that the algorithm proposed in section 4.3 to minimize the error due to overlapped tessellation of the unit sphere ensures that the acoustic pressure is accurately reconstructed at the central point of the expansion.

Firstly, a simple model based on one single reflector was tested. The calculation was performed using an absorption and scattering coefficient of 0.1 and 0.2, respectively. Because the stochastic implementation of the scattering coefficients in Catt-Acoustics, an exact reconstruction is only achieved at the central point of the expansion; beyond that, only an agreement with respect to the omnidirectional references in terms of energy is expected. 10 receivers distributed over a radius of 1 and 2 m from the central point of the expansion have been simulated (see Figure 5.5). A0 identifies the source position and 01 corresponds to the central point of the expansion. The mean error displayed in the figures was selected as a metric and it is defined as:

$$\mathrm{ME(dB)} = \frac{1}{n}\sum_{i=1}^{n}\left|10\log_{10}(|\widetilde{p}_i|^2) - 10\log_{10}(|p_i|^2)\right|, \tag{5.10}$$

in which n is the number of 1/3 octave frequency bands, $|\widetilde{p}_i|^2$ and $|p_i|^2$ are the predicted and reference energy of the acoustic pressure in the 1/3 octave band i, respectively. This error considers an equal contribution of all the 1/3 octave bands, thus being similar to a model in which pink noise is used as input signal. It was selected to provide an insight of how dissimilar on average the reconstructed field

is from the reference one. An analysis of the geometrical acoustic data in terms of phase is not presented for the to following reasons:

- Catt-Acoustics is an energy-based software in which the propagation of the cone-rays does not contain information about phase. The impulse responses are created after the calculation based on echograms and by artificially synthesizing the phase using linear and minimum phase filters.
- GA is only used at mid and high frequencies when the plane wave expansion is combined with FE results. The assumption that GA data is most likely random-phase is reasonable in this frequency range.

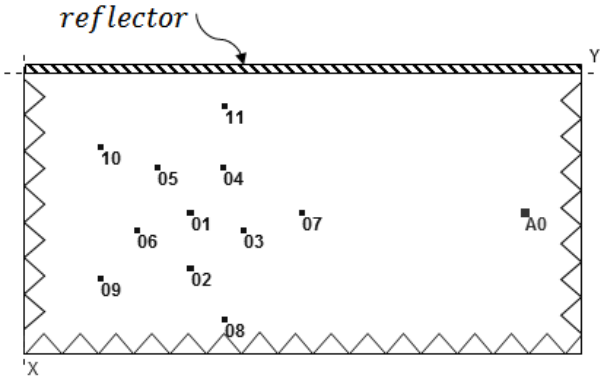


FIGURE 5.5: Sketch of the one reflector model used to evaluate the plane wave expansion from GA simulations

A translation of the PWE to the positions of the receivers was applied to compare the agreement between the translated acoustic field and the omnidirectional room impulse responses predicted by Catt-Acoustics. Figures 5.6 to 5.16 show the frequency response in narrow band (left side of the figures) and in 1/3 octave band resolution (right side of the figures). 0 dB corresponds to the acoustic pressure produced by a monopole source whose source strength generates 1 Pascal at 1 m of distance in the octave band of 1 kHz.

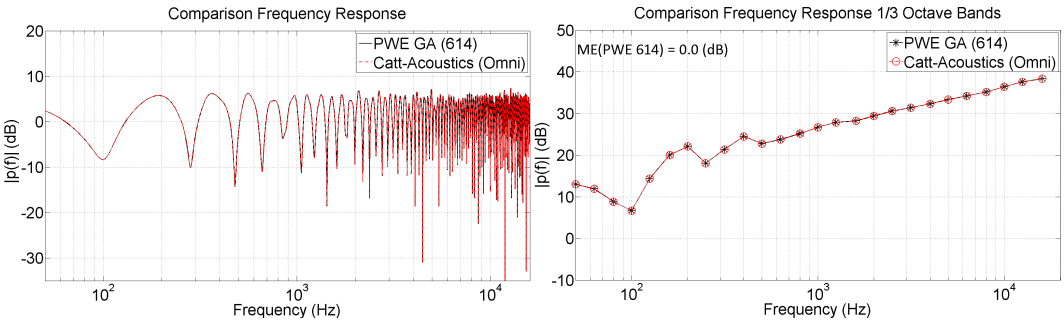


FIGURE 5.6: Comparison of frequency responses. PWE (GA) and omnidirectional RIR at the reference position 1, single wall

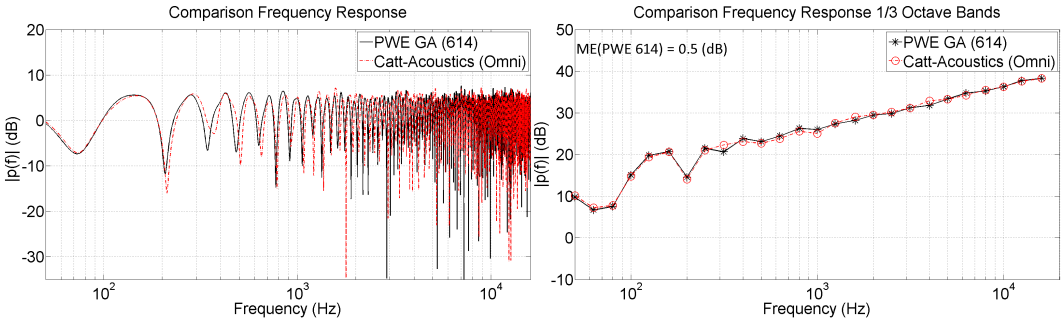


FIGURE 5.7: Comparison of frequency responses. Translated PWE (GA) and omnidirectional RIR at the translated position 2, single wall

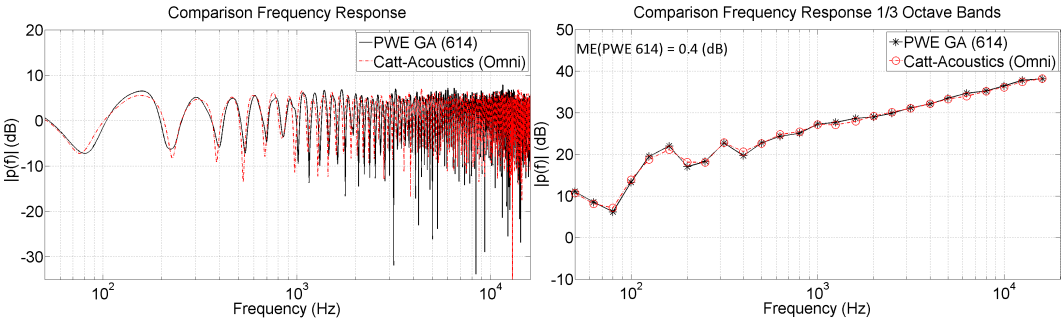


FIGURE 5.8: Comparison of frequency responses. Translated PWE (GA) and omnidirectional RIR at the translated position 3, single wall

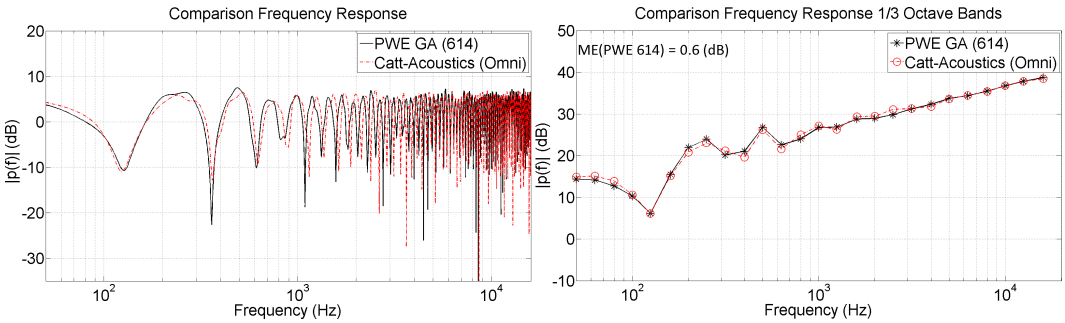


FIGURE 5.9: Comparison of frequency responses. Translated PWE (GA) and omnidirectional RIR at the translated position 4, single wall

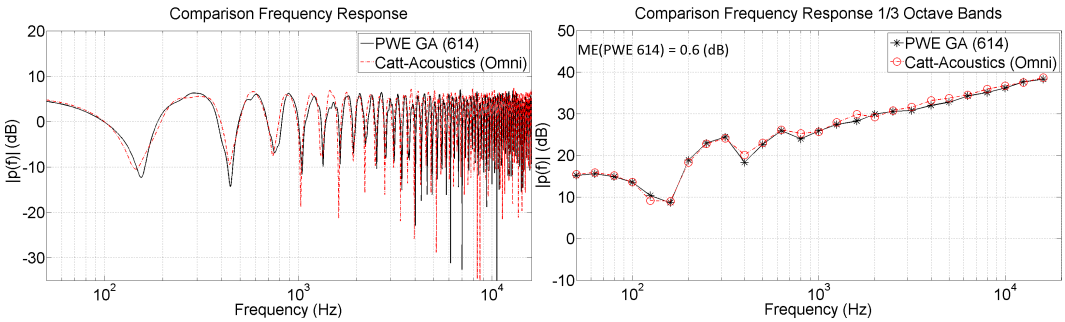


FIGURE 5.10: Comparison of frequency responses. Translated PWE (GA) and omnidirectional RIR at the translated position 5, single wall

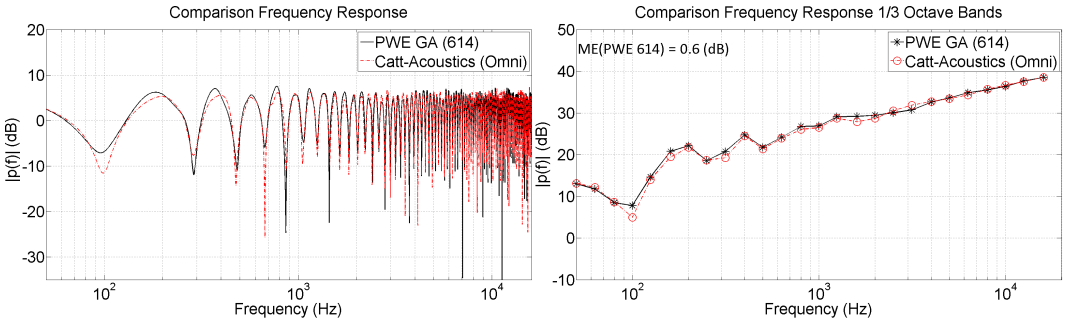


FIGURE 5.11: Comparison of frequency responses. Translated PWE (GA) and omnidirectional RIR at the translated position 6, single wall

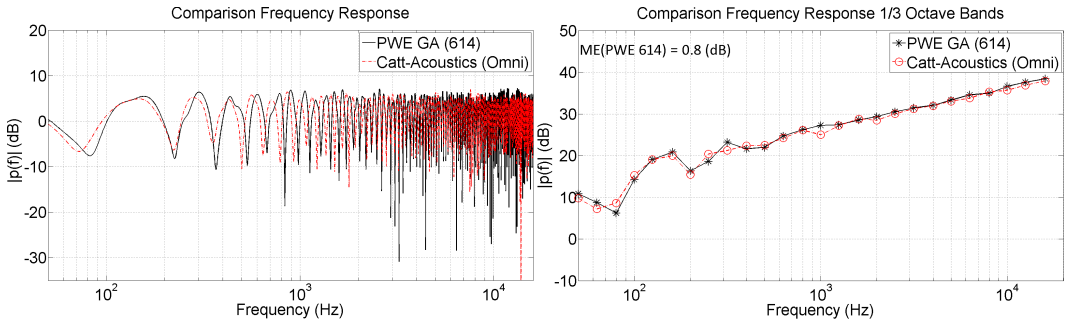


FIGURE 5.12: Comparison of frequency responses. Translated PWE (GA) and omnidirectional RIR at the translated position 7, single wall

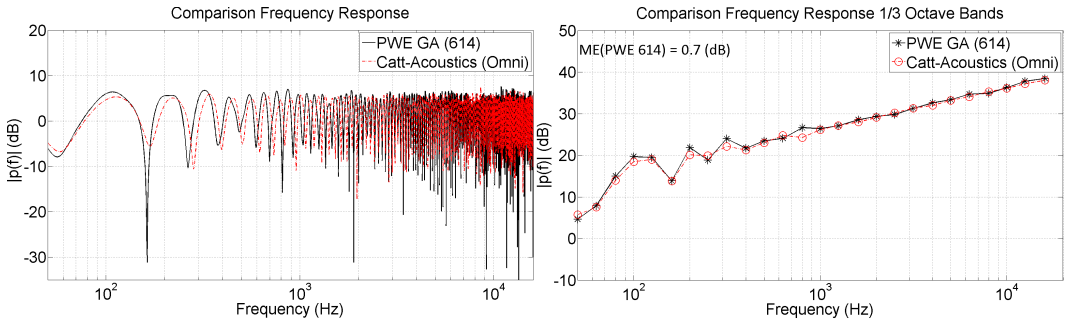


FIGURE 5.13: Comparison of frequency responses. Translated PWE (GA) and omnidirectional RIR at the translated position 8, single wall

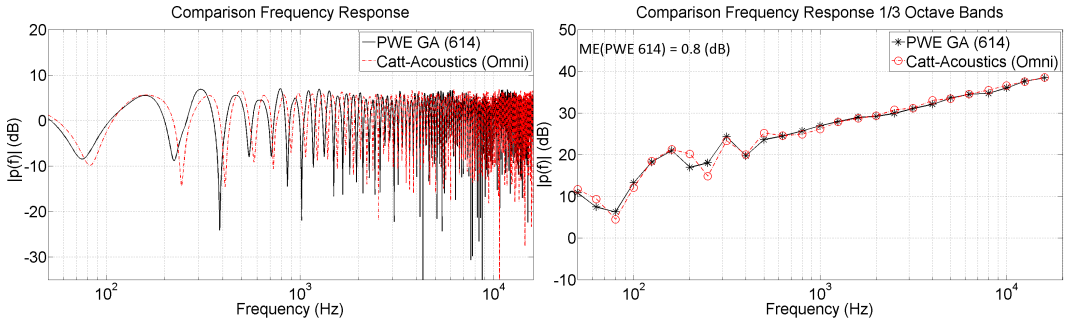


FIGURE 5.14: Comparison of frequency responses. Translated PWE (GA) and omnidirectional RIR at the translated position 9, single wall

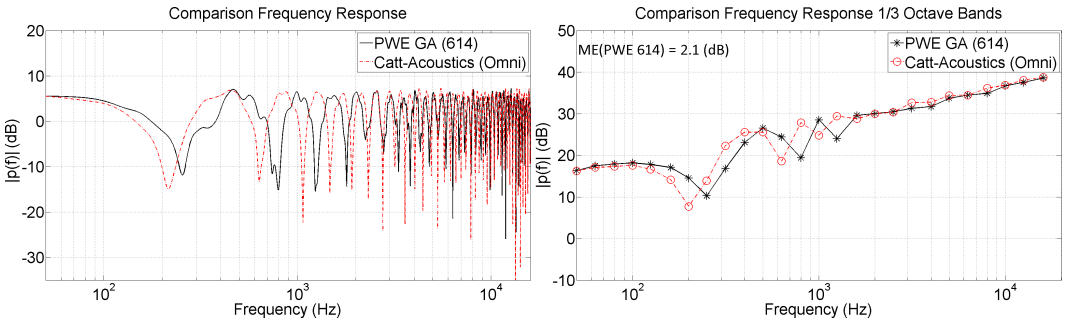


FIGURE 5.15: Comparison of frequency responses. Translated PWE (GA) and omnidirectional RIR at the translated position 10, single wall

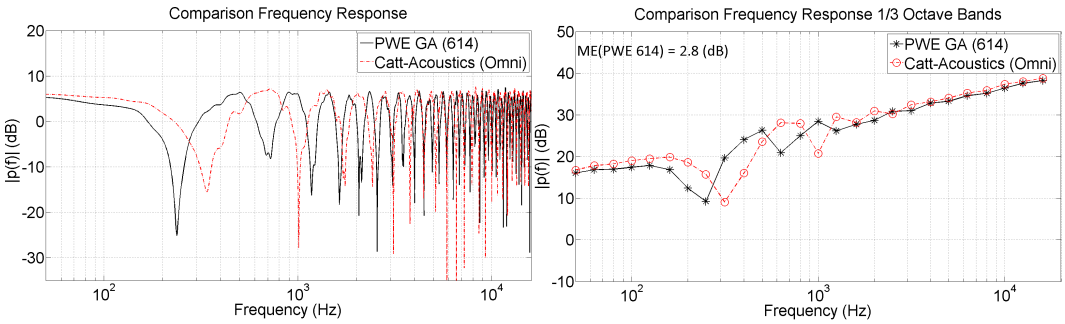


FIGURE 5.16: Comparison of frequency responses. Translated PWE (GA) and omnidirectional RIR at the translated position 11, single wall

The outcomes indicate that the implementation of the translation operator in the plane wave expansion leads to similar results in terms of energy compared to the omnidirectional impulse responses. The agreement is better at high frequencies and at receivers closest to the central point of the expansion. A higher mismatch was found for receivers 10 and 11 below 1 kHz, this disagreement may be caused by the distance between the reflector and the receivers that in this case is shorter than for the others receivers.

5.1.1.1 Rectangular room

Further analysis was conducted using a more complex enclosure. Figure 5.17 illustrates the source and receiver positions in a rectangular room of dimensions (5m, 10m, 3m) that was simulated using the same approach described above. The boundaries were characterized with an absorption and a scattering coefficient of 0.15 and 0.1, respectively. The central point of the expansion corresponds to the receiver 01.

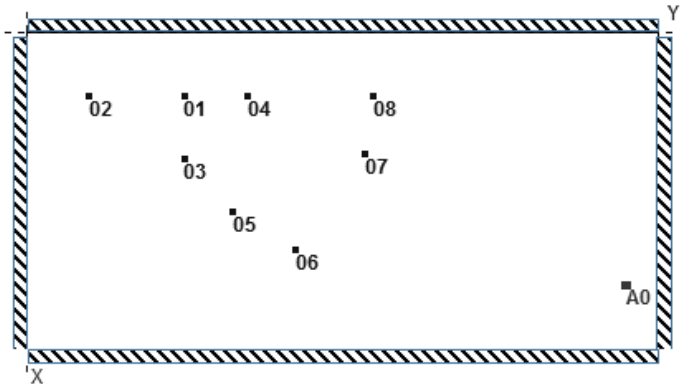


FIGURE 5.17: Sketch of the rectangular room model used to evaluate the plane wave expansion from GA simulations

A trial of 3 simulations was conducted to predict the frequency response at the receiver position 01. The calculations were performed without any variation in the simulation parameters. The results (see Figure 5.18) indicate that the synthesized frequency responses are similar but not identical. This outcome is a consequence of the stochastic implementation of the scattering coefficients in Catt-Acoustics.

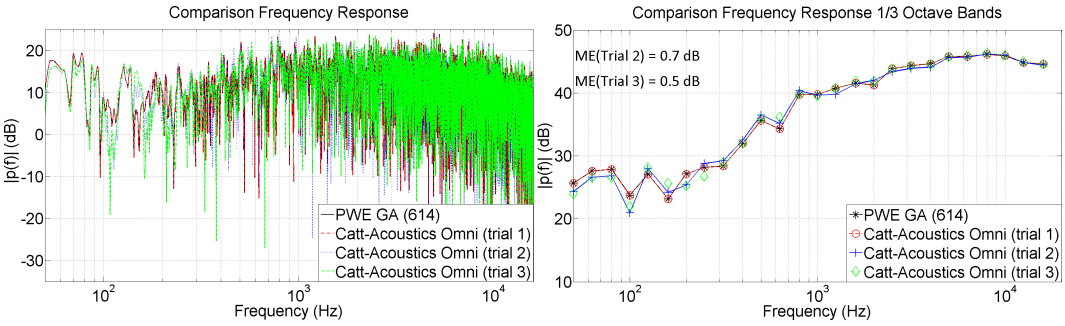


FIGURE 5.18: Comparison of frequency responses. Trial of three GA simulations at the receiver position 1, rectangular room

Figures 5.19 to 5.25 illustrate the frequency response in narrow band and in 1/3 octave band resolution at the other receivers. The cyan vertical line corresponds to the crossover frequency (355 Hz) that is used when the results are combined with FE simulations. This frequency line is displayed because the mismatch tends to be greater at low frequencies. However, the GA information in this frequency range will be replaced by FE data when the two approaches are combined.

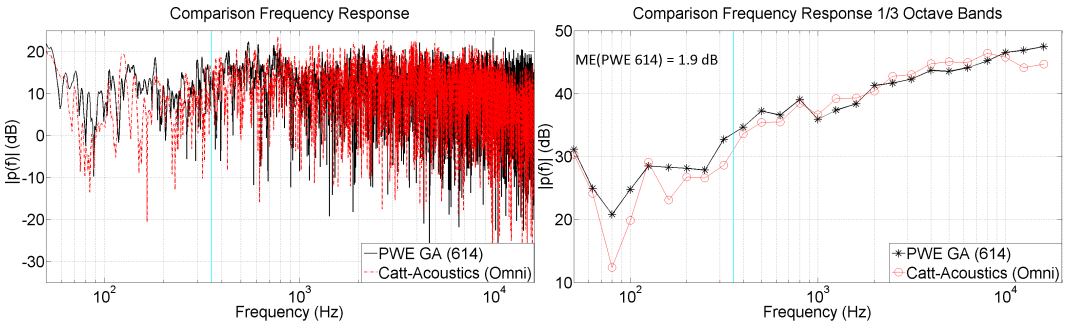


FIGURE 5.19: Comparison of frequency responses. Translated PWE (GA) and omnidirectional RIR at the translated position 2, rectangular room

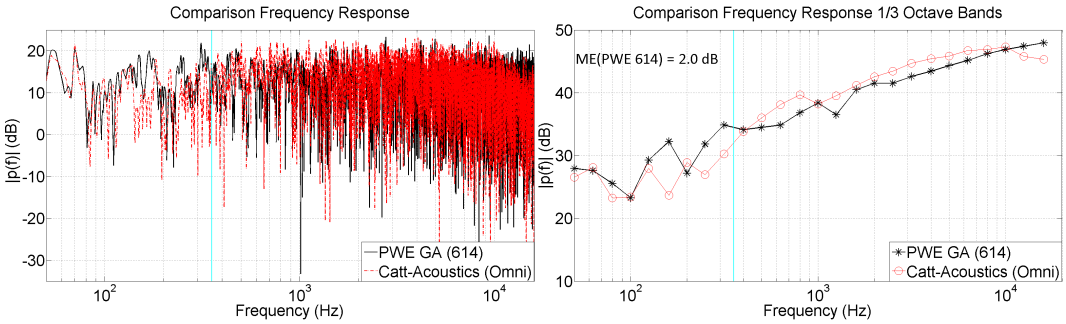


FIGURE 5.20: Comparison of frequency responses. Translated PWE (GA) and omnidirectional RIR at the translated position 3, rectangular room

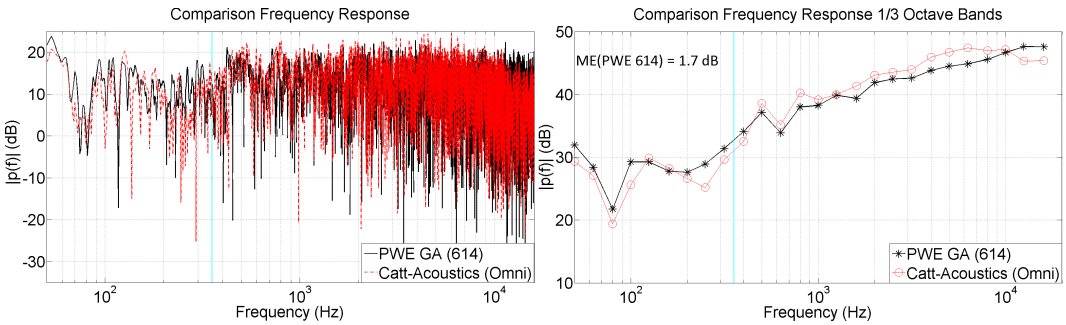


FIGURE 5.21: Comparison of frequency responses. Translated PWE (GA) and omnidirectional RIR at the translated position 4, rectangular room

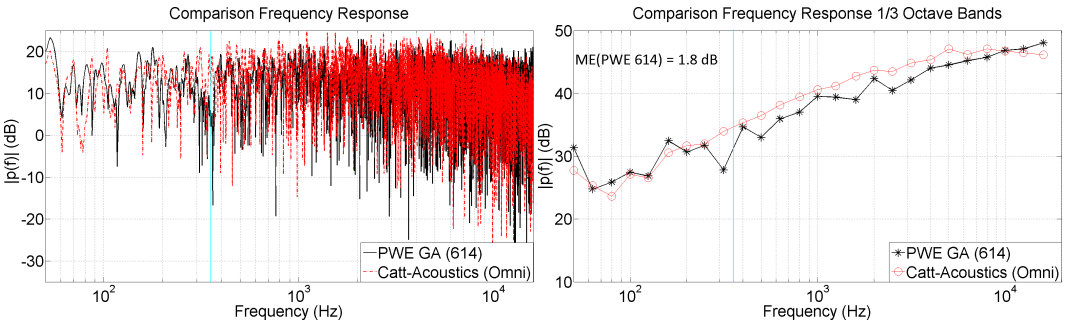


FIGURE 5.22: Comparison of frequency responses. Translated PWE (GA) and omnidirectional RIR at the translated position 5, rectangular room

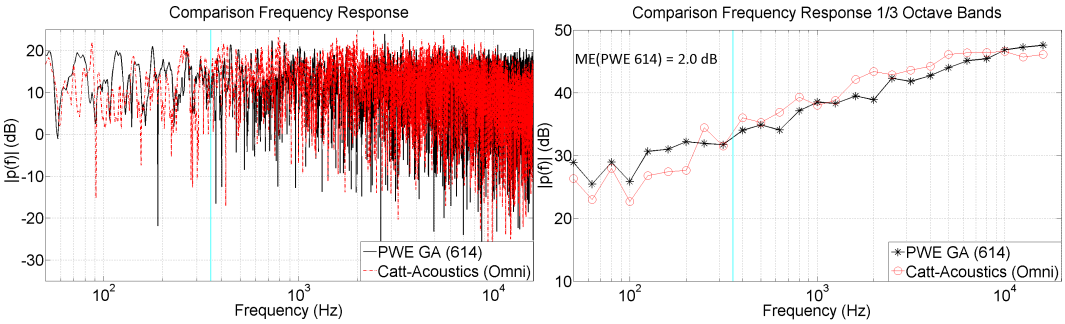


FIGURE 5.23: Comparison of frequency responses. Translated PWE (GA) and omnidirectional RIR at the translated position 6, rectangular room

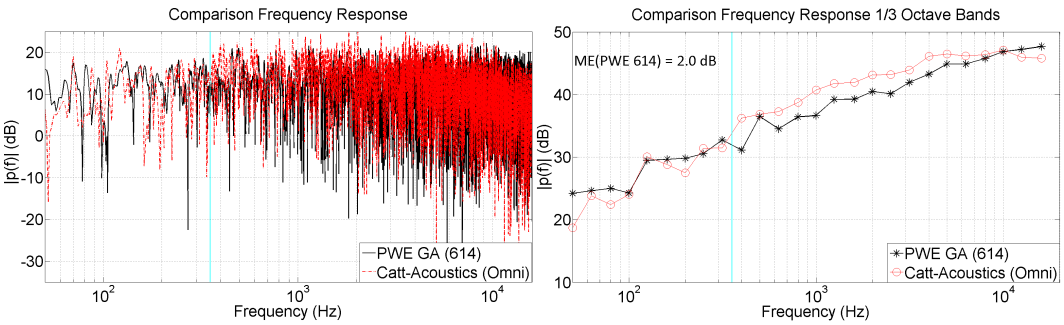


FIGURE 5.24: Comparison of frequency responses. Translated PWE (GA) and omnidirectional RIR at the translated position 7, rectangular room

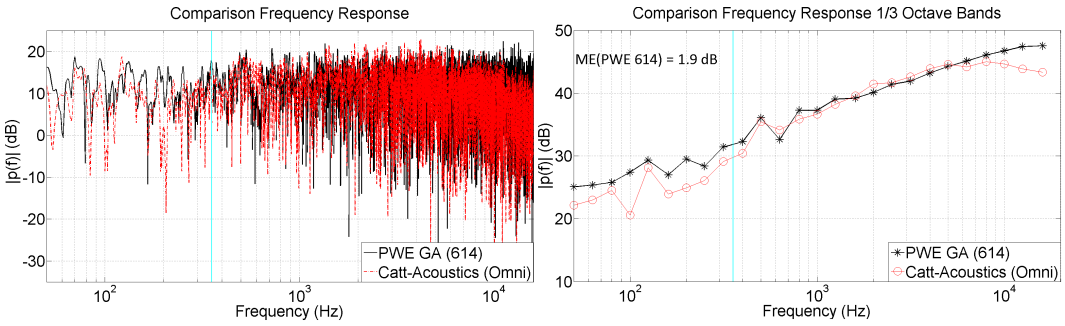


FIGURE 5.25: Comparison of frequency responses. Translated PWE (GA) and omnidirectional RIR at the translated position 8, rectangular room

The results indicate that the level of accuracy in the reconstruction is not as high as for the single reflector case. This is expected because there are more reflective boundaries, which lead to a higher order of reflections. Each reflection is affected by the stochastic implementation of the scattering coefficients, thus yielding a higher mismatch compared to the translated plane wave expansion. The outcomes for this enclosure also suggest that the accuracy does not depend on the distance from the central point of the expansion, as it is illustrated in Table 5.1. In general, it was found that the spectral shape is correctly predicted but with less energy below the 1/3 octave band of 5 kHz. In contrast, for the 1/3 octave band of 10 kHz and above, the energy predicted by the plane wave expansion is higher than the omnidirectional references. The wavelengths in this frequency range are short enough to assume a plane wave propagation, which suggests that the mismatch may be related to different reasons than the propagation model.

Receiver	**Distance (m)**	**Mean Error (dB)**
02	1.5	1.9
03	1	2.0
04	1	1.7
05	2	1.8
06	3	2.0
07	3	2.0
08	3	1.9

TABLE 5.1: Mean error according to the distance to the central point of the expansion, rectangular room

A different approach was implemented to identify whether the mismatch in the energy levels is due to the cleaning algorithm applied to compensate for the tessellation of the unit sphere when the plane wave expansion from the GA data is created (see section 4.3). The PWE used to process the finite element data corresponds to 64 plane waves uniformly distributed over a unit sphere, which leads to an average angle of 25.02 degrees between plane waves. The coordinates of the 64 plane waves were used to sample the GA data using the *sector mic* with a resolution of 25 degrees. The difference of 0.2 degrees helps to reduce the overlapping produced by the squared tessellation of the sphere when the GA data is spatially discretized. Figures 5.26 and 5.27 show the frequency response in narrow band and in 1/3 octave band resolution for the receivers 03 and 08. The findings indicate that this approach yields similar results to those obtained by the initial plane wave expansion of 614 elements, in particular in the octave bands of 8 kHz and 16 kHz. This outcome suggests that the mismatch found at these octave bands is not due to the cleaning algorithm used for the computation of the plane wave expansion of 614 elements.

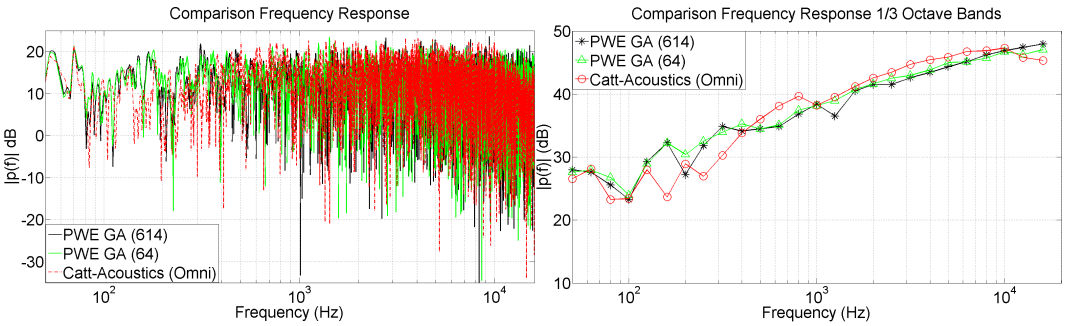


FIGURE 5.26: Comparison of frequency responses. Translated GA PWE (614 and 64) and the omnidirectional RIR at the translated position 3, rectangular room

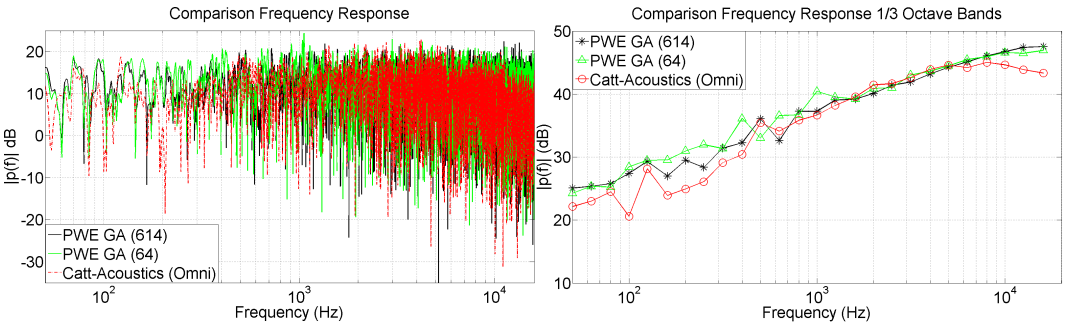


FIGURE 5.27: Comparison of frequency responses. Translated GA PWE (614 and 64) and the omnidirectional RIR at the translated position 8, rectangular room

It is important to emphasize that the use of omnidirectional impulse responses (predicted by Catt-Acoustics) as references presents two inconveniences. The first is the stochastic implementation of the scattering coefficients that changes the direction of the reflections in a random pattern. The second corresponds to the generation of the impulse response from the energy echogram, which involves the use of minimum and linear phase filters to synthesize the phase. These two internal processes carried out by Catt-Acousitcs cannot be recreated by the plane wave expansion leading to additional differences between the two approaches.

5.1.1.2 Meeting room

Figure 5.28 illustrates 8 receiver positions that were used to evaluate the plane wave propagation assumption in the meeting room. The central point of the expansion corresponds to receiver 01 and the location of acoustic source is identified as B1. Figures 5.30 to 5.36 show the comparison of the predicted frequency response in narrow band and in 1/3 octave band resolution using the translation operator and the omnidirectional receivers calculated from Catt-Acoustics.

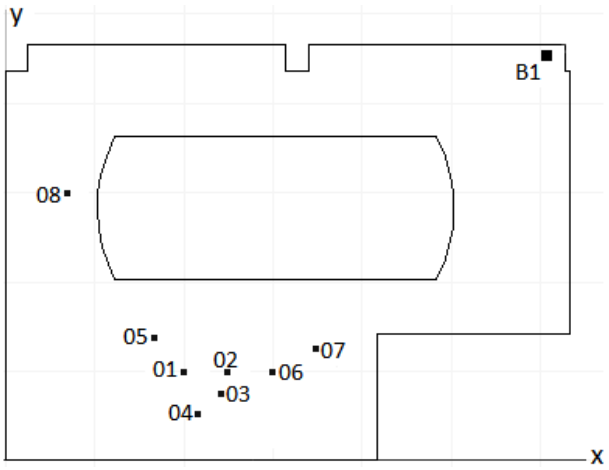


FIGURE 5.28: Sketch of the meeting room model used to evaluate the plane wave expansion from GA simulations

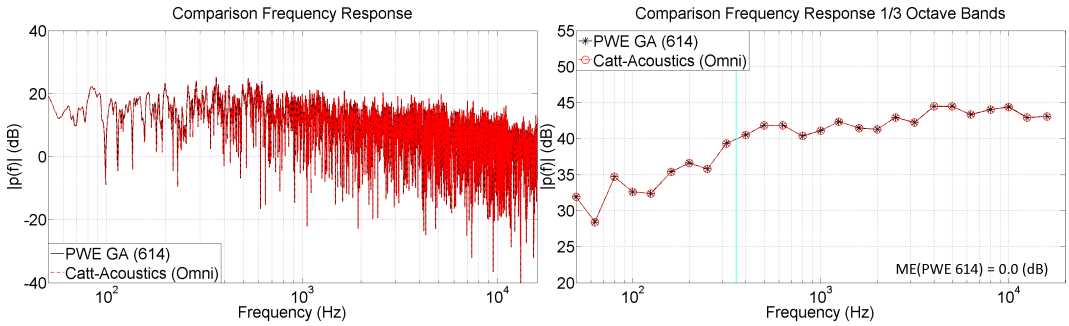


FIGURE 5.29: Comparison of frequency responses. PWE (GA) and omnidirectional RIR at the reference position 1, meeting room

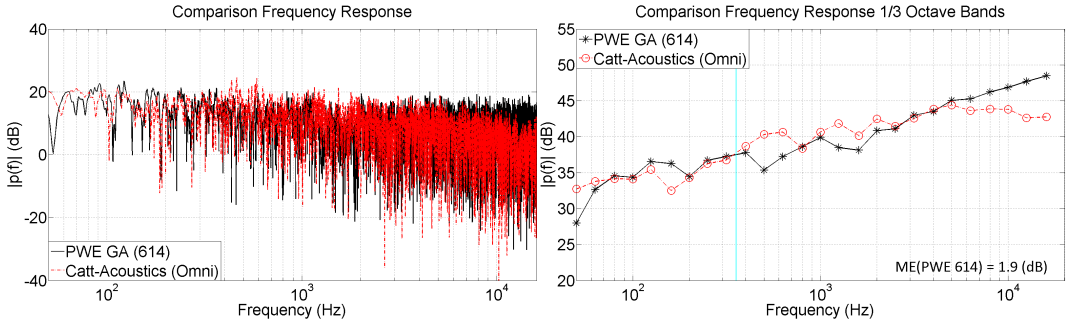


FIGURE 5.30: Comparison of frequency responses. Translated PWE (GA) and omnidirectional RIR at the translated position 2, meeting room

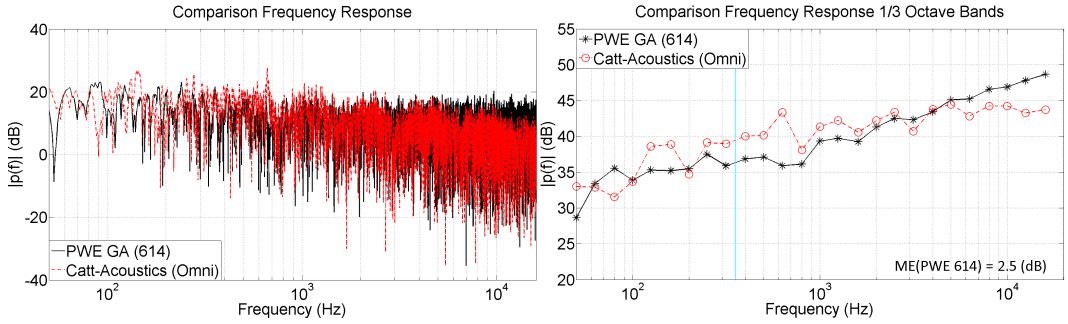


FIGURE 5.31: Comparison of frequency responses. Translated (GA) and omnidirectional RIR at the translated position 3, meeting room

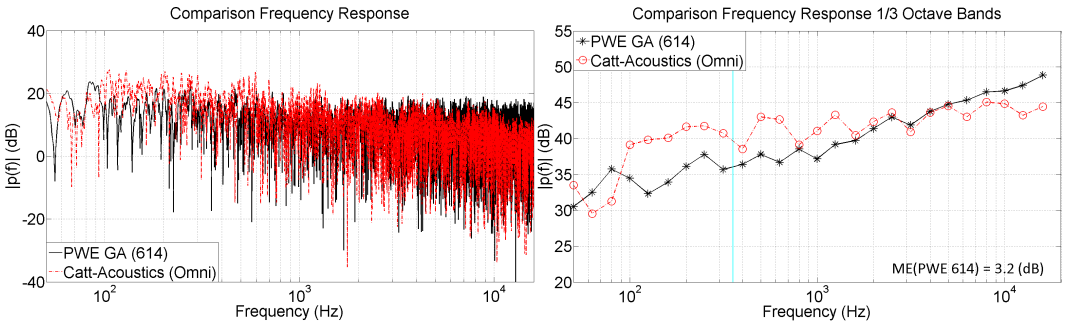


FIGURE 5.32: Comparison of frequency responses. Translated PWE (GA) and omnidirectional RIR at the translated position 4, meeting room

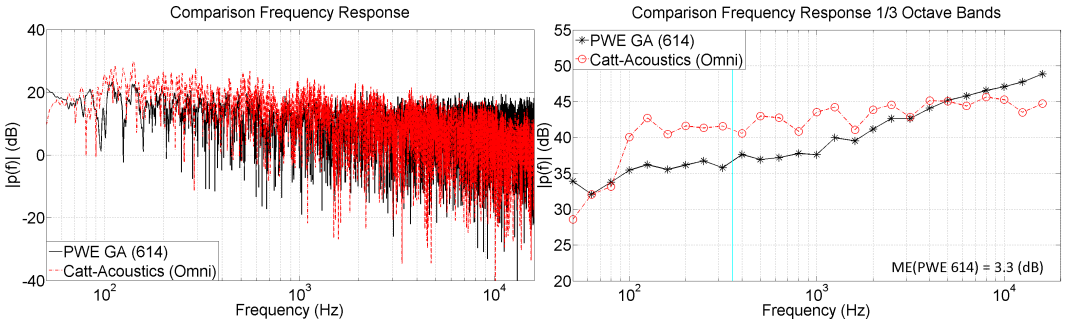


FIGURE 5.33: Comparison of frequency responses. Translated PWE (GA) and omnidirectional RIR at the translated position 5, meeting room

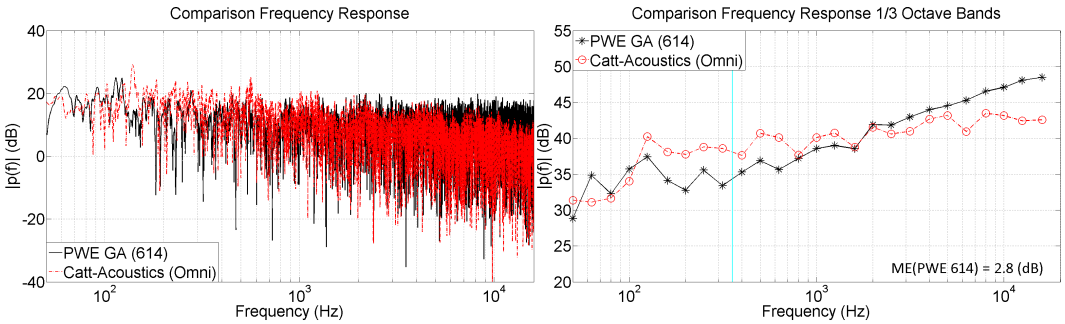


FIGURE 5.34: Comparison of frequency responses. Translated PWE (GA) and omnidirectional RIR at the translated position 6, meeting room

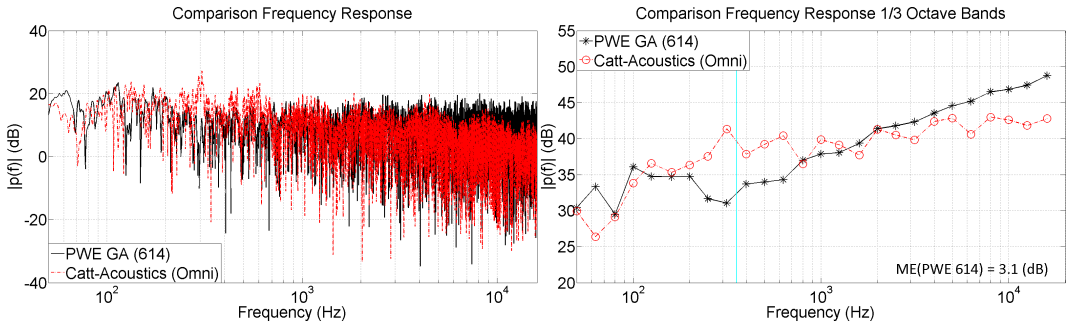


FIGURE 5.35: Comparison of frequency responses. Translated PWE (GA) and omnidirectional RIR at the translated position 7, meeting room

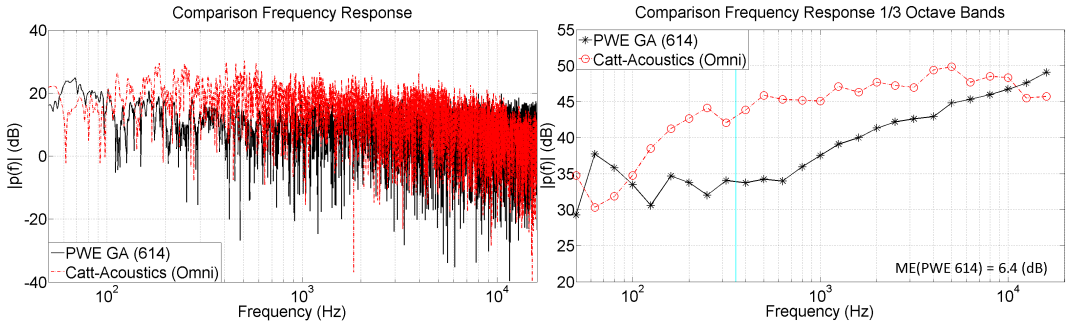


FIGURE 5.36: Comparison of frequency responses. Translated PWE (GA) and omnidirectional RIR at the translated position 8, meeting room

The figures indicate a poorer agreement between the omnidirectional frequency responses and the translation of the plane wave expansion compared to the results obtained for the rectangular room. Two frequency ranges in which the mismatch is high were found for all the receivers. The first corresponds to the 500 Hz octave band and the second to the 8 kHz and 16 kHz octave bands. The reason for this mismatch at 500 Hz can be associated with the smaller dimensions of the enclosure, which compromises the plane propagation assumption. However, for the two higher octave bands, the wavelengths are sufficiently short that the errors are not associated with the assumption of plane wave propagation. The divergence at these frequencies was also present in the case of the rectangular room, which indicates that the mismatch is due to a different reason than the plane wave assumption. Table 5.2 shows the mean error according to the distance of the receivers. The results show that error fluctuates between 1.9 dB and 3.3 dB within a radius of 1.5 m.

Receiver	Distance (m)	Mean Error (dB)
02	0.5	1.9
03	0.5	2.5
04	0.5	3.2
05	0.5	3.3
06	1	2.8
07	1.5	3.1
08	2.4	6.4

TABLE 5.2: Mean error according to the distance to the central point of the expansion, meeting room

5.1.1.3 Ightham Mote

The same approach used for the meeting room was implemented for the Ightham Mote to assess the accuracy of plane wave representation. Figure 5.37 illustrates the location of 8 receivers considered for the analysis. B0 identifies the position of the acoustic source and the central point of the expansion corresponded to receiver 01. Figures 5.39 to 5.45 show the predicted frequency response in narrow band and in 1/3 octave band resolution.

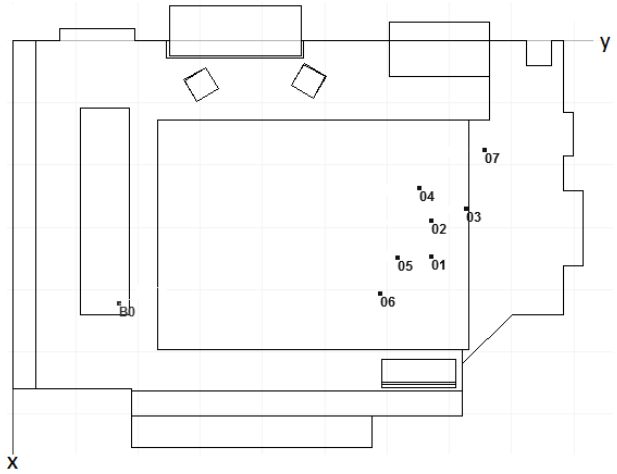


FIGURE 5.37: Sketch of the Ightham Mote model used to evaluate the plane wave expansion from GA simulations

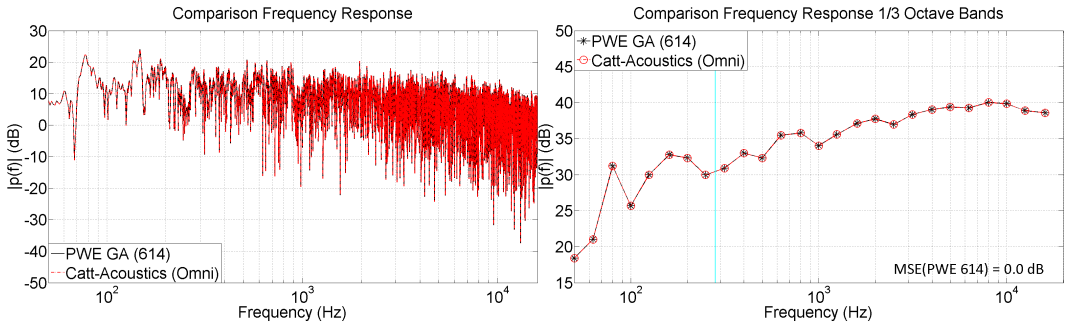


FIGURE 5.38: Comparison of frequency responses. PWE (GA) and omnidirectional RIR at the reference position 1, Ightham Mote

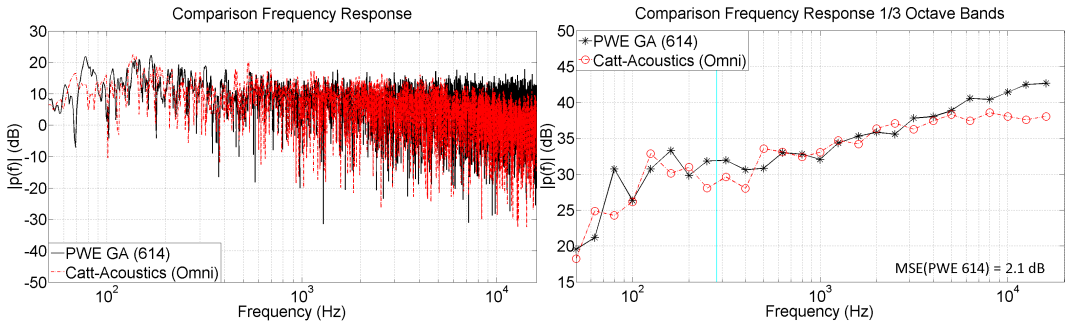


FIGURE 5.39: Comparison of frequency responses. Translated PWE (GA) and omnidirectional RIR at the translated position 2, Ightham Mote

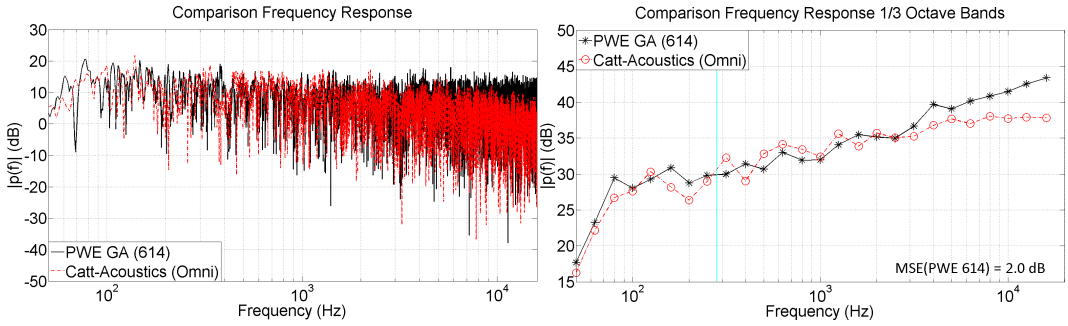


FIGURE 5.40: Comparison of frequency responses. Translated PWE (GA) and omnidirectional RIR at the translated position 3, Ightham Mote

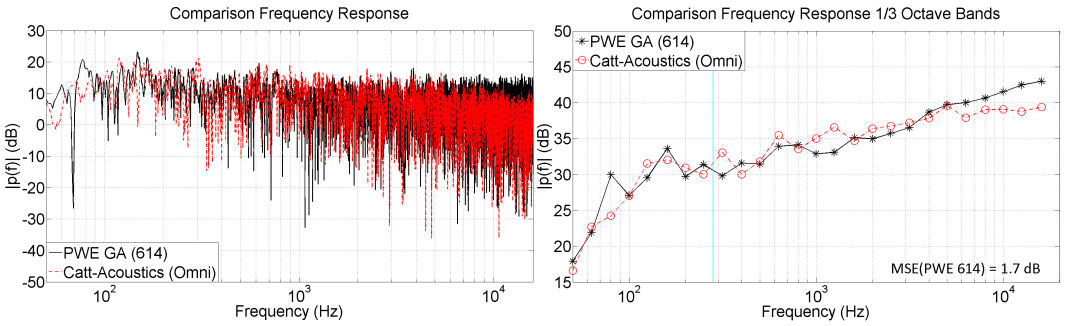


FIGURE 5.41: Comparison of frequency responses. Translated PWE (GA) and omnidirectional RIR at the translated position 4, Ightham Mote

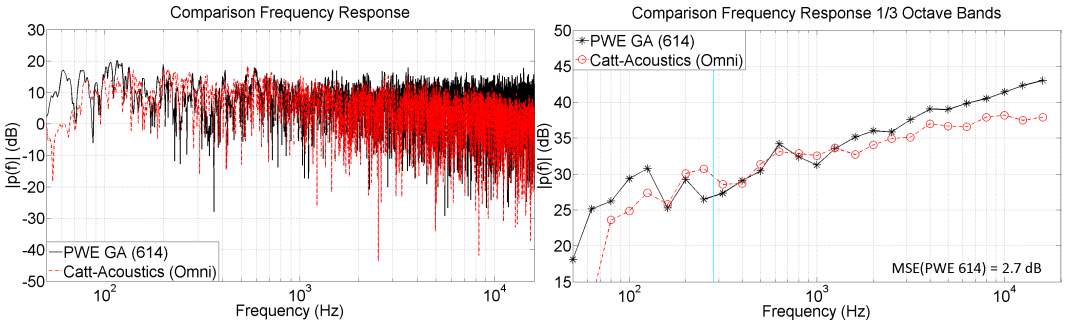


FIGURE 5.42: Comparison of frequency responses. Translated PWE (GA) and omnidirectional RIR at the translated position 5, Ightham Mote

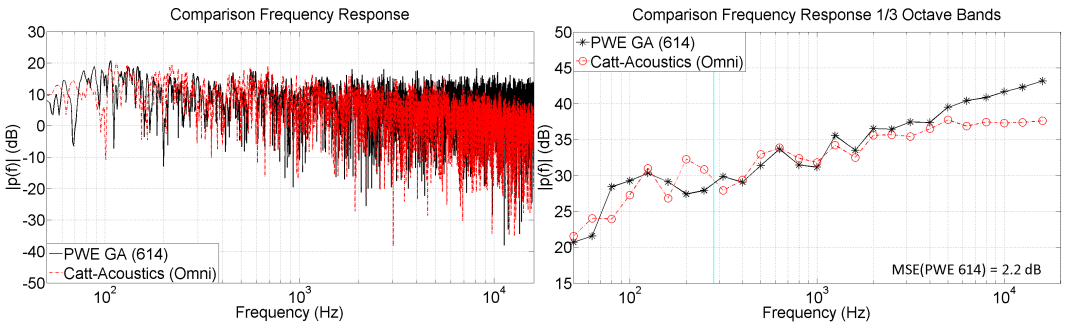


FIGURE 5.43: Comparison of frequency responses. Translated PWE (GA) and omnidirectional RIR at the translated position 6, Ightham Mote

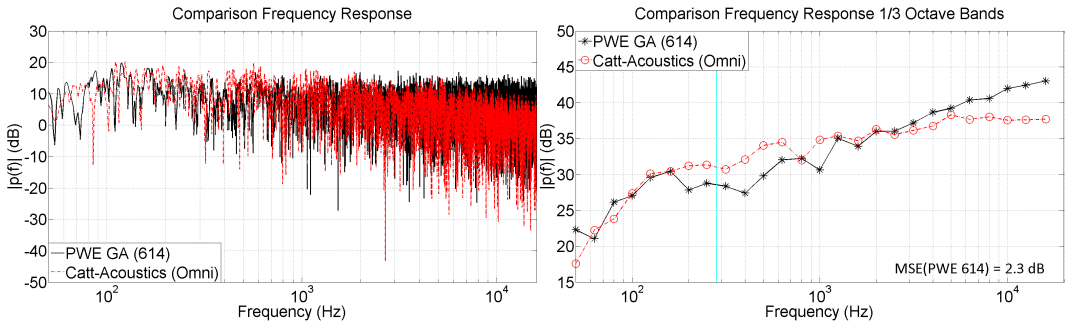


FIGURE 5.44: Comparison of frequency responses. Translated PWE (GA) and omnidirectional RIR at the translated position 7, Ightham Mote

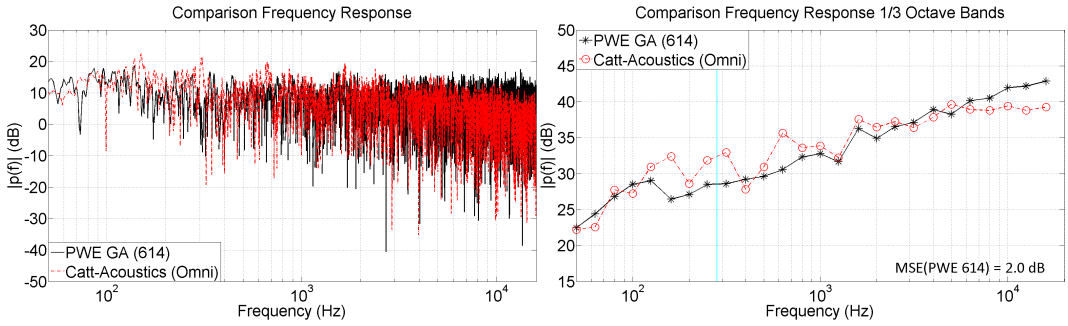


FIGURE 5.45: Comparison of frequency responses. Translated PWE (GA) and omnidirectional RIR at the translated position 8, Ightham Mote

The results show that there is a better agreement between the frequency responses predicted by the translation of the plane wave expansion and the omnidirectional references compared to the meeting room case. The Ightham Mote has a volume ≈ 5 times larger than the meeting room, which favours the plane wave model assumption. These outcomes support the postulation that the mismatch in the octave band of 500 Hz found for the meeting room may be associated with the assumption of plane wave propagation. A trend similar of the previous rooms was found at the octave bands of 8 kHz and 16 kHz for this specific enclosure. Table 5.3 summarizes the mean error according to the distance of the receivers to the central point of the expansion (receiver 01).

Receiver	Distance (m)	Mean Error (dB)
02	0.5	2.1
03	0.5	2.0
04	0.5	1.7
05	1	2.7
06	1	2.2
07	1.5	2.3
08	1.5	2.0

TABLE 5.3: Mean error according to the distance to the central point of the expansion, Ightham Mote

The analysis of the 3 previous enclosures reveals that the size of the room affects the validity of the plane wave propagation assumption. As expected, the translation of the plane wave expansion is more accurate in larger enclosures. Tables 5.1, 5.2 and 5.3 indicate that translation within a region of 1.5 m from the central point of the expansion (where the reconstruction is exact) leads to a mean error between 2.0 and 2.3 dB for the rectangular room and Ightham Mote. For the meeting room, the mean error within this radius is higher at 3.3 dB. Figure 5.18 shows that the mean error between simulations carried out without any modification in the parameters can fluctuate up to 0.7 dB, which highlights the stochastic component in the geometrical acoustic predictions. The predictions conducted without any variation in the simulation parameters corresponds to the rectangular room that is a simple case where the assumptions of geometrical acoustic algorithms are very suitable. A higher mean error is expected between simulations in more complex rooms (e.g. the meeting room) because the assumption given by GA are more constrained. Finally, it is important to emphasize that the stochastic component is only related to the second and higher orders of reflections where the cone-tracing algorithm is affected by the scattering coefficients.

5.1.1.4 Energy mapping of the plane wave expansion predcited from GA data

A common result between the three enclosures was the overestimation of the energy in the 8 kHz and 16 kHz octave bands when compared to the omnidirectional references. This result is unexpected because the wavelengths in these frequencies are short

enough to consider the plane propagation as a reasonable approach, even in small spaces as the meeting room. Furthermore, the frequencies of the 2 kHz and 4 kHz octave bands (whose wavelengths are larger) are correctly predicted, which support the hypothesis that the mismatch is not due to the chosen model. A spatial energy mapping at the 1/3 octave bands of 1 kHz and 16 kHz is presented in Figures 5.46 to 5.49 for 4 enclosures. The spaces correspond to the rectangular room, meeting room, office room and the Ightham Mote. The energy maps were generated by predicting the room impulse responses at different receiver positions (squared grid with resolution of 0.5 m) using the translation operator. Then, the energy was calculated at each receiver location and displayed as a continuous surface using an interpolation algorithm.

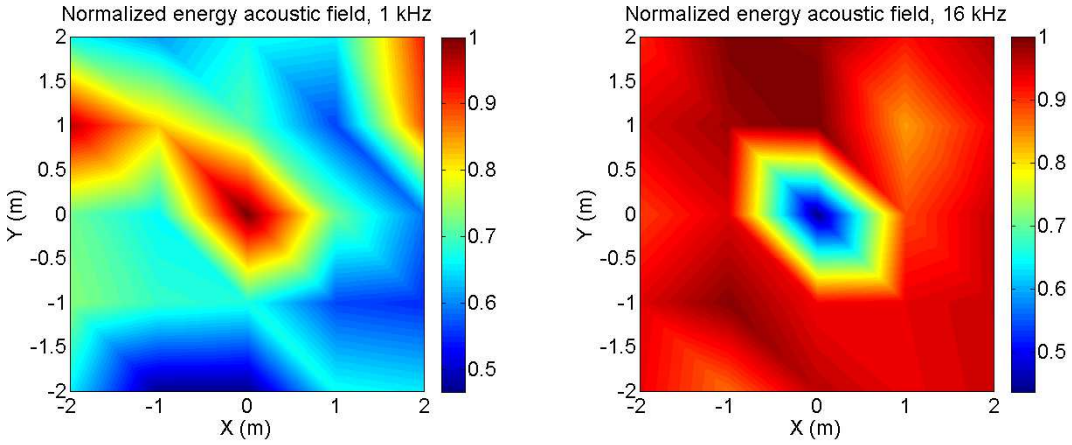


FIGURE 5.46: Spatial energy mapping, PWE (GA), rectangular room

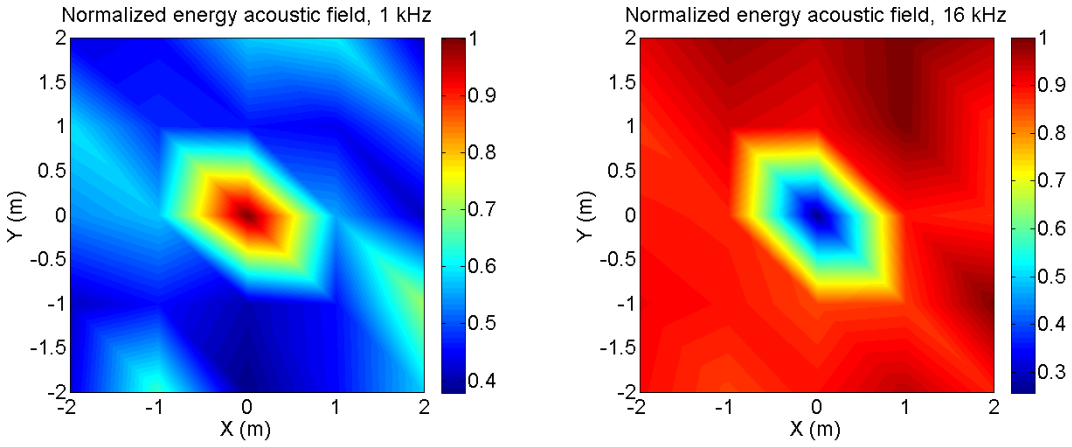


FIGURE 5.47: Spatial energy mapping, PWE (GA), meeting room

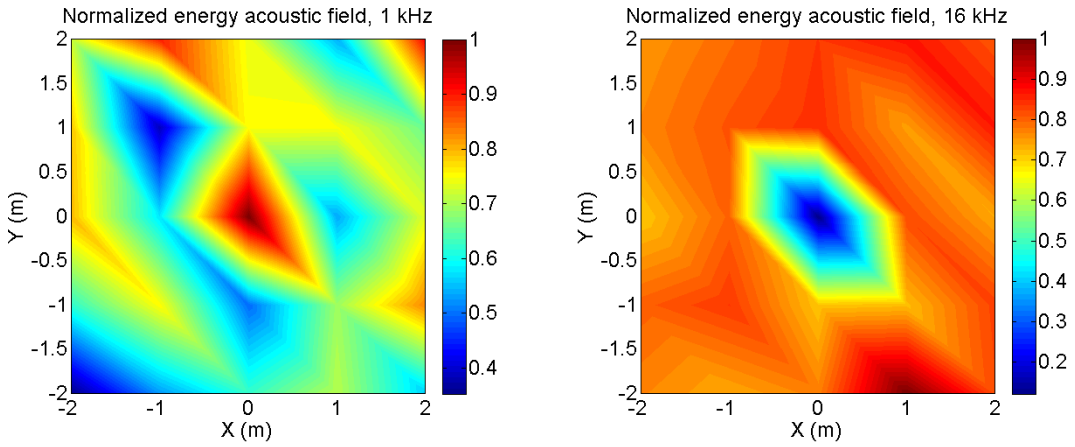


FIGURE 5.48: Spatial energy mapping, PWE (GA), office room

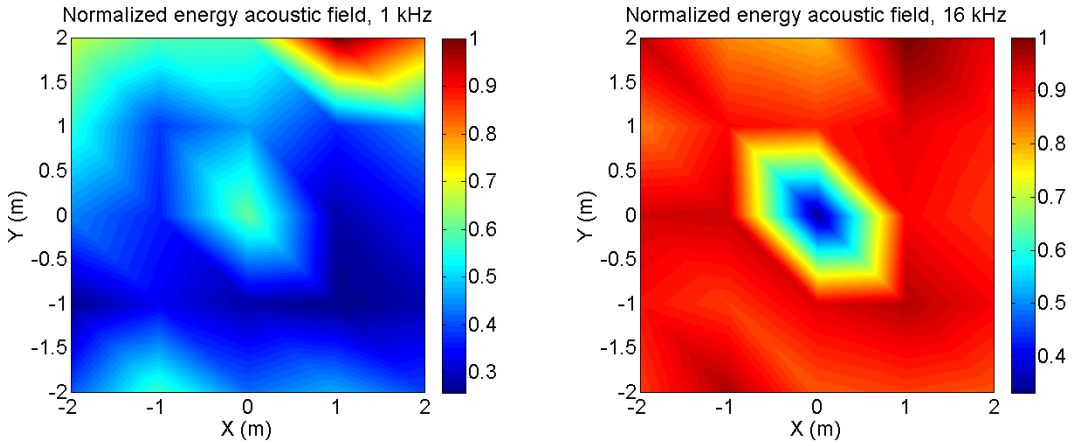


FIGURE 5.49: Spatial energy mapping, PWE (GA), Ightham Mote

The results indicate that the overestimation of the energy at the 1/3 octave band of 16 kHz is an intrinsic feature of the method used to estimate the plane wave expansion from the GA data. For all enclosures, the energy is lower at the central point of the expansion than at the rest of the evaluated field points for the 1/3 octave band of 16 kHz.

The use of a commercial package imposes limitations on the amount of information that can be accessed, thus making difficult to determine the reason for this overestimation. Possible reasons can be related to

- The synthesis of the phase in the impulse responses, which involves the use of minimum and linear filters.

- The simulations in Catt-Acoustics are carried out by defining the boundary conditions up to 4 kHz. Above that frequency band, the calculations are performed using extrapolated data.

Nevertheless, a clear explanation of this outcome is not evident and remains as an open topic for future research.

5.1.1.5 Evaluation of the closest neighbour, spherical harmonic interpolation and VBAP to resample the PWE from GA data

Three different strategies to generate a representation of the geometrical acoustic data that is suitable with the plane wave expansion used to process the finite element results were proposed in section 4.3.1. It is important to emphasize that the three proposed algorithms reduce the 614 plane waves representation to 64 plane waves only to make the approach compatible with FEM. In the following, a further analysis based on the translation of the sound field is presented to evaluate the accuracy of these three approaches. Two target fields at 250 Hz corresponding to a single plane wave and a more complex field created by the interaction of 3 different plane waves have been calculated analytically. Then, the target fields are translated from the origin of the coordinate system (cyan point) to an established evaluation position (magenta point) using the translation operator. The target sound fields are recreated using the three different algorithms proposed in section 4.3.1, namely the closest neighbour (CN), spherical harmonic interpolation (SH) and VBAP. A translation of the reconstructed sound fields is also applied to compare them with respect to the target translated fields. A general description of how the algorithms are implemented is presented as follows:

- **CN**: The incoming directions of the plane waves are modified according to the closest possible direction determined by the discretized plane wave expansion ($L = 64$) used to process the FE data (section 4.3.1.1).
- **SH**: Each plane wave is encoded using 64 complex spherical harmonic coefficients. Then, these complex coefficients are recreated using the plane wave expansion of 64 elements based on a mode-matching decoding approach (section 4.3.1.2).
- **VBAP:** Each plane wave is represented as the sum of three different plane waves whose incoming directions are defined by the 64 directions used for the

plane wave expansion. The selection of the three plane waves and the calculation of their complex amplitude are carried out by a VBAP implementation (section 4.3.1.3).

A complete description of the implementation of the previous methods is found in section 4.3.1. Figure 5.50 shows the incoming direction of the elements of the plane wave expansion as black circles. The red diamonds represent the 3 plane waves that synthesize the complex field and the blue square identifies the incoming direction of the single plane wave case. Figures 5.51 to 5.56 illustrate the real part of the acoustic pressure (in Pa) for the target and reconstructed sound fields according to the approach use to resample the plane wave expansion.

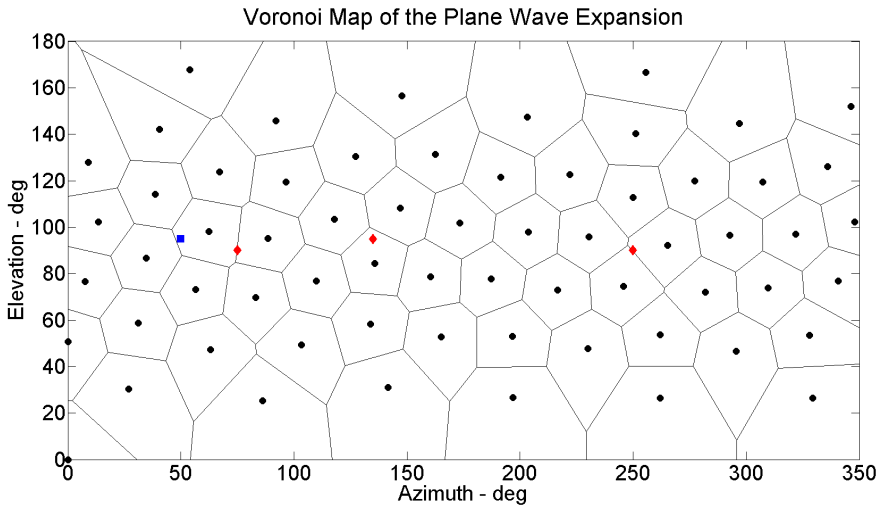


FIGURE 5.50: Voronoi map identifying the incoming direction of the elements of the plane wave expansion

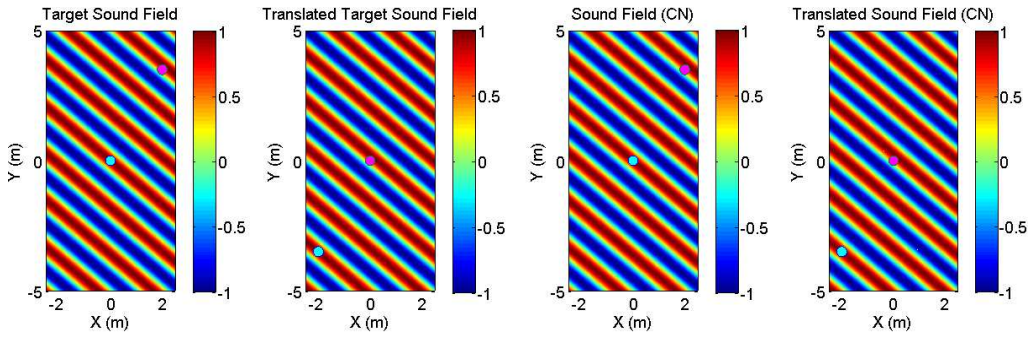


FIGURE 5.51: Target and reconstructed field, 1 plane wave, closest neighbour

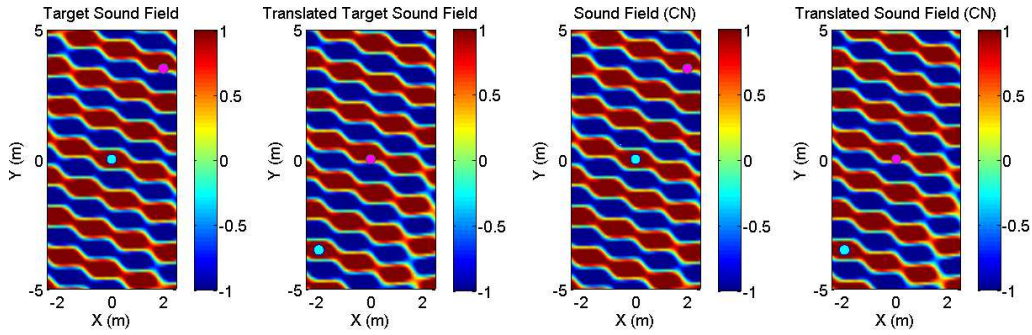


FIGURE 5.52: Target and reconstructed field, 3 plane waves, closest neighbour

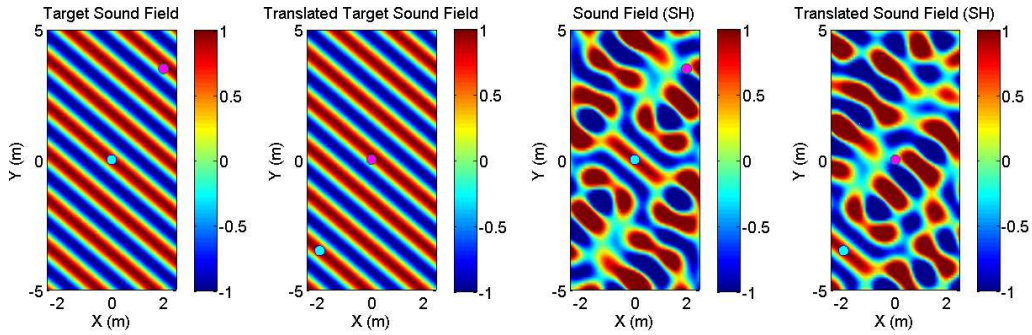


FIGURE 5.53: Target and reconstructed field, 1 plane wave, spherical harmonics

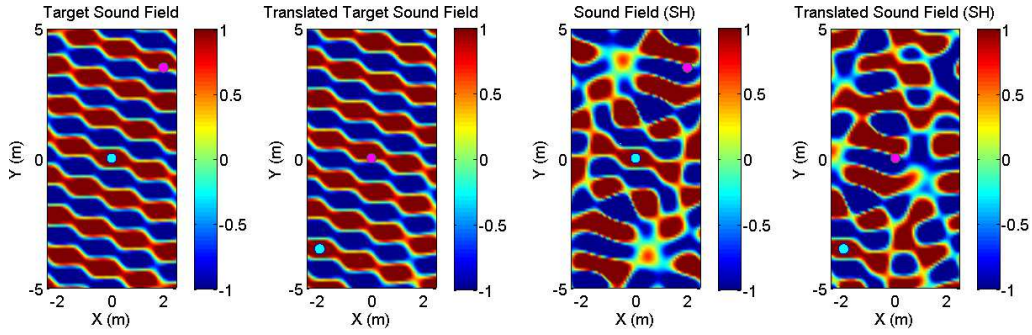


FIGURE 5.54: Target and reconstructed field, 3 plane waves, spherical harmonics

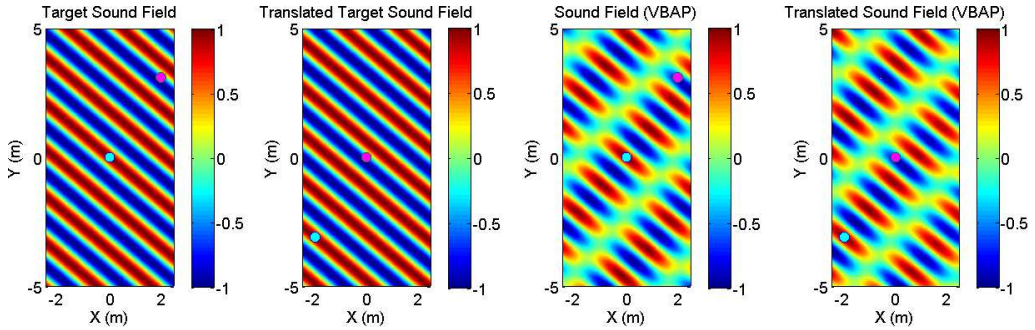


FIGURE 5.55: Target and reconstructed field, 1 plane wave, VBAP

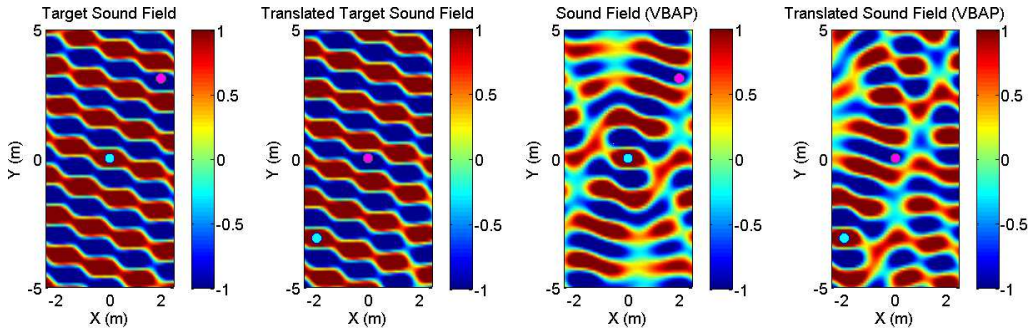


FIGURE 5.56: Target and reconstructed field, 3 plane waves, VBAP

The results show that the accuracy on the reconstructed sound fields depends on the approach used to represent the plane waves. The artifacts produced by the implementation of the spherical harmonic interpolation and VBAP are clearly illustrated in Figures 5.53 to 5.56. In the case of the closest neighbour, the reconstruction is more accurate than the other two methods for these two specific target fields. Consistently with previous results, the degree of agreement between the translated target and reconstructed fields is determined by the initial accuracy of the reconstructed (non-translated) sound fields. Figures 5.57 to 5.62 show the acoustic errors defined in section 4.1.1.2 for each specific approach and target field.

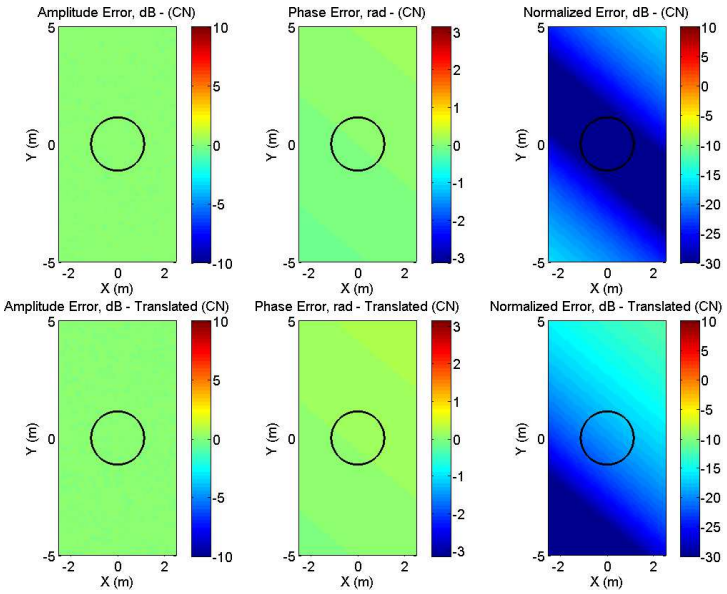


FIGURE 5.57: Acoustic errors, 1 plane wave, closest neighbour

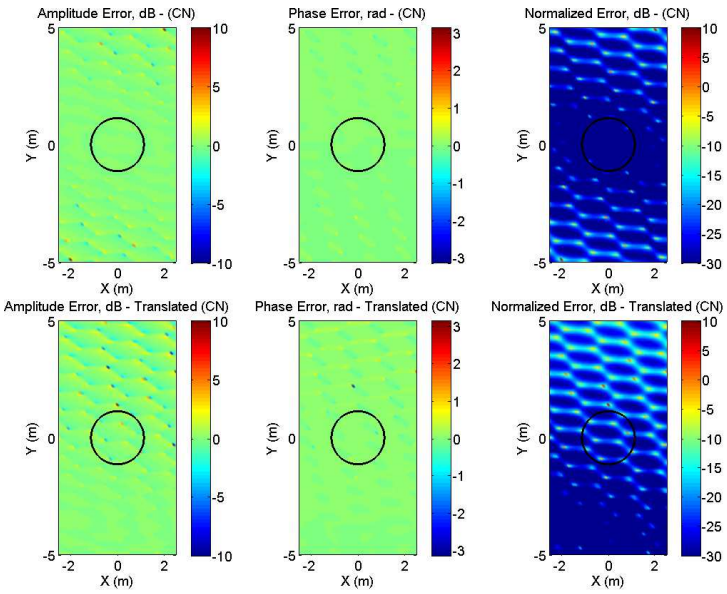


FIGURE 5.58: Acoustic errors, 3 plane waves, closest neighbour

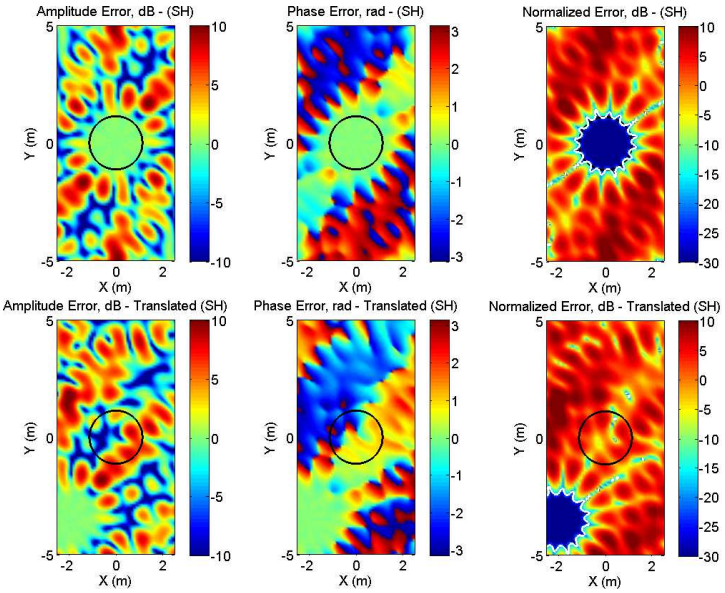


FIGURE 5.59: Acoustic errors, 1 plane wave, spherical harmonics

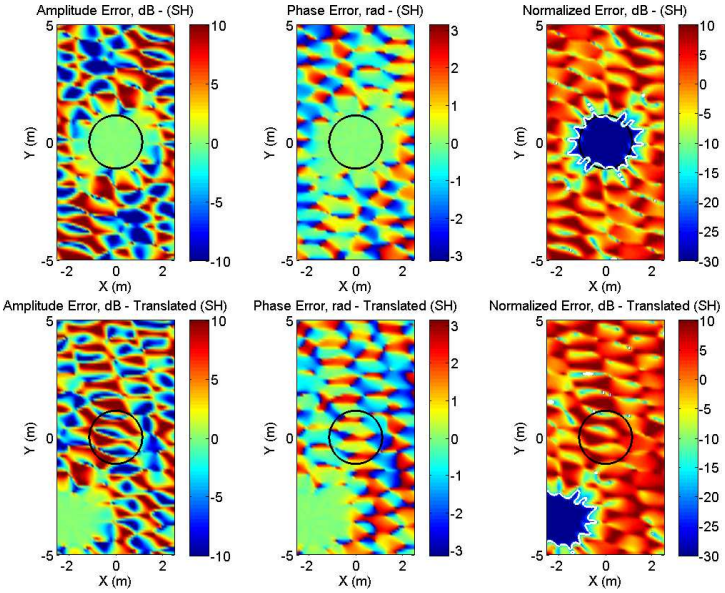


FIGURE 5.60: Acoustic errors, 3 plane waves, spherical harmonics

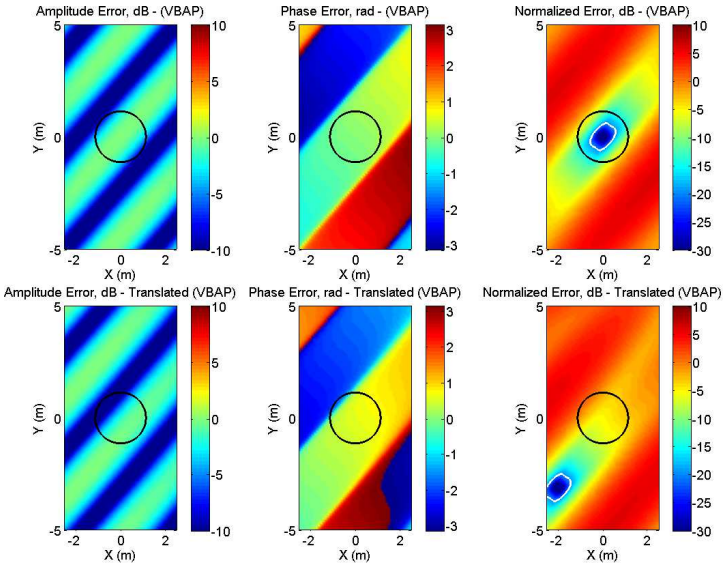


FIGURE 5.61: Acoustic errors, 1 plane wave, VBAP

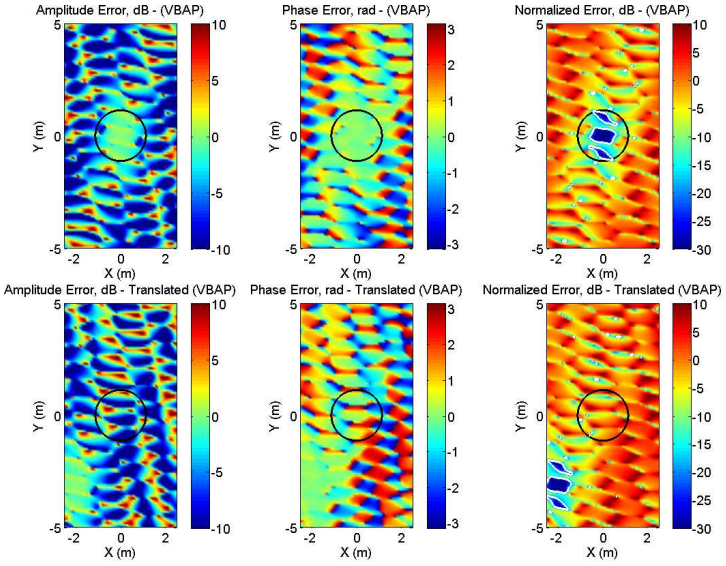


FIGURE 5.62: Acoustic errors, 3 plane waves, VBAP

The analysis of the acoustic errors suggests that the closest neighbour is the most suitable approach for the implementation of the translator operator. As expected, the spherical harmonic interpolation leads to an acoustic field whose accuracy is only correct in the radius of validity predicted by equation (4.12). This outcome implies that the translated field does not agree with the target field outside of that radius. Moreover, due to the dependency of equation (4.12) on frequency, the area of accurate reconstruction is very small at high frequencies. Regarding VBAP, the results suggest

that is the least accurate of the three approaches. This may be due to the interference effect generated by the representation of one plane wave as the interaction of 3 others.

An additional study was conducted to evaluate the consistency of the results in a more complex scenario. The plane wave expansion of 614 elements from the rectangular room was selected as reference. The three algorithms (closest neighbour CN, spherical harmonics SH, and VBAP) were applied to generate an equivalent plane wave expansion of 64 elements. Then, the sound field was calculated at different receiver positions using the translation operator (see Figure 5.17). Figures 5.63 to 5.66 illustrate the frequency response in 1/3 octave band resolution. The central point of the expansion is the receiver position 01.

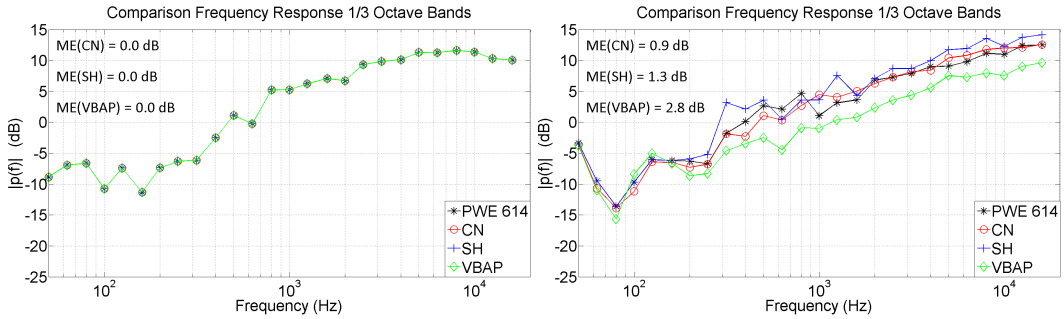


FIGURE 5.63: Comparison of the frequency response in 1/3 octave band resolution, PWE (614) CN, SH and VBAP, translated receiver position 1 and 2

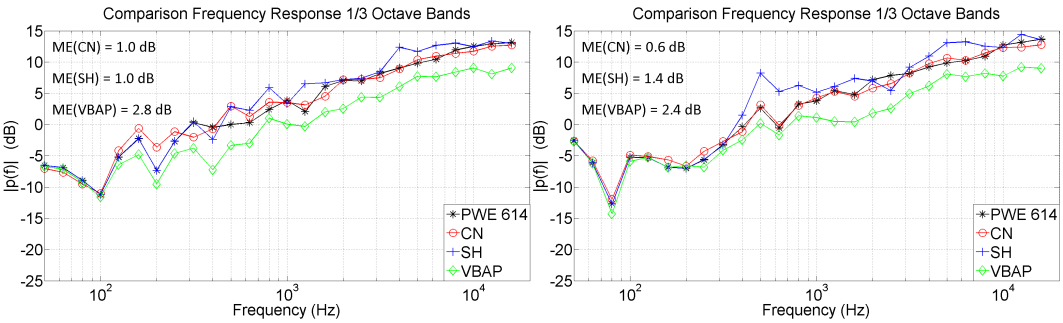


FIGURE 5.64: Comparison of the frequency response in 1/3 octave band resolution, PWE (614) CN, SH and VBAP, translated receiver position 3 and 4

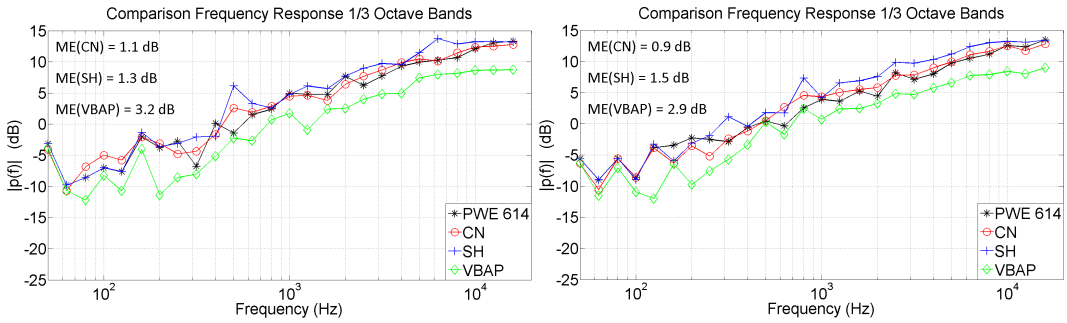


FIGURE 5.65: Comparison of the frequency response in 1/3 octave band resolution, PWE (614) CN, SH and VBAP, translated receiver position 5 and 6

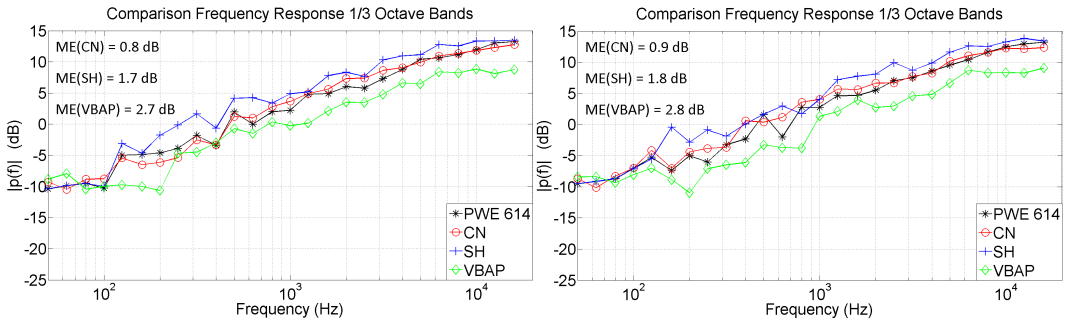


FIGURE 5.66: Comparison of the frequency response in 1/3 octave band resolution, PWE (614) CN, SH and VBAP, translated receiver positions 7 and 8

The results confirm that the closest neighbour is the most accurate algorithm. The worst agreement was found for VBAP for which lower energy than the reference is predicted at high frequencies at all the receivers. Further analysis of these outcomes is conducted in the next Chapter when an interactive auralization of different enclosures is discussed.

5.2 Rotation of the sound field

In this section two different alternatives are discussed that enable the listener to rotate their head. Similarly to the sound field translation operator, the rotation is not performed for the listener per se, instead, a rotation of the sound field is applied. Although a rotation of the acoustic field can be generated by interpolating HRTFs, this methodology has the limitation that it is restricted to binaural reproduction systems only. The two approaches considered in this section are suitable to be implemented with several audio reproduction techniques, which increases the

flexibility of the auralization system. Firstly, a spherical harmonic transformation is addressed. This is achieved by taking advantage the Jacobi-Anger expansion, which allows for the interoperability between the plane wave and spherical harmonic sound field representations. The second approach corresponds to the use of VBAP as an interpolation tool to determine a set of plane waves whose amplitudes synthesize a rotated sound field. Results of the accuracy of these two approaches as algorithms for the rotation of acoustic fields are discussed and numerical simulations are used to evaluate the suitability of the proposed approaches.

5.2.1 Spherical harmonic transformation as a rotation operator

Based on equation (2.56), the plane wave expansion can be represented as a linear superposition of spherical harmonics and spherical Bessel functions that allows for the rotation of the sound field in this domain. After the rotation is performed, a mode matching decoding approach (section 4.1.3) is implemented to return to the plane wave domain. The derivation of the sound field rotation operator in the spherical harmonic domain is presented as follows: the acoustic pressure at the point $\mathbf{r}$ is given by

$$p(r,\theta,\phi,\omega) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} A_{nm}(\omega) j_n(kr) Y_n^m(\theta,\phi). \tag{5.11}$$

Likewise, a rotation in the azimuth angle ϕ_0 can be expressed as

$$p(r,\theta,\phi-\phi_0,\omega) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} A_{nm}(\omega) j_n(kr) Y_n^m(\theta,\phi-\phi_0). \tag{5.12}$$

Expanding the right side of equation (5.12)

$$p(r,\theta,\phi-\phi_0,\omega) = \\ \sum_{n=0}^{\infty}\sum_{m=-n}^{n} A_{nm}(\omega) j_n(kr) \sqrt{\frac{(2n+1)}{4\pi}\frac{(n-m)!}{(n+m)!}} P_n^m(\cos\theta) e^{jm\phi} e^{-jm\phi_0}, \tag{5.13}$$

yields

$$p(r,\theta,\phi-\phi_0,\omega)=\sum_{n=0}^{\infty}\sum_{m=-n}^{n} j_n(kr)Y_n^m(\theta,\phi)A_{\phi_0 nm}(\omega), \tag{5.14}$$

in which

$$A_{\phi_0 nm}(\omega)=A_{nm}(\omega)e^{-jm\phi_0}. \tag{5.15}$$

Equation (5.15) indicates that the rotation of the sound field can be performed by taking the product between the complex spherical harmonic coefficients and a complex exponential whose argument depends on the angle of rotation. A decoding stage (section 4.1.3) can be implemented to return to the plane wave representation after the rotation has been carried out in the spherical harmonic domain. It is important to mention that apart from the initial error given by the discretization of the plane wave expansion, the truncation of the spherical harmonic series leads to additional errors in the sound field reconstruction. In order to keep the extension of the region in which the reconstruction is accurate, the same number of spherical harmonic coefficients as the number of plane waves must be implemented.

An evaluation of the rotation operator is presented below for different rotation angles (45°, 60° and 135°). For this, a plane wave expansion of 64 elements was used to synthesize the acoustic field generated by a single plane wave at 250 Hz coming from ($\theta = 90°, \phi = 45°$). The procedure for the encoding and decoding of the plane wave expansion in terms of spherical harmonics is described in section 4.1.3. Figures 5.67, 5.68 and 5.69 show the reconstructed acoustic fields for the reference and rotated cases. The three plots in the figures correspond to:

- (A) the reconstructed acoustic field generated by a plane wave expansion that synthesizes a single plane wave coming from the reference direction ($\theta = 90°, \phi = 45°$).
- (B) the reconstructed acoustic field generated by a plane wave expansion that synthesizes a plane wave coming from the desired rotated directions ($\theta = 90°, \phi = 90°, 135°, 180°$). These plane wave expansions are created to have a reference to compare the reconstructed fields that have been rotated using the rotation operator.

- (C) the reconstructed acoustic field generated by the plane wave expansion (A) whose complex amplitudes have been modified by the implementation of the rotation operator in the spherical harmonic domain. The complex amplitudes are calculated based on a mode-matching decoding approach.

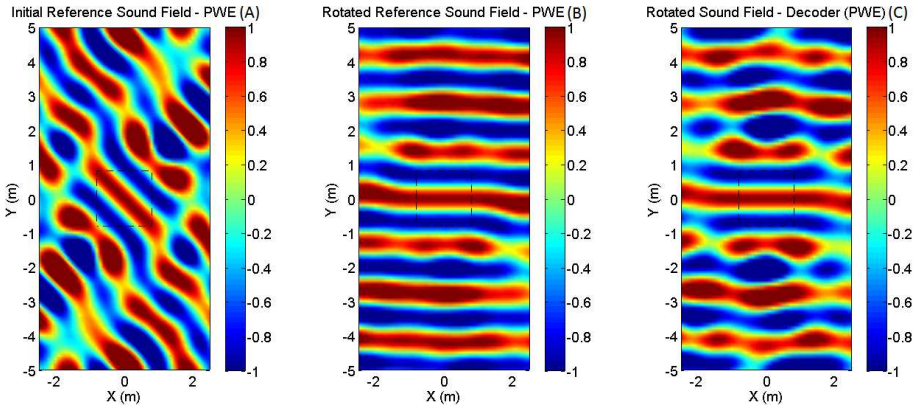


FIGURE 5.67: Rotation of the sound field, SH, 45° with respect to the initial incoming direction

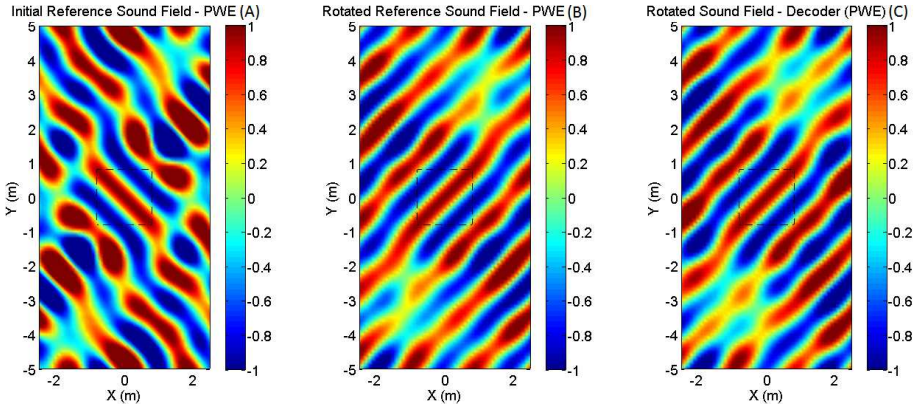


FIGURE 5.68: Rotation of the sound field, SH, 90° with respect to the initial incoming direction

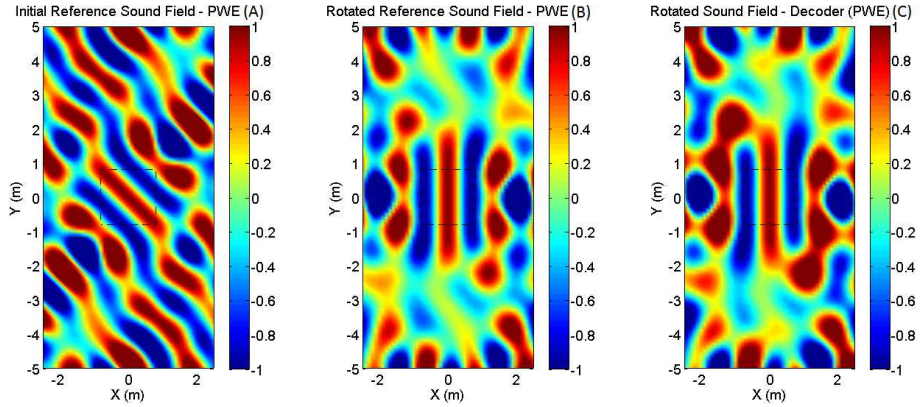


FIGURE 5.69: Rotation of the sound field, SH, 135° with respect to the initial incoming direction

These figures confirm the suitability of the spherical harmonic transformation to rotate the acoustic field. No relevant differences were found between the acoustic fields of the plots (B) and (C) close to the central point of the expansion, which indicates that rotation of the sound field using spherical harmonics does not affect the initial accuracy achieved by discretized plane wave expansion, namely the radius predicted by equation (4.12). However, this statement is true only if the number of spherical harmonic coefficients is equal to the number of plane waves ($(N+1)^2 = L$). Otherwise, the area of accurate reconstruction is reduced according to the number of spherical harmonic coefficients implemented. The acoustic errors defined in section 4.1.1.2 are presented in Figures 5.70, 5.71 and 5.72. The errors are calculated using as target field plane waves analytically estimated whose incoming directions correspond to the rotated angles.

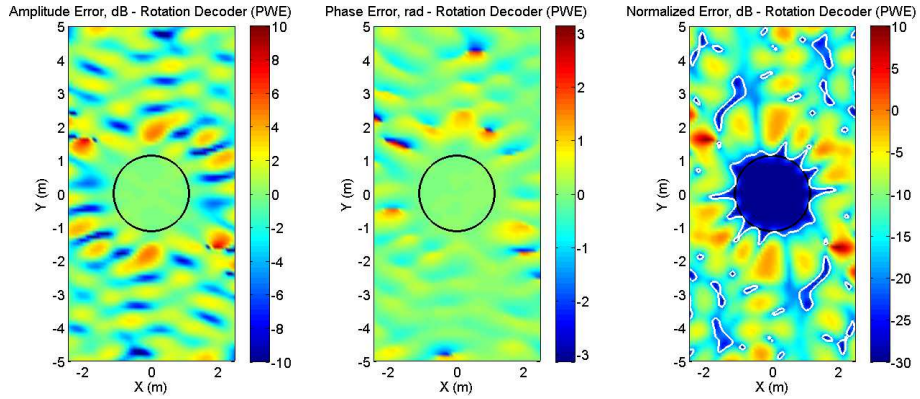


FIGURE 5.70: Acoustic errors, rotation of the sound field, SH, 45° with respect to the initial incoming direction

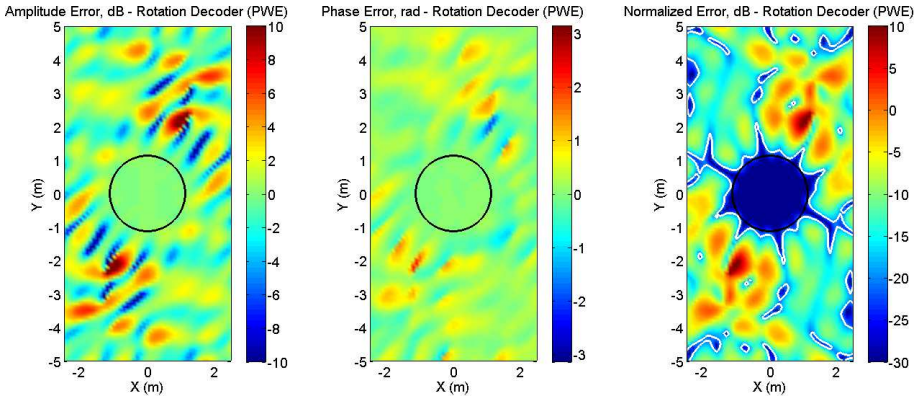


FIGURE 5.71: Acoustic errors, rotation of the sound field, SH, 90° with respect to the initial incoming direction

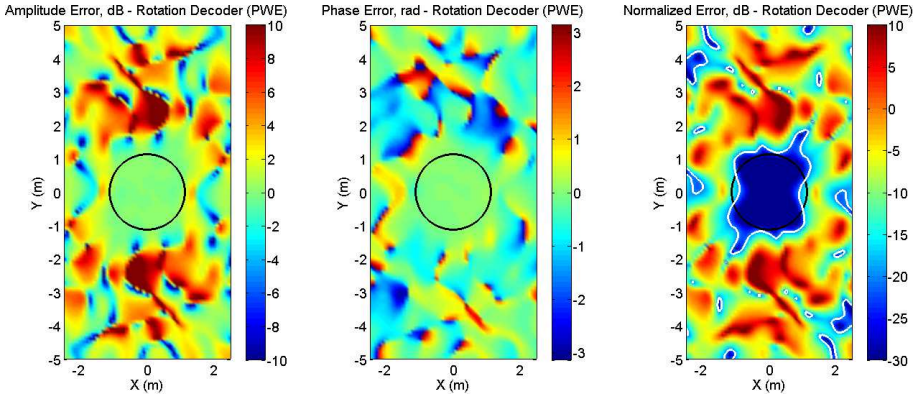


FIGURE 5.72: Acoustic errors, rotation of the sound field, SH, 135° with respect to the initial incoming direction

An analysis of the acoustic errors reveals that the rotation of the acoustic field leads to an accurate reconstruction of the target field within the radius predicted by equation (4.12). Outside this radius, the reconstruction leads to a different sound field reconstruction. Consistently with the results obtained for the translation operator, the mismatch between the reconstructed and target sound fields is not due to the rotation operator. Instead, the accuracy is determined by the number of planes waves used in the first instance to reconstruct the acoustic field. This is true provided that the number of spherical harmonic coefficients is equal to number of plane waves.

5.2.2 VBAP as a rotation operator

A different approach based on VBAP was implemented to perform the rotation of the acoustic field directly in the plane wave domain. The rotation is generated simply by shifting all the elements of the plane wave expansion in the opposite direction of the orientation of the listener. This rotated set of plane waves can be recreated using a VBAP algorithm where three different plane waves whose incoming directions are restricted to the directions established by the original discretized plane wave expansion (equation (5.8)) are used to generate each of them. The solution can be obtained by the implementation of an inverse method as follows: the directions of the three plane waves selected to synthesize a plane wave coming from (θ_{target}, ϕ_{target}) are established as unit vectors ($\mathbf{l}_n$) whose origin is the centre of a unit sphere. The direction of the target plane wave is defined by the unit vector $\tilde{\mathbf{l}}$. By expressing $\tilde{\mathbf{l}}$ as a combination of the three plane waves the following relation is determined

$$\tilde{\mathbf{l}} = (g_1\mathbf{l}_1 + g_2\mathbf{l}_2 + g_1\mathbf{l}_3)\,, \tag{5.16}$$

in which g_n represents the amplitude of the plane waves. A matrix notation yields

$$\tilde{\mathbf{l}} = \mathbf{L}\mathbf{g}, \tag{5.17}$$

where $\mathbf{L} = [\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3]$ and $\mathbf{g} = [g_1, g_2, g_3]^T$ .The amplitudes are calculated by solving equation (5.17) for $\mathbf{g}$, namely

$$\mathbf{g} = \mathbf{L}^{-1}\tilde{\mathbf{l}}, \tag{5.18}$$

In addition, the amplitudes are normalized based on a coherent summation where their sum leads to unity, namely

$$\mathbf{g_{normalized}} = \frac{\mathbf{g}}{g_1 + g_2 + g_3}, \tag{5.19}$$

This process is conducted for the 64 plane waves according to the shifting angle, which leads to a rotated acoustic field. Figures 5.73, 5.74, and 5.75 show the acoustic fields reconstructed by a plane wave expansion of 64 elements. The initial sound field corresponds to a plane wave at 250 Hz coming from ($\theta = 90°, \phi = 45°$). Then, the plane wave expansion is rotated by 45°, 60° and 135° with respect to the initial

incoming direction using the VBAP algorithm. The three plots illustrated in each figure correspond to:

- (A) the reconstructed acoustic field generated by a plane wave expansion that synthesizes a plane wave coming from the reference direction ($\theta = 90°, \phi = 45°$).
- (B) the reconstructed acoustic field generated by a plane wave expansion that synthesizes a plane wave coming from the desired rotated directions ($\theta = 90°, \phi = 90°, 135°, 180°$). These plane wave expansions are created to have a reference to compare the reconstructed fields that have been rotated using the VBAP algorithm.
- (C) the reconstructed acoustic field generated by the plane wave expansion whose complex amplitudes hwere modified based on the VBAP algorithm to perform the rotation.

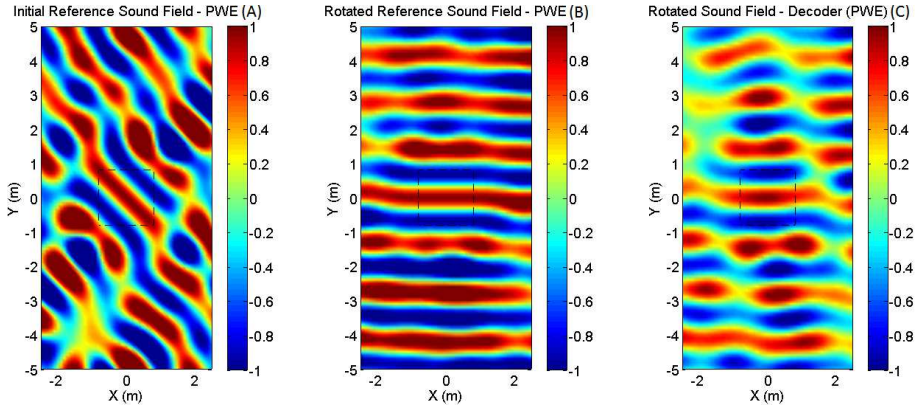


FIGURE 5.73: Rotation of the sound field, VBAP, 45°

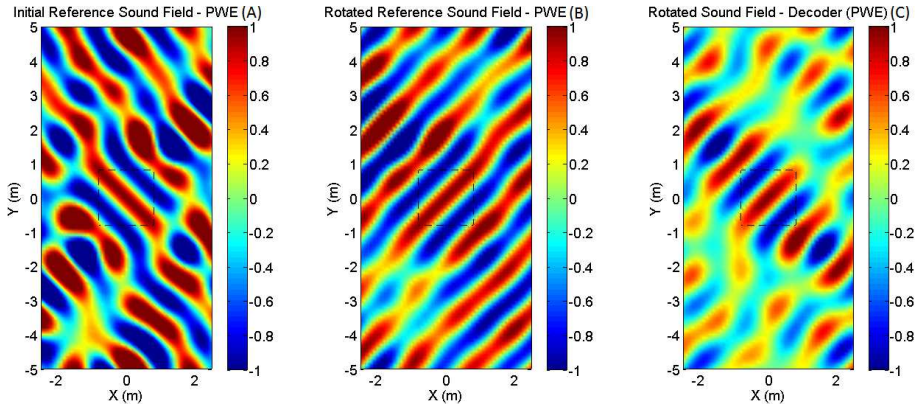


FIGURE 5.74: Rotation of the sound field, VBAP, 90°

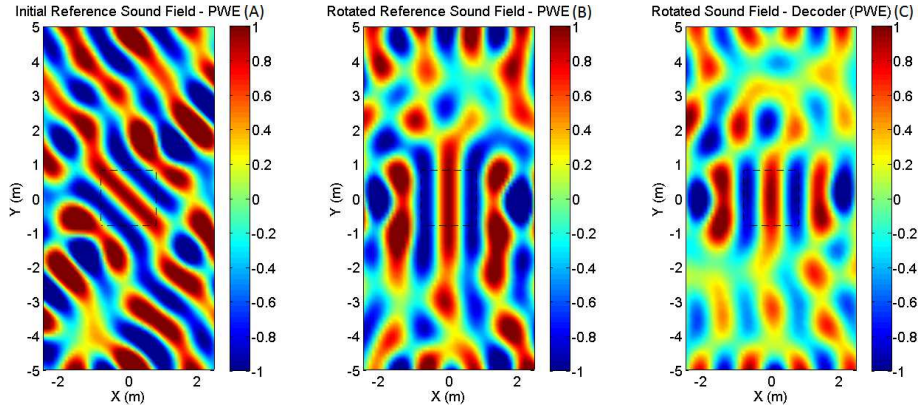


FIGURE 5.75: Rotation of the sound field, VBAP, 135°

The figures show that the use of VBAP is a suitable approach to perform the rotation of acoustic fields. A further analysis in terms of the acoustic errors is presented in Figures 5.76 to 5.78.

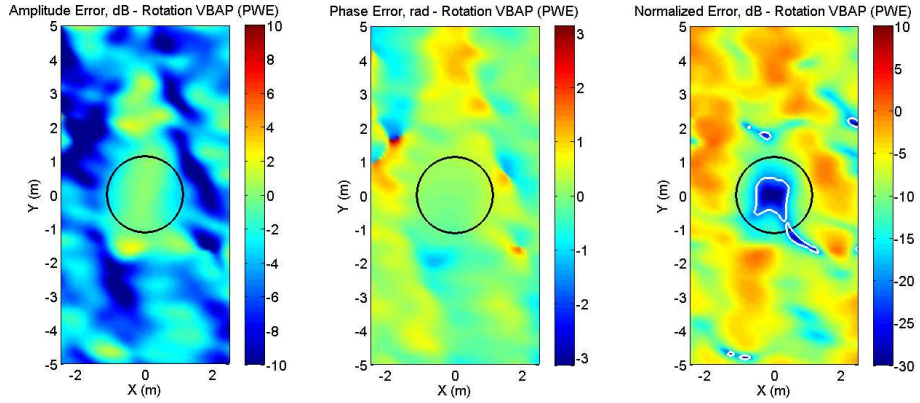


FIGURE 5.76: Acoustic errors, rotation of the sound field, VBAP, 45°

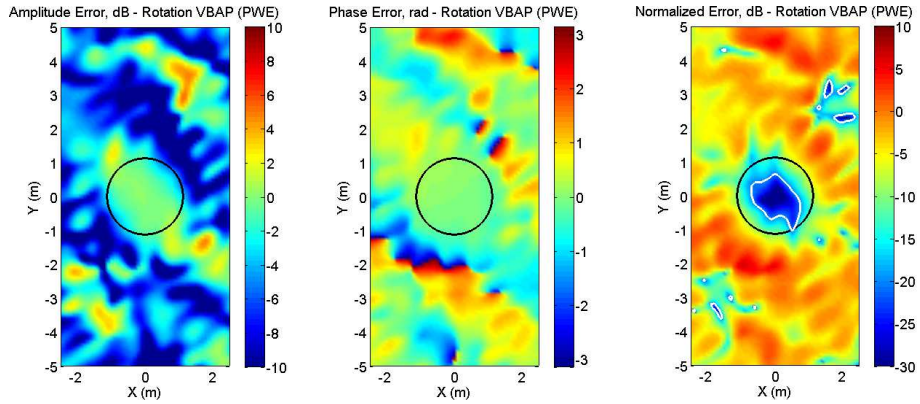


FIGURE 5.77: Acoustic errors, rotation of the sound field, VBAP, 90°

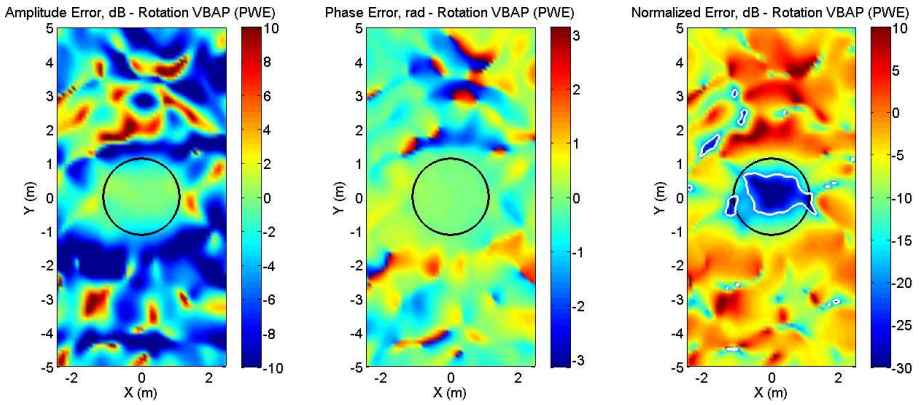


FIGURE 5.78: Acoustic errors, rotation of the sound field, VBAP, 135°

An analysis of the acoustic errors indicates that the area where the reconstruction is accurate does not match the radius predicted by equation (4.12). These findings suggest that the implementation of VBAP as a rotation operator is less accurate than the spherical harmonics approach. As frequency increases the area where the rotation using VBAP yields an accurate reconstruction decreases, which constrains furthermore its implementation at high frequencies. These findings reveal that a lower order of spherical harmonics is required to achieve the same accuracy as VBAP for use as rotation operators.

5.3 Discussion

Previous sections describe different approaches to generate interactive features in an auralization system based on a plane wave representation. The error given by the discretization of the plane wave expansion leads to a local description of the acoustic field, which in turn restricts the amount of translation that can be applied. Regarding rotation, the number of spherical harmonic coefficients is limited by the number of plane waves used in the expansion, and therefore, the accuracy at high frequencies is also determined by this discretization. Further analysis of the area where the rotation of the sound field has to be performed correctly is carried out in the next Chapter in the light of the implementation results.

The manipulation of the sound field to generate an interactive auralization based on a plane wave expansion has been discussed. The interactive features correspond to the translation and rotation of the listener within the enclosure. An analytical

expression for the translation of acoustic fields has been derived. From an implementation point of view, the use of a finite plane wave expansion generates an initial error that constrains the distance over which the translation accurately reconstructs a reference target field. For the low frequency content of the plane wave expansion (FE data), this maximum distance can be estimated from equation (5.9). In the case of the high frequency content of the plane wave expansion (GA data), the distance cannot be predicted by equation (5.9) due to the method used for the generation of the plane wave expansion. Nevertheless, experiments carried out using the translation operator indicate a good agreement in terms of the frequency response in 1/3 octave band resolution with respect to the omnidirectional references predicted by Catt-Acoustics (within a radius of 1.5 m).

The translation operator has been implemented to evaluate three different approaches to resample the GA plane wave expansion from 614 to 64 elements. This resampling is required only to make the approach compatible with the plane wave expansion from FEM. The results indicate that the closest neighbour is the most accurate technique to resample the data. The least accurate method is the use of VBAP.

Two different approaches have been proposed to perform the rotation of the acoustic field. An implementation of VBAP as an interpolation tool validates the suitability of this method to rotate sound fields in the plane wave domain. However, a comparison with a rotation algorithm based on a spherical harmonic transformation reveals that the latter approach is more accurate in terms of sound field reconstruction. Nevertheless, the number of spherical harmonic coefficients must be the same as the number of plane waves to preserve the area in which the reconstruction is accurate.

Chapter 6

Real-Time Implementation of an Auralization System

The aim of this chapter is to implement the theoretical background described in previous chapters in order to create an interactive auralization system based on the plane wave expansion. The interactive auralization is performed based on a direct acoustic rendering (see section 2.2.5.2) meaning that the directional impulse responses, which constitute the plane wave expansion, are previously calculated (off-line). This methodology allows for a real-time acoustic rendering of enclosures using synthesized directional impulse responses whose low frequency and high frequency components have been calculated in advance using FEM and GA, respectively. Due to the precomputation of the impulse responses, the proposed auralization system enables interactive features such as translation and rotation of the listener within the enclosure. However, changes in the boundary conditions or modifications of the acoustic source in terms of its directivity and spatial location are not feasible without recalculating the RIRs. The proposed auralization system combines a real-time acoustic rendering with a graphical interface based on a video game environment. Experiments about the accuracy and limitations of the proposed approach are presented and discussed for several reference cases on the basis only of the translation operator.

Figure 6.1 shows the general architecture of the signal processing chain. It is composed of 4 main blocks. The first block corresponds to the convolution of anechoic audio material with a plane wave expansion that has been calculated from FE-GA simulations (Chapter 4). The second module refers to the implementation of the translation operator in the plane wave domain (section 5.1). The third block corresponds to the application of a rotation operator in the spherical harmonic domain

(section 5.2.1) and finally, the last stage is the reproduction of the auralization using a headphone-based binaural system. The implementation has been made using the commercial packages Max v.7.2 and Unity v.5.0. Max is a visual programming language oriented to audio and video processing. In contrast, Unity is a programming language dedicated to video game development and it is used in the current research to create graphical interfaces for interactive auralization.

In the following two general setups are presented. The first one corresponds to the auralization at a fixed position and the other one is for an interactive sound field reconstruction. Each setup is described individually based on the blocks illustrated in Figure 6.1.

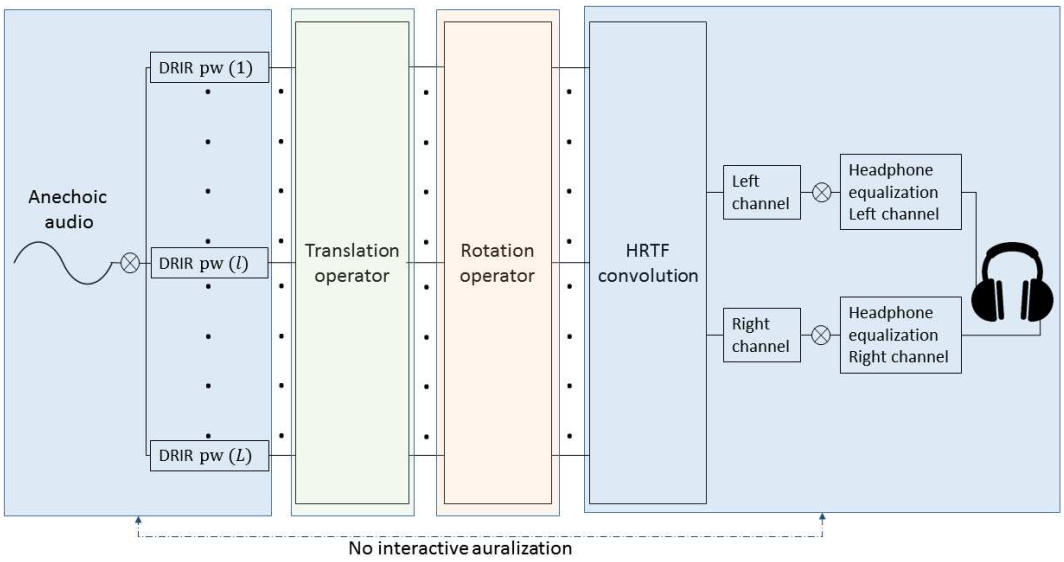


FIGURE 6.1: General architecture for a real-time auralization based on the plane wave expansion

6.1 Auralization at fixed listener positions

Figure 6.1 shows two blocks highlighted in blue which correspond to the components used for auralizations at a fixed position. The acoustic field is reproduced using a headphone-based binaural system. The first block corresponds to the estimation of the directional impulse responses based on plane wave expansion. For this, numerical simulations of the acoustics of enclosures are carried out using a combination of the finite element method and geometrical acoustics as explained in Chapter 3.

Then, the plane wave expansion is generated from the simulated data following the methodology proposed in Chapter 4. To reduce the number of convolutions required in real-time, the anechoic audio material is previously convolved with the directional impulse responses. The only limitation of this approach is that the audio material cannot be changed in real-time. Regarding the second block, the convolution with the HRTFs can be performed offline but in the current implementation is executed in real-time. Each directional impulse response is convolved with the HRTF that corresponds to the incoming direction given by the plane wave expansion. The implemented HRTF database was provided by Technology Arts Sciences TH Koln [146] and corresponds to a Neumann dummy head KU100.

A non-individualised headphone equalization is applied at the end of the signal processing chain to reduce the frequency colouration produced by the transducers in the reproduced sound. The headphone equalization was made based on the methodology proposed by Masiero and Fels [147]. For this, measurements of Headphone Impulse Responses (HPIRs) from an AKG K-702 were conducted using a Neumann dummy head KU100 in the small anechoic chamber of the ISVR (see Figure 6.2). The HPIRs were measured using a logarithmic sine sweep from 20 Hz to 20 kHz with a duration of 10 s. Because the HPIRs vary according to how the headphones fit on over the listener, a set of 10 trials was carried out by removing and placing the headphones again for each measurement. Figure 6.3 shows the measured headphone frequency responses for both channels (left and right). 0 dB corresponds to the magnitude measured at 1 kHz.

FIGURE 6.2: Measurement of headphone frequency responses, AKG K-702

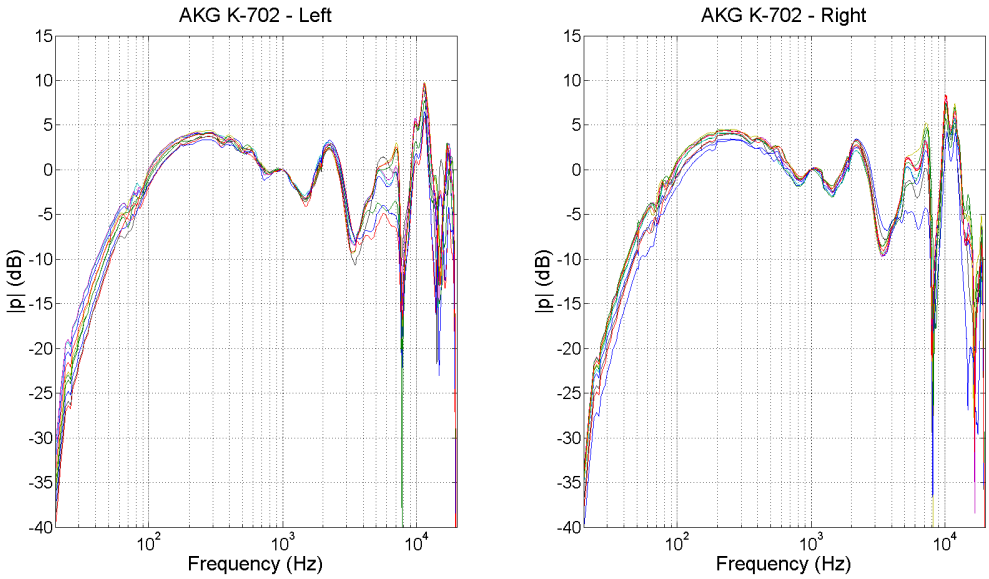


FIGURE 6.3: Measured headphone frequency responses, AKG K-702

The assumption made in Masiero's approach is that peaks in the inverted frequency response are the most subjectively disturbing component. Therefore, the aim is to reduce the notches in the frequency response to be inverted and thus minimize the resulting peaks of the inverted one. One approach is to take the highest magnitude among the measured headphone frequency responses at each frequency, which leads to all the local notches being reduced. This approach yields an inverted frequency response whose irregularities will most probably be in the form of valleys and not peaks.

However, instead of taking the maximum of the magnitude corresponding to all the measurements for every frequency, a more conservative approach based on a statistical analysis was used in this work. Assuming that the measured headphone impulse responses are normally distributed, the curve obtained from the mean μ plus 2 times the standard deviation σ of the magnitude of the measured frequency responses will be above the magnitude of the measured frequency responses with over 95% chance. This approach yields a relatively smooth curve and it is more robust to outliers compared to taking the maximum of the magnitude of all measurements [147]. Finally, the phase is synthesized again using minimum phase filters. Figure 6.4 illustrates the target frequency responses to be inverted for both channels. The inverted frequency responses are calculated as

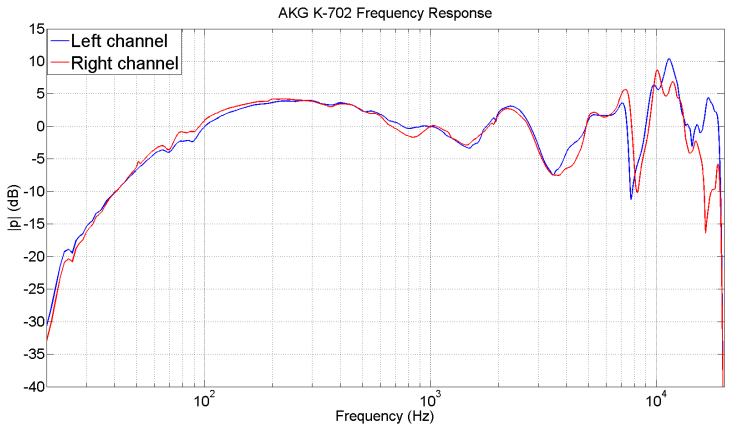


FIGURE 6.4: Target headphone frequency responses to be equalized, AKG K-702

$$\text{inv}(H(\omega)) = \frac{H(\omega)^*}{H(\omega)^* H(\omega) + \beta}, \tag{6.1}$$

in which $\text{inv}(H(\omega))$ is the inverted frequency response, $(\cdot)^*$ indicates the complex conjugate operation and β is a regularization parameter implemented to reduce the boost produced by the inversion at low frequencies. Figure 6.5 shows the frequency responses and their inverted versions. The impulse responses of the inverted filters are illustrated in Figure 6.6.

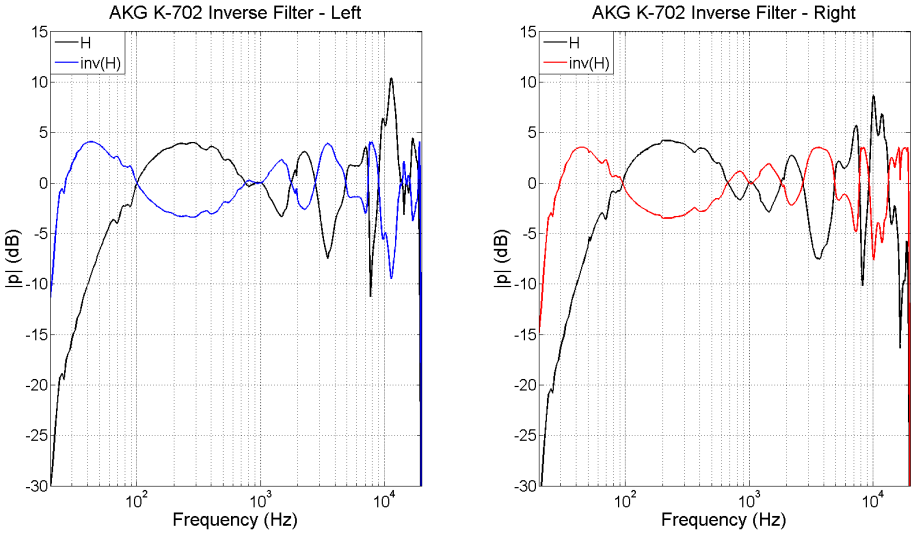


FIGURE 6.5: Inverted headphone frequency responses, AKG K-702

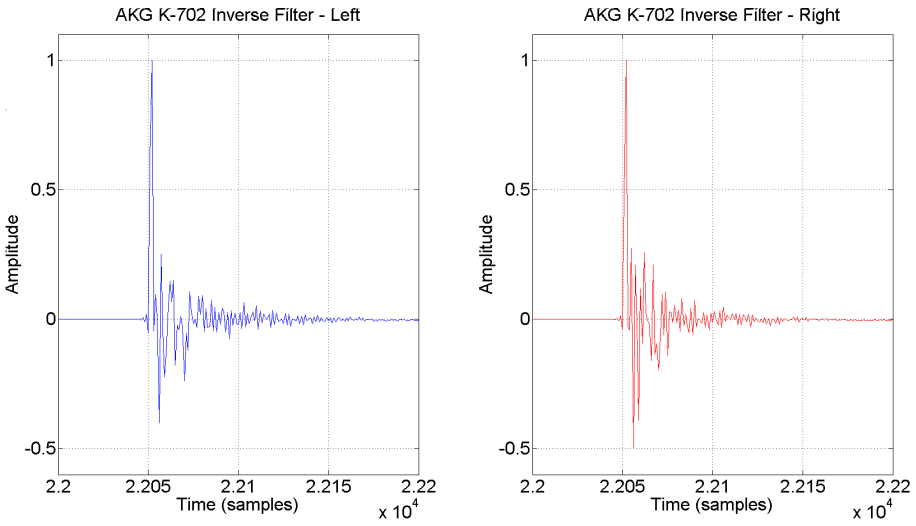


FIGURE 6.6: Inverted headphone filters, AKG K-702, $\beta = 10^7$

The filters are convolved with the HRTFs at the end of the signal processing chain of the auralization system.

6.2 Interactive auralization

All the modules in Figure 6.1 are now included in the auralization system to generate an interactive sound field reconstruction. The translation of the acoustic field is performed in the plane wave domain based on the procedure presented in section 5.1. The spherical harmonic approach was selected for the rotation because it provides a higher accuracy than VBAP (see section 5.2). The plane wave expansion used for the auralization corresponds to 64 directional impulse responses whose low and high frequency content have been calculated using FEM and GA, respectively. A detailed description of the Max patches used for the auralization system is presented in Appendix A.

6.2.1 Translation of the acoustic field

The translation of the sound field is carried out by the implementation of delay objects in Max. The values of the delays are calculated according to the relative position of the listener inside the virtual enclosure. In the current instance, the translation is performed only in the azimuthal plane. The values of the delays are given by

$$D_{(l)} = \left\lceil \left[-f_s \left(\frac{\mathbf{x}' \cdot \widehat{\mathbf{y}}_l}{c} \right) \right] + \zeta \right\rceil , \tag{6.2}$$

where $D_{(l)}$ is the delay applied to the lth directional impulse response, f_s is the sampling frequency, c is the speed of sound, $\mathbf{x}'$ is a vector identifying the coordinates of the listener position, $\widehat{\mathbf{y}}_l$ is the unitary vector that indicates the incoming direction of the plane wave l, ζ is an arbitrary constant to prevent the occurrence of negative delay values and $\lceil \cdot \rceil$ indicates the rounding operation. Figure 6.7 shows the general delay algorithm in Max.

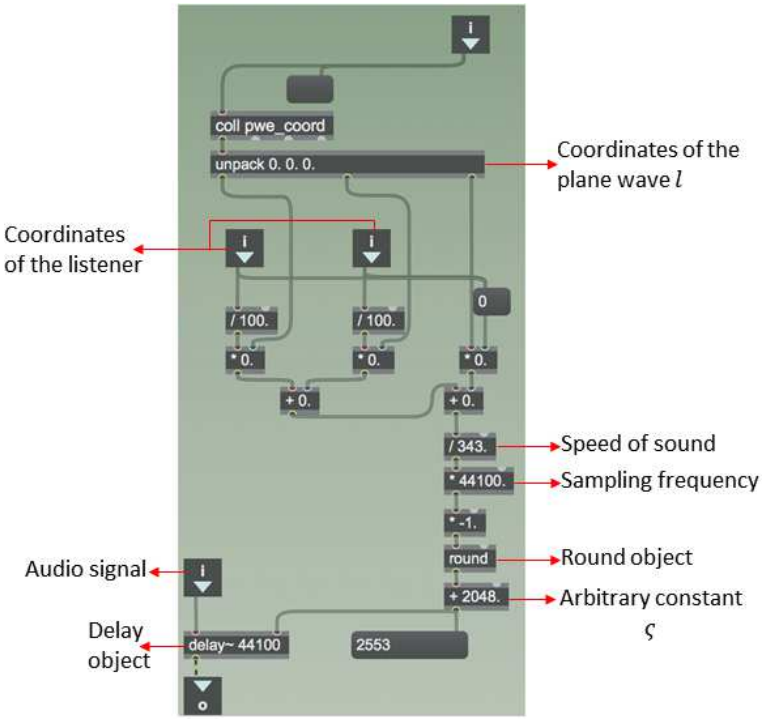


FIGURE 6.7: General delay algorithm in Max

6.2.2 Rotation of the acoustic field

The implementation of the rotation in the auralization system using a spherical harmonic transformation can be divided into two main steps: the encoding and decoding stage (see section 4.1.3), respectively. The general concept can be defined as the encoding of each of the directional impulse responses into a finite number of spherical harmonic coefficients, the application of the rotation and, finally, the return to the plane wave domain again through a decoding process. Nevertheless, to reduce the computational cost required by the generation of the rotation in the

acoustic field, instead of multiplying the spherical harmonic coefficients by a rotation operator, the encoding stage is computed using as an input the difference given by the angles of the directional impulse responses ((θ_l, ϕ_l)) and the rotation angle. Two different approaches to compute the encoding and decoding stages were considered as shown in Figures 6.8 and 6.9. In the first case, the encoding of the directional impulse responses includes the multiplication of the spherical harmonic coefficients with the audio signals, which allows for all the coefficients of the same mode and order to be added together, and therefore, the use of only one decoding block. In contrast, in the second approach, just the directions of the directional impulse responses are encoded leading to a unique decoding stage for each plane wave.

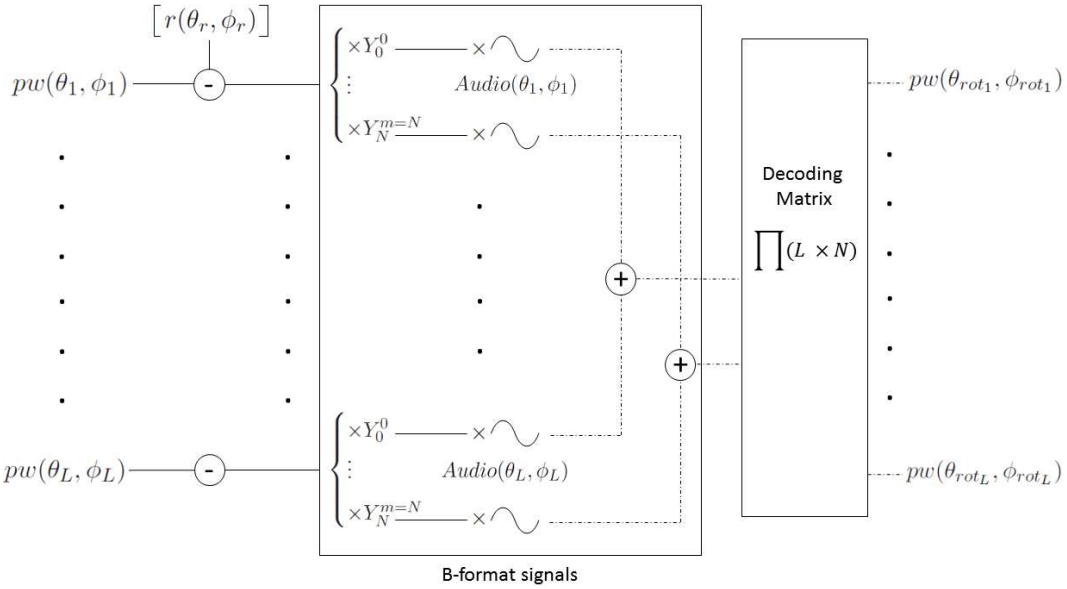


FIGURE 6.8: Encoding and decoding using spherical harmonics, approach 1

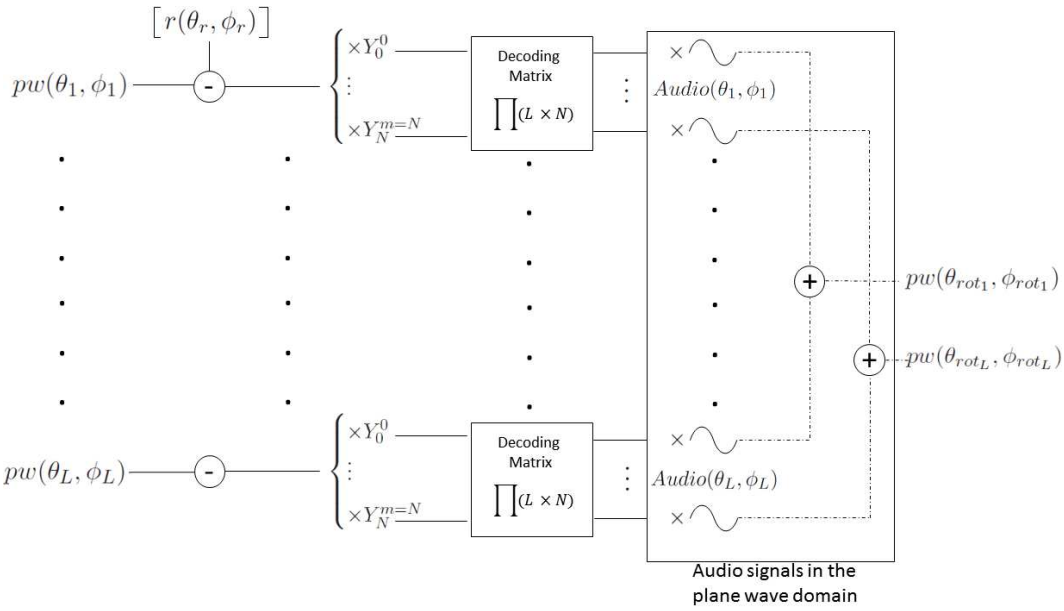


FIGURE 6.9: Encoding and decoding using spherical harmonics, approach 2

Although the second approach requires more mathematical operations, the rate of update is lower because the multiplication with the audio signals is performed at the last step of the process. Experiments conducted in Max revealed that approach 2 uses approximately 30% less computational resources than approach 1 and for that reason it has been implemented in the current research. However, the number of audio channels to be processed at some stage of the signal processing chain is $(L \times N)$ regardless of the approach. This was not possible to achieve in Max without the creation of an external and more efficient object due to the very large amount of operations. The processing of this number of audio channels in Max would involve the development of external tools and is not considered in this investigation. Based on the discussion above, the number of spherical harmonic coefficients in equation (4.25) was limited to the fifth order of spherical harmonics (36 coefficients) and this has been implemented directly in Max for the encoding and decoding stages. The encoding is performed using real-valued spherical harmonics [148] which are calculated from their complex pairs using equation (2.58).

The decoding stage is performed using a mode-matching approach. One consequence of the reduced spherical harmonic order is that the area of accurate reconstruction is reduced following the relation $N = kr$. In order to preserve the area where the

translation is correct, the translation operator is applied every time before the rotation. In this case, the reduction given by the lower spherical harmonic order does not reduce the region where the translation operator accurately reconstructs the target field. This means that the outcome of the rotation operation has only to be accurate in the radius corresponding to the listener's head (r_l). The truncation of the spherical harmonic series up to 5th order leads to a maximum frequency of $\approx$ 2 kHz for a listener radius of 0.1 m. Above that frequency, the sound field reconstruction is not performed accurately and alternative Ambisonics approaches to optimize the high frequency content may be used. A complete description of the implementation of the rotation in Max can be found in Appendix A. Furthermore, Appendix B reports the results of the evaluation of different Ambisonics decoding methods, which may be used to improve the high frequency sound field reconstruction.

6.3 Graphical interfaces

Virtual environments were created in Unity to generate a platform where the listener can move using a first-person avatar and hear the changes in the acoustic field based on its relative position with respect to the enclosure. This is performed by sending from Unity to Max the location and orientation of the avatar. The interaction between these two softwares was achieved using the Max-Unity Interoperability Toolkit [149], which is an external Max object developed by the Virginia Tech Department of Music. Figures 6.10 to 6.13 illustrate the models made in Unity to generate the interactive auralizations. The models correspond to 4 different spaces: a virtual rectangular room of dimensions 5m x 10m x 3m and the enclosures considered in Chapter 3.

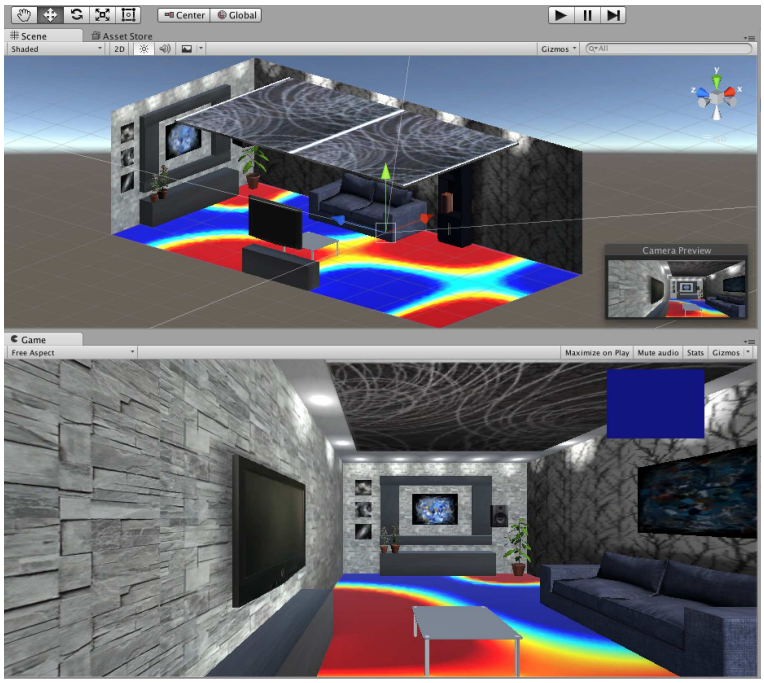


FIGURE 6.10: Rectangular room created in Unity

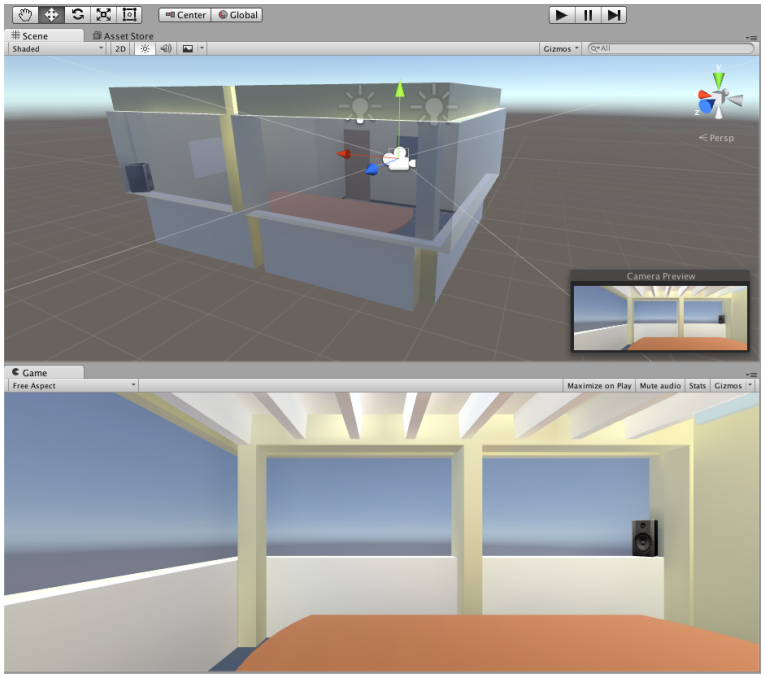


FIGURE 6.11: Meeting room model created in Unity

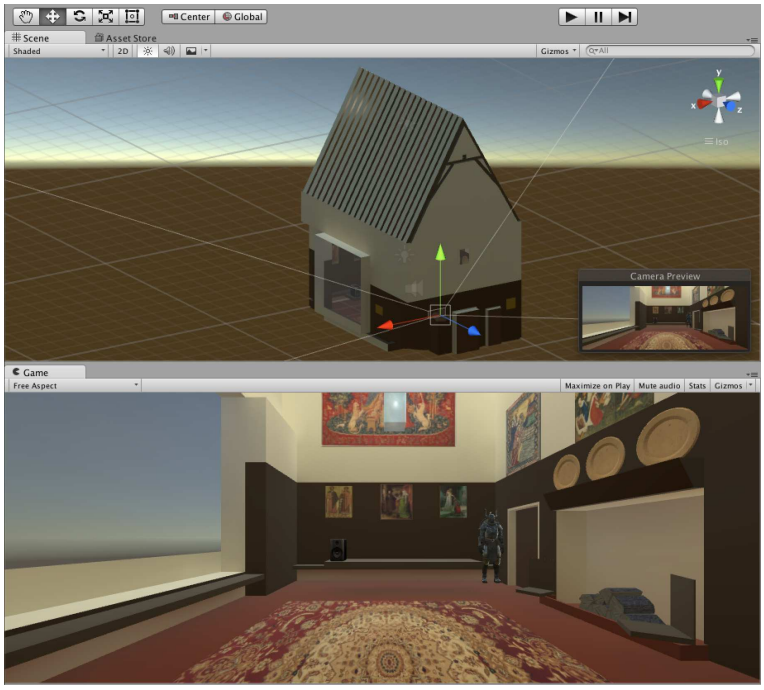


FIGURE 6.12: Ightham Mote model created in Unity

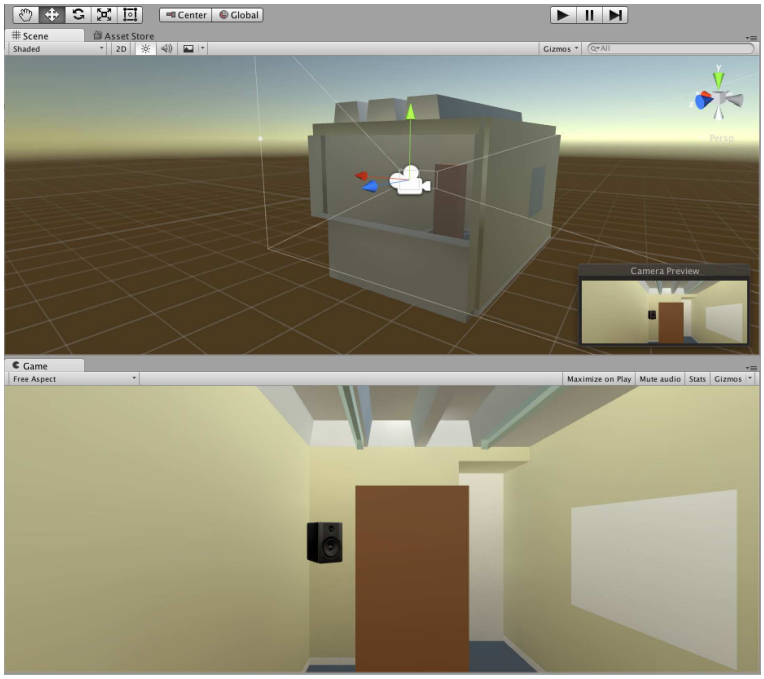


FIGURE 6.13: Office room model created in Unity

6.4 Evaluation of the auralization system using the Max implementation

The aim of this section is to present a series of experiments to assess the accuracy of the sound field reconstruction given by the auralization system presented in section 6.2. For this, the real-time implementation in Max was used in conjunction with plane wave expansions of the different rooms, which have been predicted by means of the finite element method and geometrical acoustics. Two types of analysis have been performed: monaural (based on omnidirectional signals) and a spatial evaluation based on first order B-format signals, respectively. The procedure is based on the recording of the output signals from the real-time auralization system implemented in Max and their comparison to numerical references.

6.4.1 Monaural analysis

A comparison of the predicted omnidirectional frequency responses at different receiver positions is considered for different enclosures. This is performed by rendering the sound field in real-time using an auralization system developed in Max whose directional impulse responses were estimated based on the methodology established in Chapter 4. The omnidirectional frequency responses from the interactive auralization system were obtained by adding all the directional impulse responses and recording the total output. This information is compared to omnidirectional frequency responses that were synthesized individually at the receiver locations using a combination of the finite element method and geometrical acoustics. These omnidirectional references do not use the plane wave expansion information. They correspond to the frequency response of omnidirectional receivers obtained directly from the commercial packages Comsol and Catt-Acoustics, respectively. The use of the numerical information as reference is due to the lack of measurements across the enclosures to evaluate different positions. The assumption given by this approach is that the numerical simulations provide an adequate description of the acoustics of the rooms. A calibration between reference measurements and simulations has been presented in Chapter 3, which makes this assumption reasonable. Nevertheless, it is important to emphasize that the main objective of this research is not to prove that a combination of FEM and GA leads to an accurate prediction of room impulse response, but how to generate a plane wave representation of FEM-GA simulations,

which allows for the generation of an interactive auralization. Four spaces have been taken into consideration for the analysis as described in section 6.3.

6.4.1.1 Rectangular room

The positions within the room were selected based on the setup illustrated in Figure 5.17. Figures 6.14 to 6.21 show the frequency response in narrow band and in 1/3 octave band resolution at 8 receiver locations based on the method selected to resample the GA data, i.e. closest neighbour, spherical harmonic interpolation and VBAP. The cyan line indicates the crossover frequency (355 Hz) and the black line indicates the predicted maximum frequency at which the translation should be correctly performed. The maximum frequency was estimated by solving equation (4.12) for R. This is done taking into account the distance between the central point of the expansion and each receiver's position. The data plotted in magenta corresponds to the initial plane wave expansion of 614 elements that contains only geometrical acoustic information. This plot is only presented in the frequency range in which the GA data is used. The use of this information is to point out the initial divergence between the plane wave expansion and the omnidirectional receiver predicted by Catt-Acoustics. Finally, Table 6.1 presents a summary of the mean errors according to the approach used to resample the GA data from 614 to 64 plane waves.

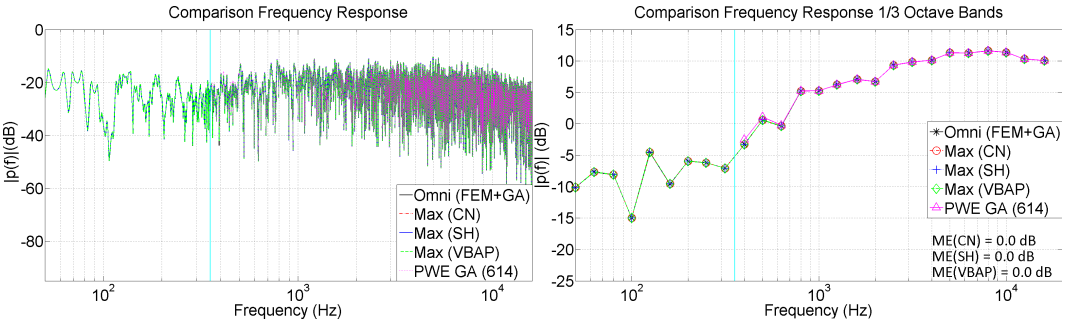


FIGURE 6.14: Comparison of frequency responses. PWE (FEM+GA) and omnidirectional RIR (FEM+GA) at the reference position 1, rectangular room

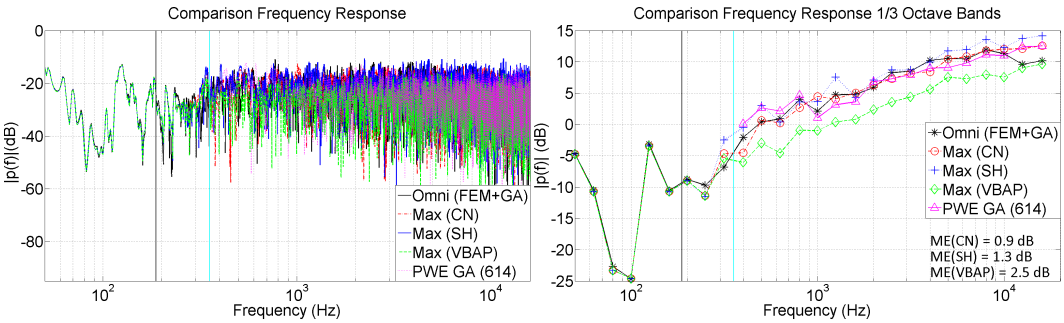


FIGURE 6.15: Comparison of frequency responses. Translated PWE (FEM+GA) and omnidirectional RIR (FEM+GA) at the translated position 2, rectangular room

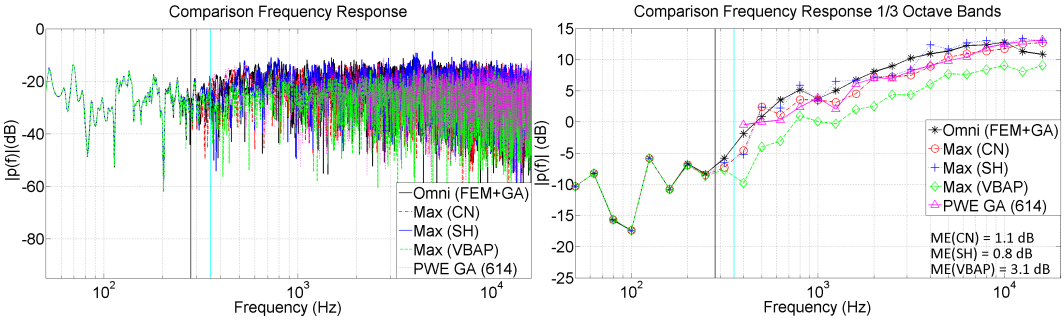


FIGURE 6.16: Comparison of frequency responses. Translated PWE (FEM+GA) and omnidirectional RIR (FEM+GA) at the translated position 3, rectangular room

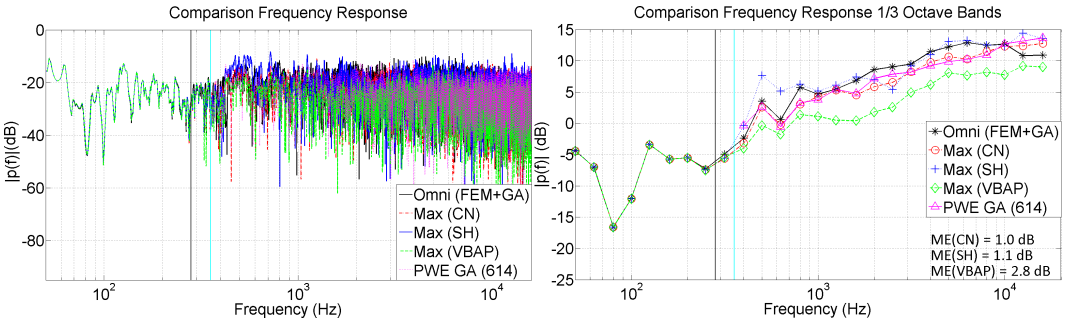


FIGURE 6.17: Comparison of frequency responses. Translated PWE (FEM+GA) and omnidirectional RIR (FEM+GA) at the translated position 4, rectangular room

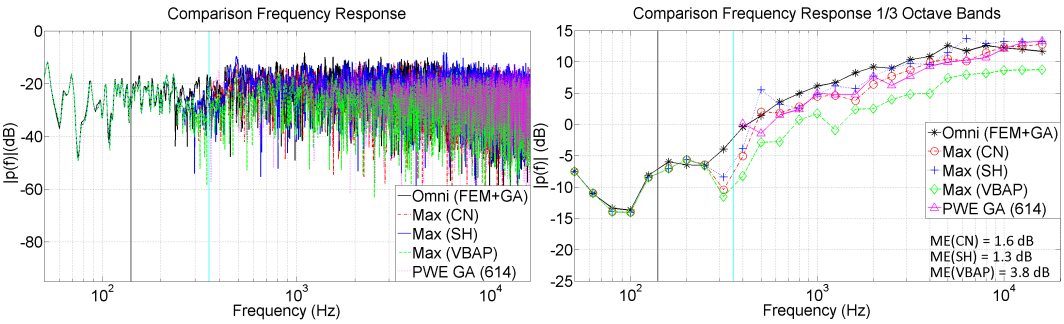


FIGURE 6.18: Comparison of frequency responses. Translated PWE (FEM+GA) and omnidirectional RIR (FEM+GA) at the translated position 5, rectangular room

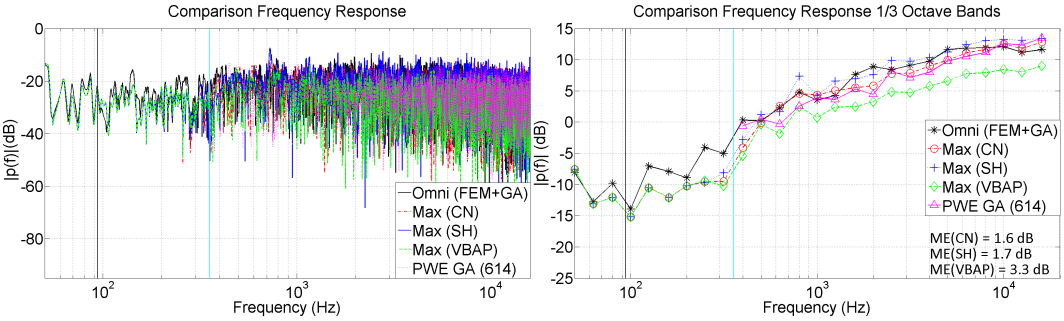


FIGURE 6.19: Comparison of frequency responses. Translated PWE (FEM+GA) and omnidirectional RIR (FEM+GA) at the translated position 6, rectangular room

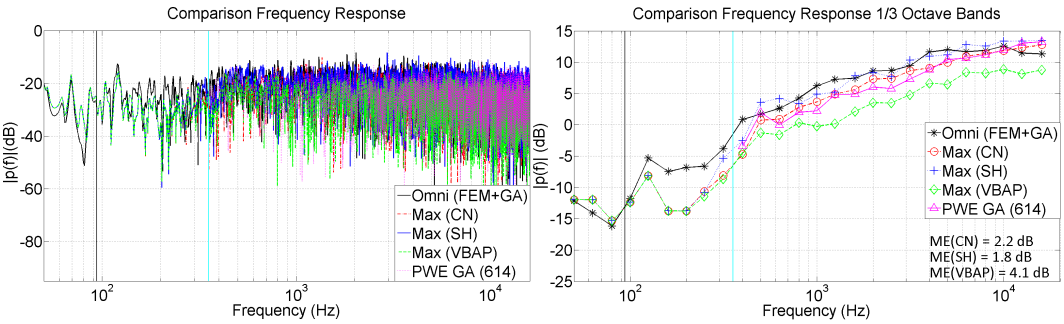


FIGURE 6.20: Comparison of frequency responses. Translated PWE (FEM+GA) and omnidirectional RIR (FEM+GA) at the translated position 7, rectangular room

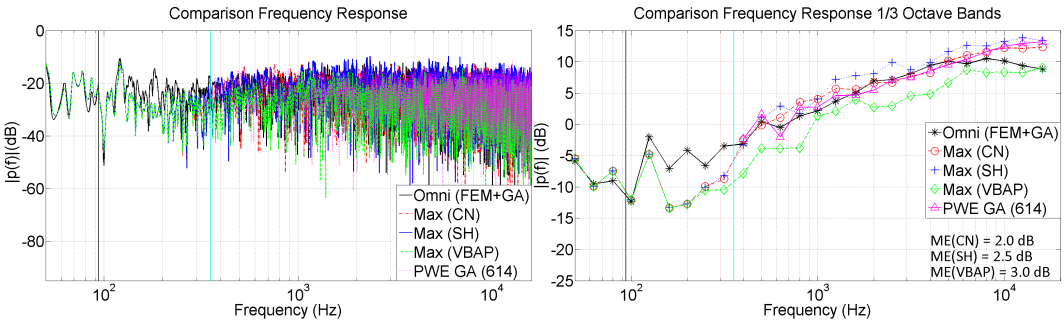


FIGURE 6.21: Comparison of frequency responses. Translated PWE (FEM+GA) and omnidirectional RIR (FEM+GA) at the translated position 8, rectangular room

Receiver	Dist. (m)	Freq. (Hz)	CN (dB)	SH (dB)	VBAP (dB)
2	1.5	≈ 187	0.9	1.3	2.5
3	1	≈ 281	1.1	0.8	3.1
4	1	≈ 281	1.0	1.1	2.8
5	2	≈ 141	1.6	1.3	3.8
6	3	≈ 94	1.6	1.7	3.3
7	3	≈ 94	2.2	1.8	4.1
8	3	≈ 94	2.0	2.5	3.0

TABLE 6.1: Mean errors for the interactive auralization at different receiver locations, rectangular room. The distance to the central point of the expansion and the maximum frequency in which is expected to achieve an accurate reconstruction is reported for each receiver

The results indicate a good agreement to the reference data below the frequency predicted by equation (4.12). Smaller differences found in this frequency range are associated with three reasons: the implementation of integer delays in Max, the numerical accuracy used by Max to perform mathematical operations (summing the directional impulse responses), and the application of regularization in the inverse problem, which decreases the matching between the radius of validity and the effective area where the reconstruction is accurate. Above the crossover frequency, a better agreement in terms of energy was found for the closest neighbour and spherical harmonic approach. Based on Table 6.1, the accuracy of the closest neighbour and the spherical harmonics methods depends on the spatial location of the receiver and there is not a clear difference when using one or the other. Consistently with previous results, VBAP was the least accurate method to represent the GA data. An important mismatch in the frequency range between 160 Hz and 315 Hz was found in Figures 6.19, 6.20 and 6.21. This outcome is related to the situation where the

frequency predicted by equation (4.12) is much lower than the crossover frequency. In this case, there is no control over the energy in this frequency range because the translation operator is applied on the low frequency content, which corresponds to the FE data, and the distance is much further from the maximum region where the translation leads to the correct sound field reconstruction.

6.4.1.2 Meeting room

Eight receivers were selected according to the description given in Figure 5.28. The predicted frequency response in narrow band and in 1/3 octave band resolution for each receiver are illustrated in Figures 6.22 to 6.29. The cyan line indicates the crossover frequency (355 Hz) and the black line indicates the maximum frequency in which is expected that the translation is correctly performed. Finally, Table 6.2 shows the mean errors according to the approach used to resample the GA data.

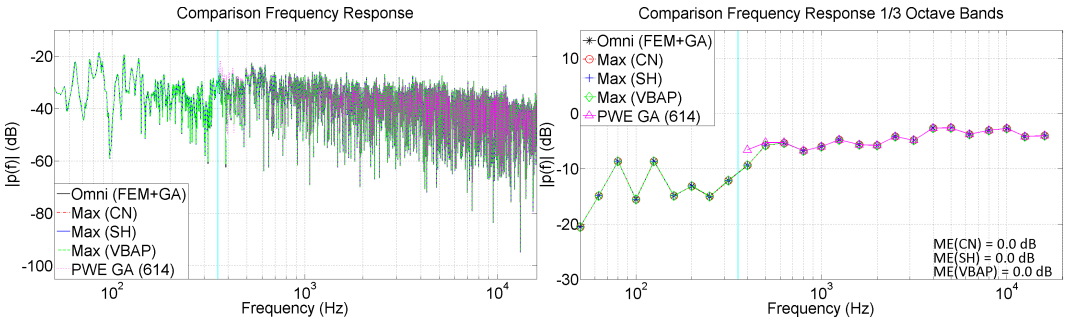


FIGURE 6.22: Comparison of frequency responses. PWE (FEM+GA) and omnidirectional RIR (FEM+GA) at the reference position 1, meeting room

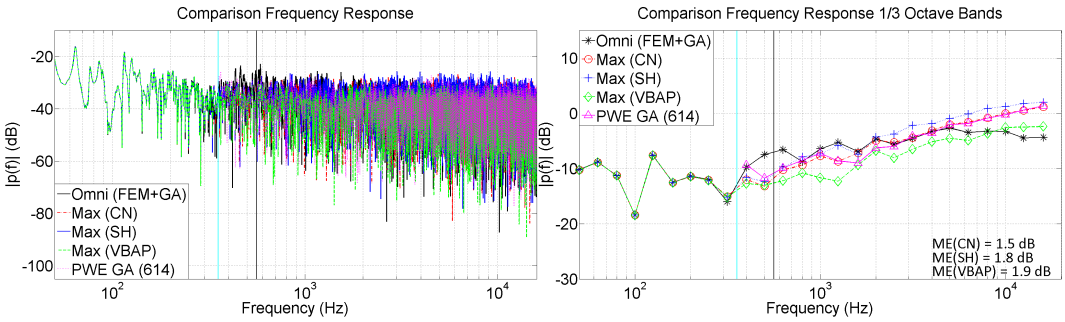


FIGURE 6.23: Comparison of frequency responses. Translated PWE (FEM+GA) and omnidirectional RIR (FEM+GA) at the translated position 2, meeting room

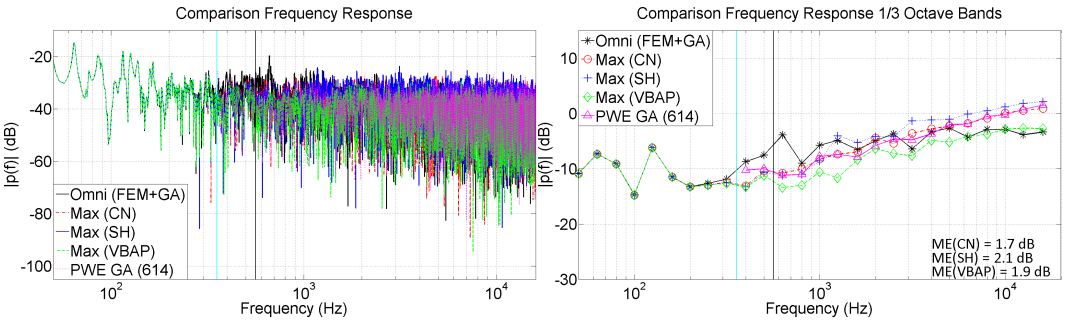


FIGURE 6.24: Comparison of frequency responses. Translated PWE (FEM+GA) and omnidirectional RIR (FEM+GA) at the translated position 3, meeting room

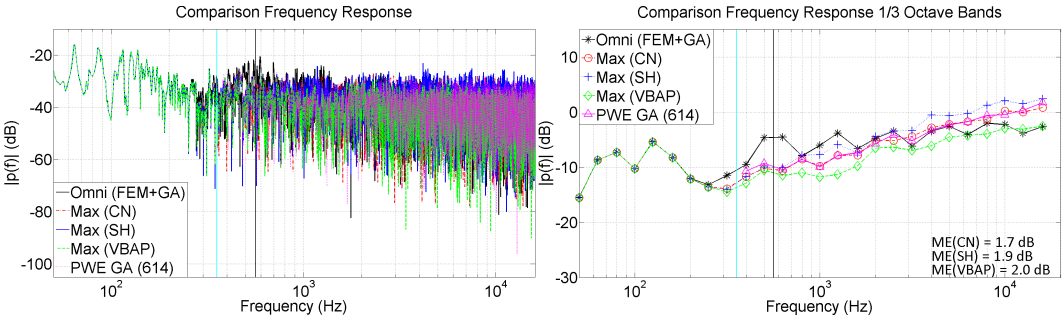


FIGURE 6.25: Comparison of frequency responses. Translated PWE (FEM+GA) and omnidirectional RIR (FEM+GA) at the translated position 4, meeting room

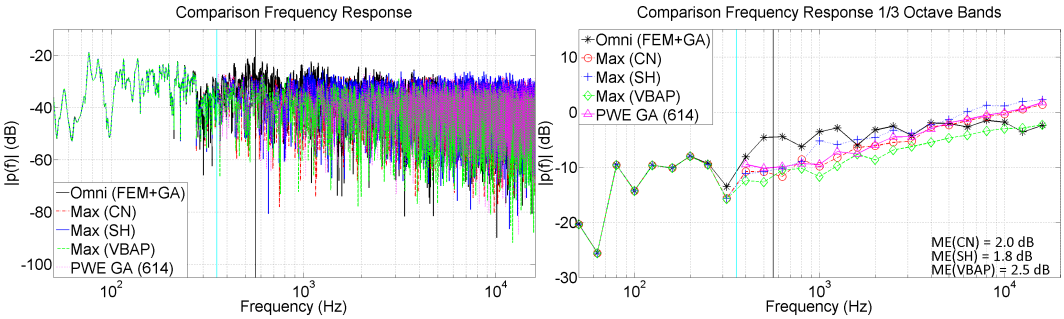


FIGURE 6.26: Comparison of frequency responses. Translated PWE (FEM+GA) and omnidirectional RIR (FEM+GA) at the translated position 5, meeting room

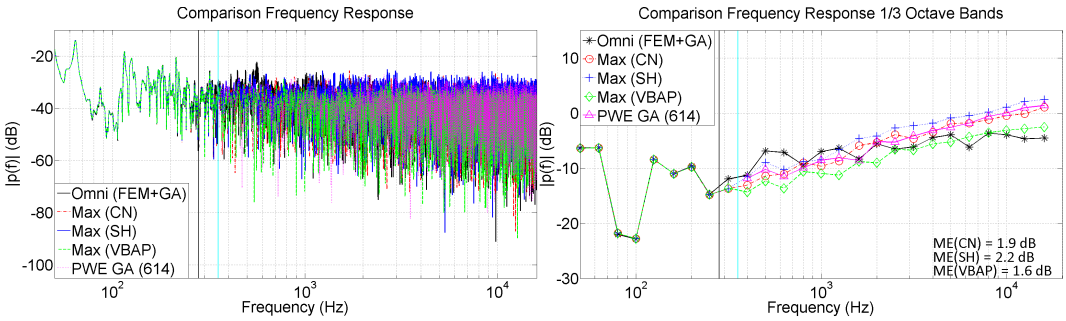


FIGURE 6.27: Comparison of frequency responses. Translated PWE (FEM+GA) and omnidirectional RIR (FEM+GA) at the translated position 6, meeting room

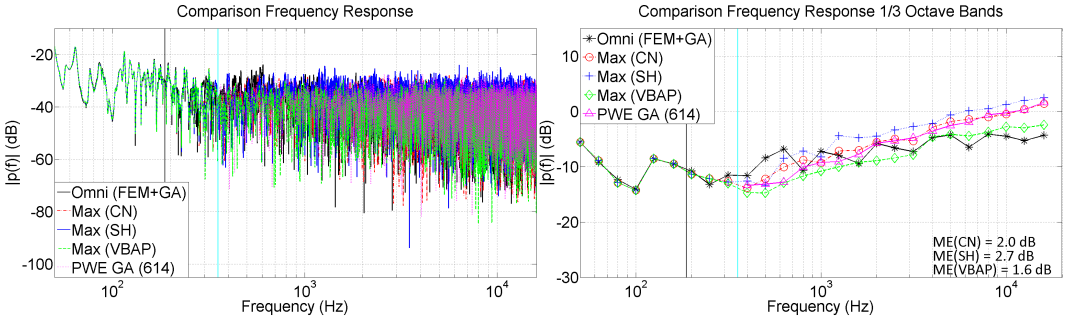


FIGURE 6.28: Comparison of frequency responses. Translated PWE (FEM+GA) and omnidirectional RIR (FEM+GA) at the translated position 7, meeting room

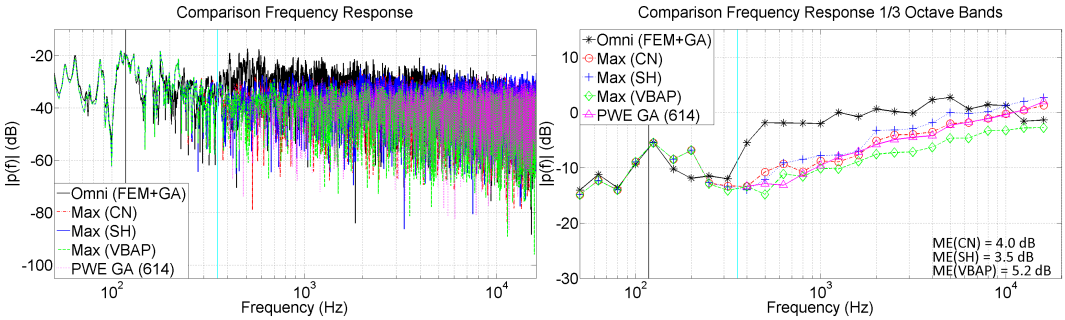


FIGURE 6.29: Comparison of frequency responses. Translated PWE (FEM+GA) and omnidirectional RIR (FEM+GA) at the translated position 8, meeting room

Receiver	Dist. (m)	Freq. (Hz)	CN (dB)	SH (dB)	VBAP (dB)
2	0.5	≈ 562	1.5	1.8	1.9
3	0.5	≈ 562	1.7	2.1	1.9
4	0.5	≈ 562	1.7	1.9	2.0
5	0.5	≈ 562	2.0	1.8	2.5
6	1	≈ 281	1.9	2.2	1.6
7	1.5	≈ 187	2.0	2.7	1.6
8	2.4	≈ 117	4.0	3.5	5.2

TABLE 6.2: Mean errors for the interactive auralization at different receiver locations, meeting room. The distance to the central point of the expansion and the maximum frequency in which is expected to achieve an accurate reconstruction is reported for each receiver

The figures show a similar tendency to the findings from subsection 5.1.1.2 in which the plane wave expansion from geometrical acoustics data showed a higher disagreement with the omnidirectional references compared to the rectangular room. At high frequencies, the analysis of the reconstructed sound fields points out that the translation is more accurately performed for the closest neighbour and spherical harmonic approaches if the initial PWE of 614 elements are selected as reference. The mismatches found at high frequencies are associated not to the translation operator but, to the disagreement between the initial plane wave expansion of 614 elements and the omnidirectional reference (see subsection 5.1.1.2). Nevertheless, the mean errors within a radius of 1.5 m never exceeded 2.0 dB, 2.7 dB and 2.5 dB for the closest neighbour, spherical harmonic interpolation and VBAP, respectively. At low frequencies, as expected, very good agreement between the reconstructed sound fields and the omnidirectional references was achieved. In this frequency range, the changes in the modal response of the enclosure are correctly predicted by the translation operator. Furthermore, the figures show a high degree of accuracy in the sound field reconstruction up to the frequencies predicted by equation (4.12) as long as these frequencies are below the crossover frequency. If not, the reconstruction is very accurate up to the frequency where the sound field is predicted using FE data, namely, the crossover frequency.

6.4.1.3 Ightham Mote

Figure 5.37 shows the location of 8 receivers that have been used for the current analysis. The predicted frequency response in narrow band and in 1/3 octave band

resolution for each receiver are illustrated in Figures 6.30 to 6.37. The crossover frequency between the finite element and geometrical acoustic data corresponds to a 282 Hz for this enclosure. Table 6.3 shows the mean errors according to the approach used to resample the GA information.

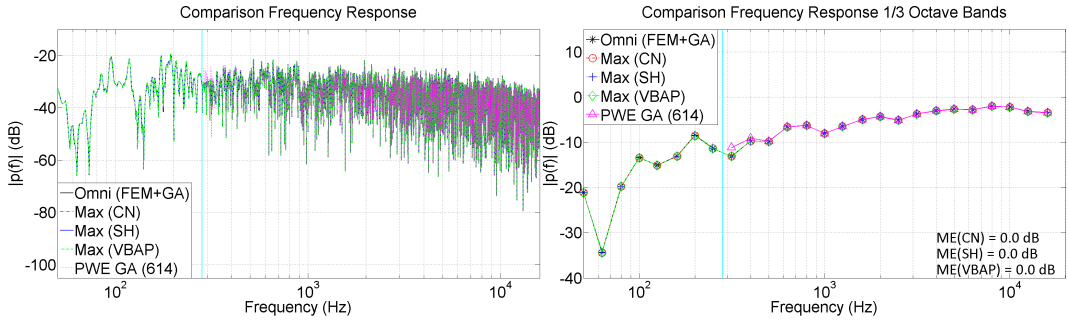


FIGURE 6.30: Comparison of frequency responses. PWE (FEM+GA) and omnidirectional RIR (FEM+GA) at the reference position 1, Ightham Mote

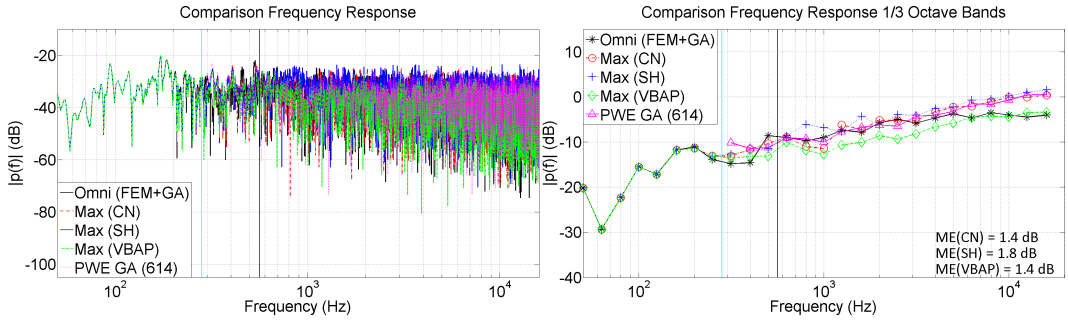


FIGURE 6.31: Comparison of frequency responses. Translated PWE (FEM+GA) and omnidirectional RIR (FEM+GA) at the translated position 2, Ightham Mote

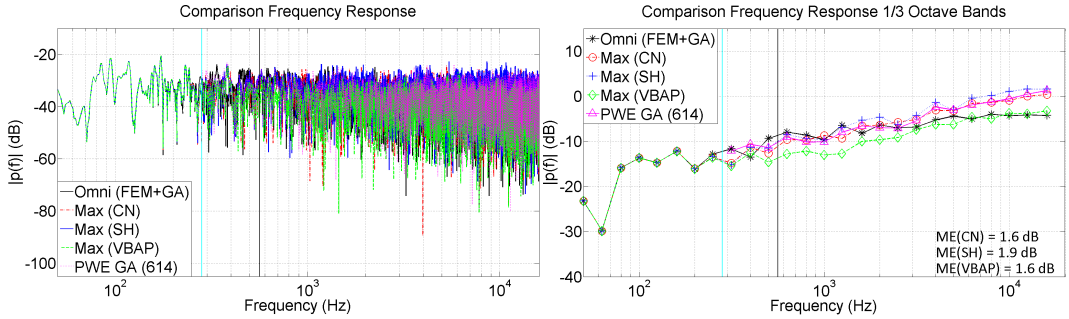


FIGURE 6.32: Comparison of frequency responses. Translated PWE (FEM+GA) and omnidirectional RIR (FEM+GA) at the translated position 3, Ightham Mote

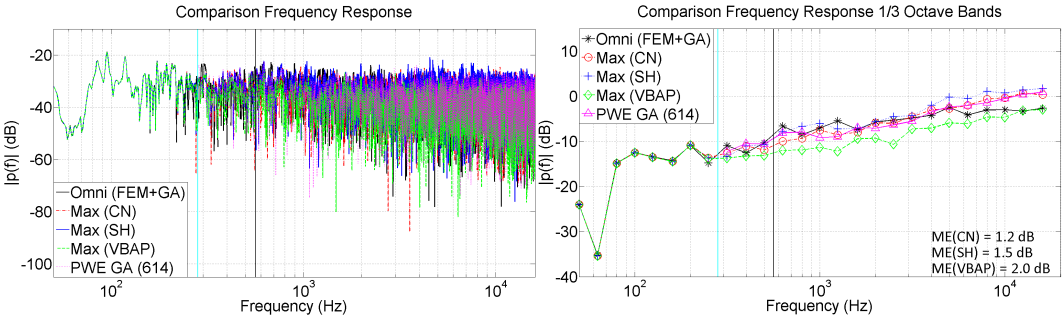


FIGURE 6.33: Comparison of frequency responses. Translated PWE (FEM+GA) and omnidirectional RIR (FEM+GA) at the translated position 4, Ightham Mote

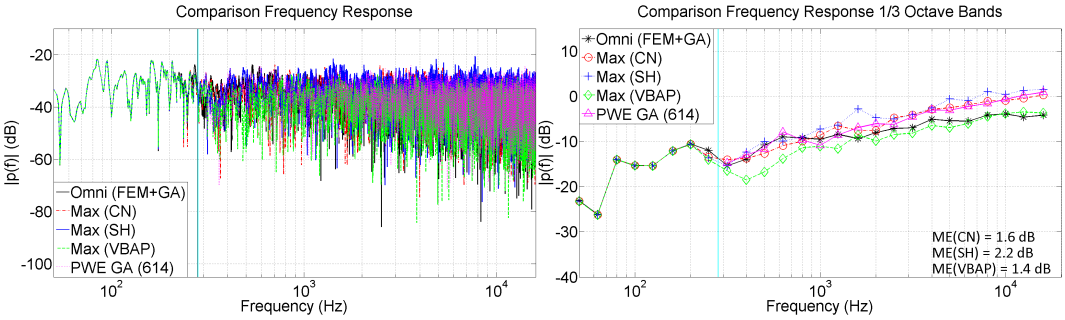


FIGURE 6.34: Comparison of frequency responses. Translated PWE (FEM+GA) and omnidirectional RIR (FEM+GA) at the translated position 5, Ightham Mote

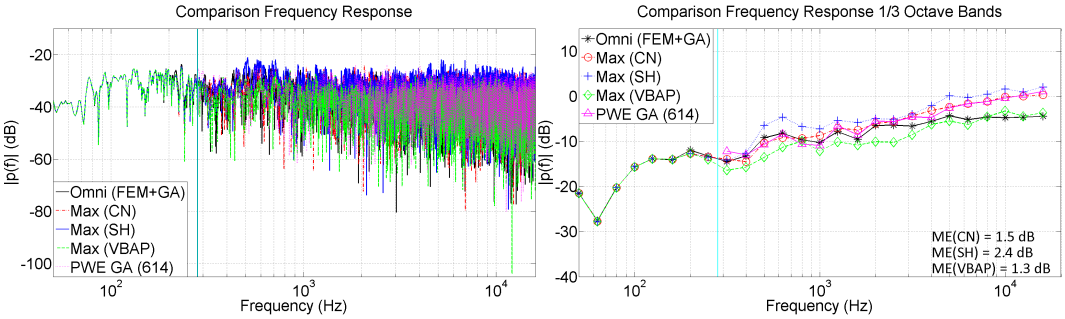


FIGURE 6.35: Comparison of frequency responses. Translated PWE (FEM+GA) and omnidirectional RIR (FEM+GA) at the translated position 6, Ightham Mote

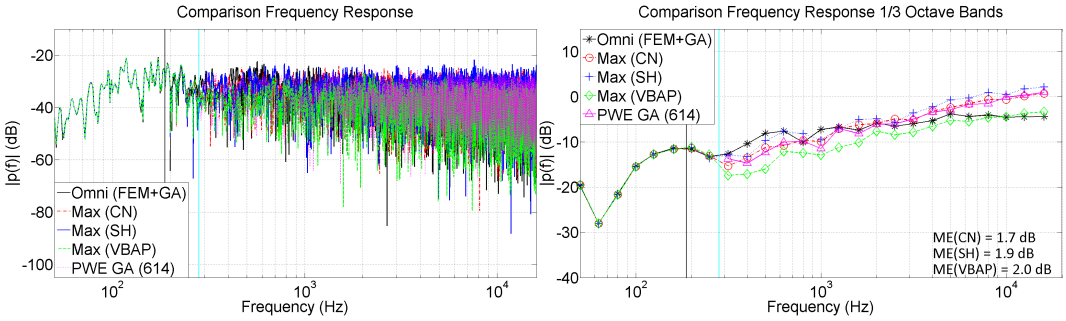


FIGURE 6.36: Comparison of frequency responses. Translated PWE (FEM+GA) and omnidirectional RIR (FEM+GA) at the translated position 7, Ightham Mote

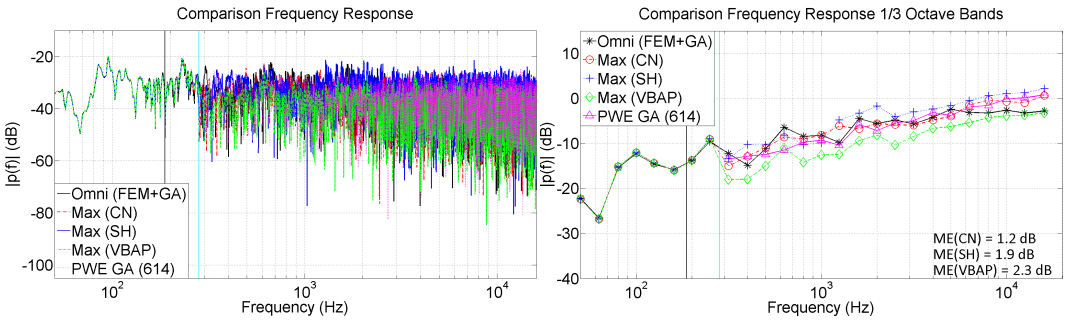


FIGURE 6.37: Comparison of frequency responses. Translated PWE (FEM+GA) and omnidirectional RIR (FEM+GA) at the translated position 8, Ightham Mote

Receiver	Dist. (m)	Freq. (Hz)	CN (dB)	SH (dB)	VBAP (dB)
2	0.5	≈ 562	1.4	1.8	1.4
3	0.5	≈ 562	1.6	1.9	1.6
4	0.5	≈ 562	1.2	1.5	2.0
5	1	≈ 281	1.6	2.2	1.4
6	1	≈ 281	1.5	2.4	1.3
7	1.5	≈ 187	1.7	1.9	2.0
8	1.5	≈ 187	1.2	1.9	2.3

TABLE 6.3: Mean errors for the interactive auralization at different receiver locations, Ightham Mote. The distance to the central point of the expansion and the maximum frequency in which is expected to achieve an accurate reconstruction is reported for each receiver

The outcomes are consistent with the results from the two previous enclosures. The translation of the plane wave expansion leads to an accurate reconstruction of the reference acoustic field at low frequencies, specifically below the crossover frequency.

An interesting finding was obtained from the receiver positions 7 and 8 where the synthesis of the sound field is correctly performed at frequencies higher than the threshold given by equation (4.12). At high frequencies (octave bands of 8 kHz and 16 kHz), the energy predicted by the translation of the plane wave expansion is higher than the omnidirectional references. This result is consistent in all the enclosures and it is an intrinsic feature of the method used to calculate the plane wave expansion from GA simulations. Regarding the three methods used to resample the PWE from geometrical acoustic data, Table 6.3 indicates that closest neighbour is the most accurate approach. The mean errors using this technique are smaller than 1.7 dB for all the receivers considered.

6.4.1.4 Office room

Figure 6.38 shows three receiver positions that were used to evaluate the performance of the auralization system in the office room. This enclosure is the smallest space of the 4 reference cases, thus providing a good scenario to assess the constraints given by the chosen sound propagation model. The central point of the plane wave expansion is at the receiver 01. The other two receivers were selected to compare the synthesized acoustic pressure predicted by the translation of the plane wave expansion to omnidirectional references. The distance of receiver 03 to receiver 01 is approximately 3 times the distance between receiver 02 and 01 allowing for a more global analysis of the space. Furthermore, due to its smaller dimensions, the crossover frequency between the FE and GA data corresponded to 562 Hz, which is the highest among the crossover frequencies used in this work. Figures 6.39, 6.40 and 6.41 illustrate the predicted frequency response in narrow band and in 1/3 octave band resolution for each receiver. Table 6.4 summarizes the mean error according to the approaches selected to resample the GA data.

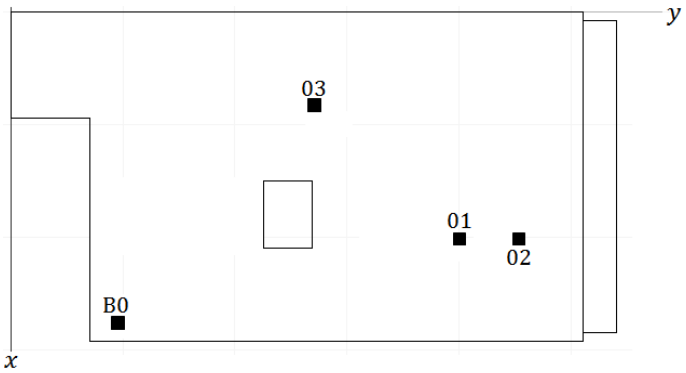


FIGURE 6.38: Sketch of the office room, validation of the auralization system

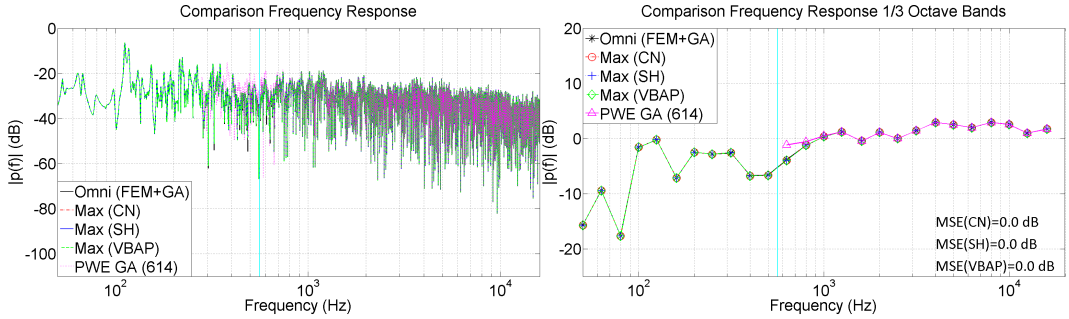


FIGURE 6.39: Comparison of frequency responses. PWE (FEM+GA) and omnidirectional RIR (FEM+GA) at the reference position 1, office room

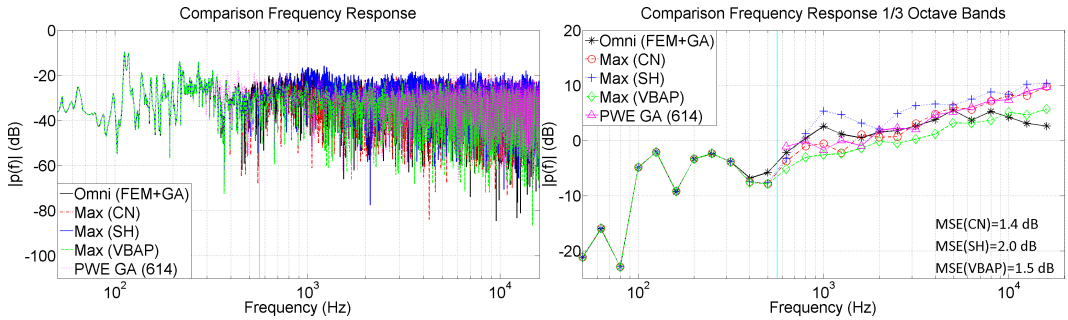


FIGURE 6.40: Comparison of frequency responses. Translated PWE (FEM+GA) and omnidirectional RIR (FEM+GA) at the translated position 2, office room

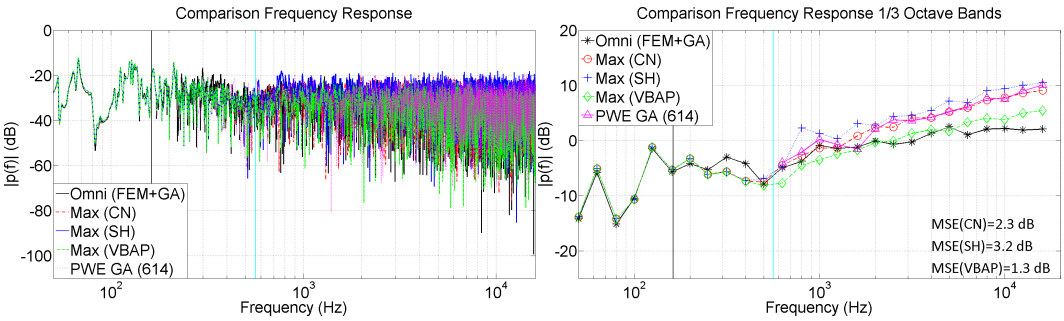


FIGURE 6.41: Comparison of frequency responses. Translated PWE (FEM+GA) and omnidirectional RIR (FEM+GA) at the translated position 3, office room

Receiver	Dist. (m)	Freq. (Hz)	CN (dB)	SH (dB)	VBAP (dB)
2	0.5	≈ 562	1.4	2.0	1.5
3	1.7	≈ 162	2.3	3.2	1.3

TABLE 6.4: Mean errors for the interactive auralization at different receiver locations, office room

Figure 6.40 indicates that the frequency predicted by equation (4.12) and the crossover frequency between FE and GA match at receiver 02. An excellent agreement between the synthesized and reference frequency responses was found up to 400 Hz for this receiver. Beyond that, a good agreement in terms of the frequency response in 1/3 octave band resolution was achieved. Regarding receiver 03, an accurate sound field reconstruction was achieved up to the frequency predicted by equation (4.12). From that frequency and up to the crossover frequency, a mismatch between the predicted and reference frequency responses was found. This disagreement is related to the initial error given by the discretization of the plane wave expansion, which leads the translation operator to reconstruct a different acoustic field from the reference one. Above the crossover frequency, an overestimation of the energy was found for this receiver. Nevertheless, the mean errors values are 2.3 dB, 3.2 dB and 1.3 dB for the closest neighbour, spherical harmonic interpolation and VBAP, respectively. An interesting result was found for this room that is consistent with the outcomes for the meeting room. The overestimation of the energy with respect to the omnidirectional references given the closest neighbour and spherical harmonic interpolation tends to be higher in these smaller rooms. However, in the case of VBAP, the lack of energy provided by this method (compared to the initial plane wave expansion of 614 element) reduced the gap between the predicted and reference frequency responses. This particular condition makes VBAP the most accurate approach for this specific

room.

The monaural analysis of the 4 enclosures reveals that, as expected, the plane wave expansion provides a description that is locally accurate of the acoustic fields. The extent of the region where an accurate reconstruction is performed is mainly determined by the number of directional impulse responses but also depends on the specific enclosure. In large rooms in which the plane wave model is suitable a wider low-error region was found. The implementation of the translation operator in the auralization system leads to an accurate representation of the modal response of the rooms at low frequencies. At high frequencies, a good agreement in terms of energy was found up to a distance of 1.5 m from the central point of the expansion. From the three different approaches used to resample the geometrical acoustic data (PWE of 614 elements), the closest neighbour provides the lowest mean error in 3 of the 4 rooms. In this case, the mean error never exceeds 2.3 dB within a radius of 1.5 m. Finally, a good agreement between the frequency determined by equation (4.12) and the frequency up to which the sound field reconstruction is accurately performed was found for all rooms. This agreement remains as long as the frequency determined by equation (4.12) is lower than the crossover frequency.

6.4.2 Spatial analysis

The previous experiment evaluated the auralization system in terms of the accuracy to reconstruct the acoustic pressure at different reference locations. Although, this analysis provides a useful insight into the performance of the auralization system, it does not give any information about the spatial characteristics of the synthesized sound fields. An evaluation of spatial information is carried out in this section based on B-format (first order) signals, namely, the signals from a monopole ($Y_0^0 = W$) and three orthogonal dipoles oriented over the coordinate axes ($Y_1^{-1} = Y$, $Y_1^0 = Z$, $Y_1^1 = X$). The assessment of the sound field by means of B-format signals allows an investigation of how the energy is spatially distributed, which is important for binaural and multichannel sound field reproduction. It has been shown by Fazi and Menzies [150] that is possible to extract the acoustic pressure and particle velocity from B-format signals. This information can be used to accurately estimate the localization cues at low frequencies [151]. The procedure to generate the B-format signals is described below.

The low frequency content of the B-format reference signals was estimated by the implementation of an inverse method according the formulation given in section 4.1.2. For this, virtual microphone arrays were used to sample the FE data at the positions where the B-format signals were intended to be synthesized. At high frequencies, B-format impulse responses were directly exported from Catt-Acoustics. These two sound field representations were combined based on the approach established in section 3.4 leading to unified reference B-format signals. The B-format signals from the auralization system were obtained by using the rotation module (see Figure 6.1), which is based on a spherical harmonic transformation. The signals corresponding to the zero and first order were recorded after the encoding stage. In the following, a comparison between the reference (numerical) and the synthesized (by the auralization system) B-format signals is presented for the four enclosures previously considered.

6.4.2.1 Rectangular room

Receivers 4 and 7 illustrated in Figure 5.17 were selected for the analysis in the rectangular room. The selection of these receiver positions was motivated by the distance from the central point of the expansion (rcv(4)=1 m and rcv(7)=3 m) and the mean errors from the monaural evaluation, which provide two different levels of agreement. Figure 6.42 illustrates a comparison between the omnidirectional frequency response predicted by FEM and the W signal estimated by the inverse method. The W signal was compensated by the factor ($\sqrt{2}$) in order to compare it with respect to the omnidirectional reference. The result indicates that the inverse method is able to generate the correct spherical harmonic coefficients and therefore can be used for the subsequent analysis. Figures 6.43 to 6.48 show the frequency response in narrow band and in 1/3 octave band resolution of the reference and synthesized B-format signals. Table 6.5 reports the mean error according to the approach selected to resample the GA data.

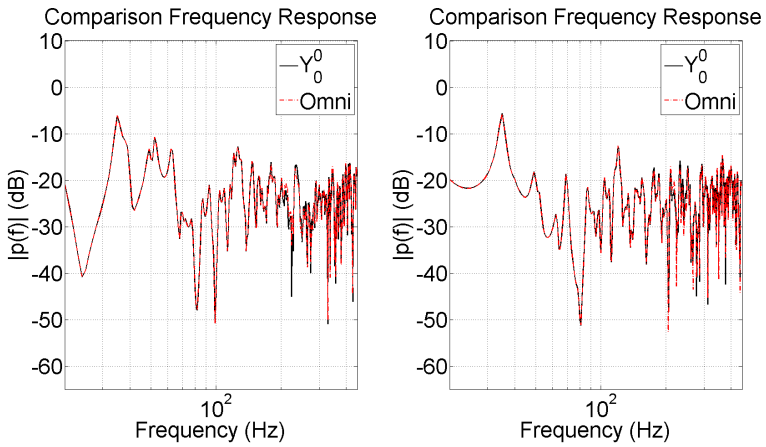


FIGURE 6.42: Comparison of frequency responses. W and omidirectional (FEM), positions 4 (left) and 7 (right), rectangular room

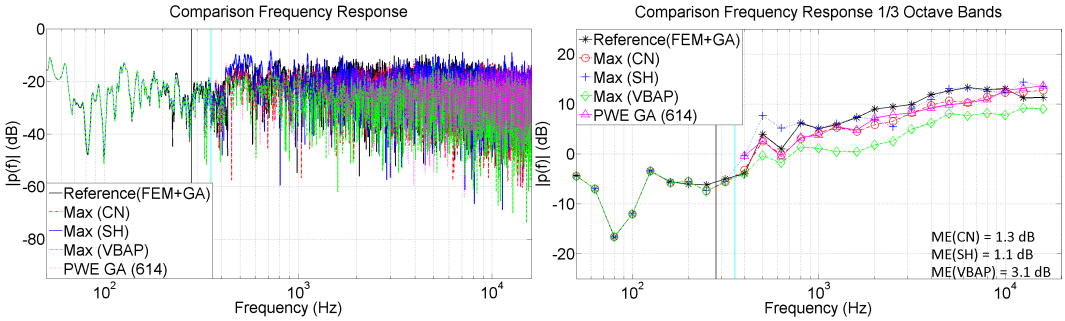


FIGURE 6.43: Comparison of frequency responses. Translated PWE (W) and reference (W) at the translated position 4, rectangular room

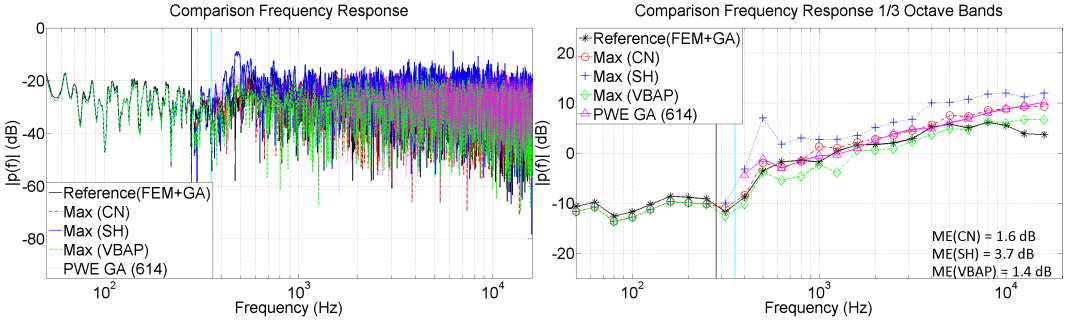


FIGURE 6.44: Comparison of frequency responses. Translated PWE (X) and reference (X) at the translated position 4, rectangular room

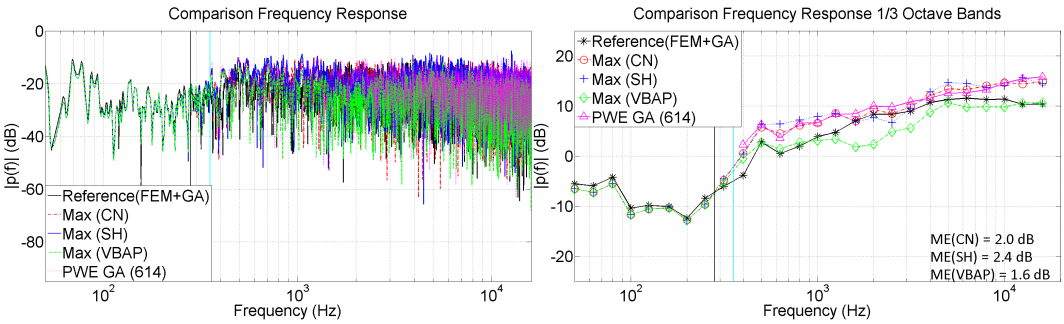


FIGURE 6.45: Comparison of frequency responses. Translated PWE (Y) and reference (Y) at the translated position 4, rectangular room

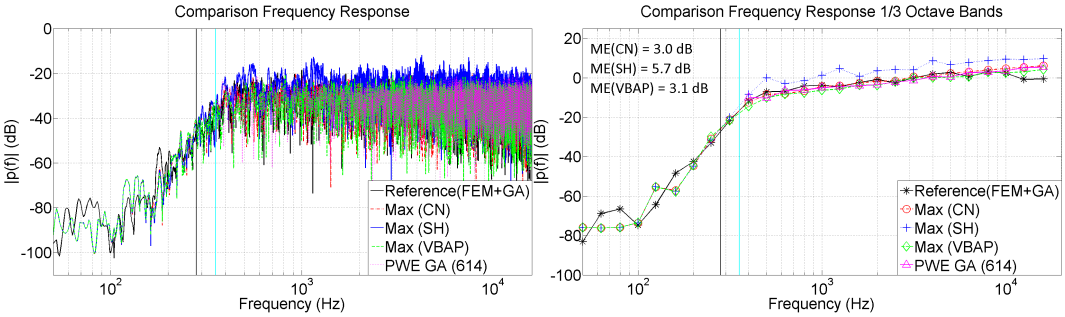


FIGURE 6.46: Comparison of frequency responses. Translated PWE (Z) and reference (Z) at the translated position 4, rectangular room

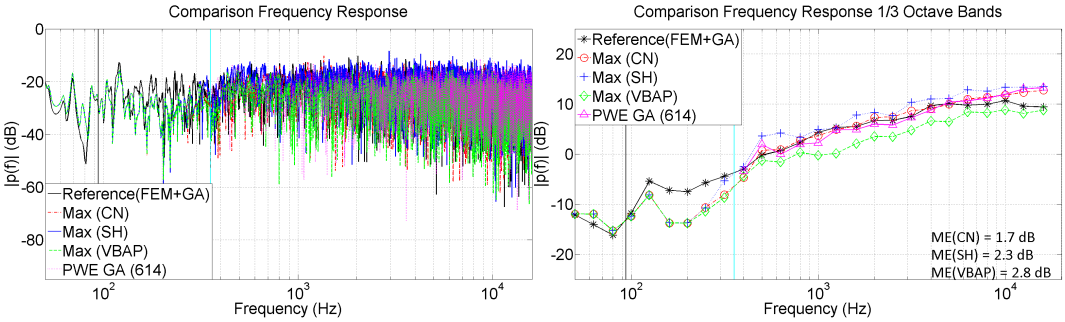


FIGURE 6.47: Comparison of frequency responses. Translated PWE (W) and reference (W) at the translated position 7, rectangular room

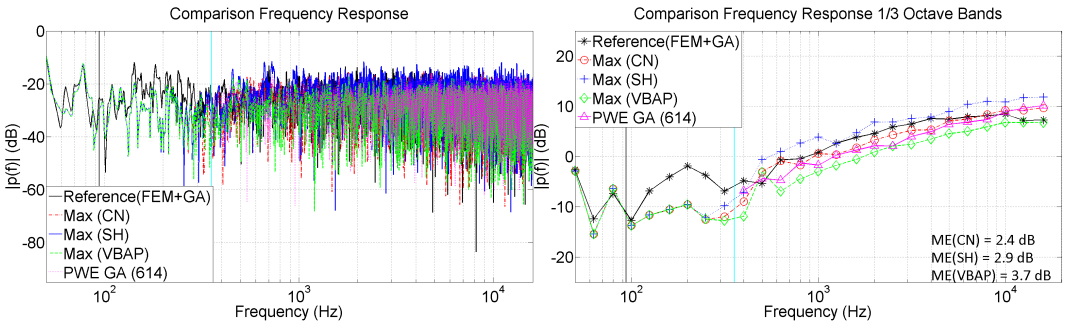


FIGURE 6.48: Comparison of frequency responses. Translated PWE (X) and reference (X) at the translated position 7, rectangular room

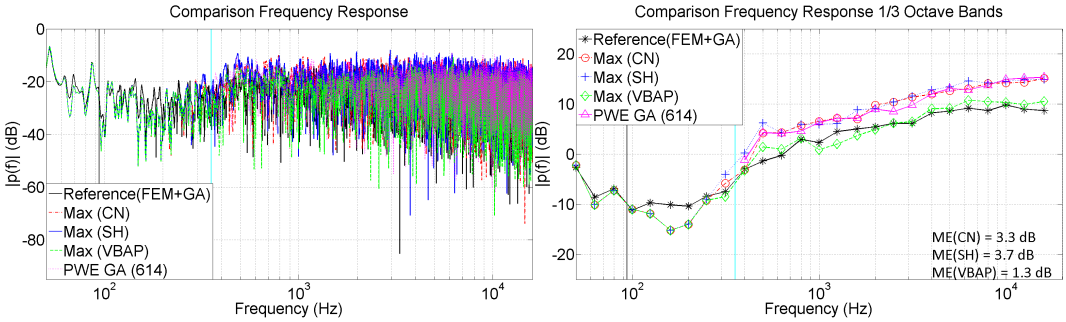


FIGURE 6.49: Comparison of frequency responses. Translated PWE (Y) and reference (Y) at the translated position 7, rectangular room

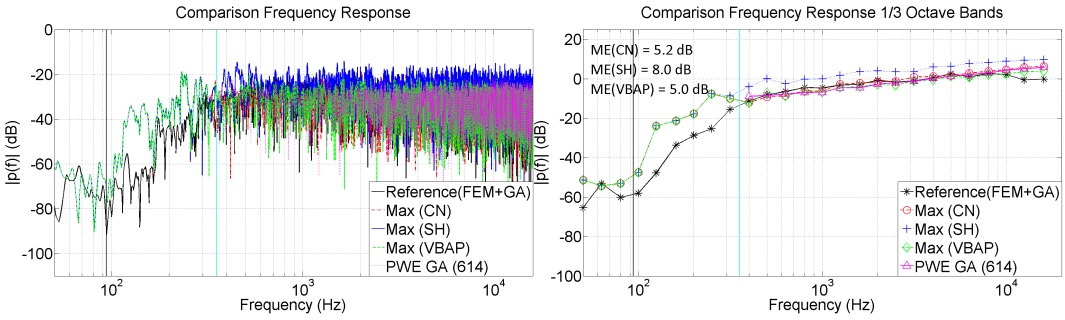


FIGURE 6.50: Comparison of frequency responses. Translated PWE (Z) and reference (Z) at the translated position 7, rectangular room

Receiver	**Approach**	**W** (dB)	**Y** (dB)	**Z** (dB)	**X** (dB)	**Average** (dB)
4 (1m)	CN	1.3	2.0	3.0	1.6	2.0
	SH	1.1	2.4	5.7	3.7	3.2
	VBAP	3.1	1.6	3.1	1.4	2.3
7 (3m)	CN	1.7	3.3	5.2	2.4	3.2
	SH	2.3	3.7	8.0	2.9	4.2
	VBAP	2.8	1.3	5.0	3.7	3.3

TABLE 6.5: Mean errors in the B-format signals for the interactive auralization at different receiver locations, rectangular room

A comparison of the mean errors indicates that receiver 4 leads to smaller values than receiver 7. The reason for this result can be mainly associated with the fact that the distance of the receivers to the central point the expansion is smaller for receiver 4. Regarding the B-format signals, the outcomes indicate that the reconstruction is more accurately performed for the W signals. In this case, a very good agreement between the frequency response synthesized by the auralization system and the reference signal was found up to the frequency established by equation (4.12). For the other orders, the match at low frequencies is not as good as the zero order, but with a good agreement in terms of the frequency response in 1/3 octave band resolution. The largest differences were found for the Z signals. Nevertheless, it is important to point out that the magnitude of the Z signals are very low so the mean errors may not be as perceptually relevant as for the other signals. Above the crossover frequency (GA data), the results are consistent with the monaural results (section 6.4.1.1) showing that the mismatch is dominated by the initial discrepancy between the plane wave expansion of 614 elements and the omnidirectional references. This is confirmed by the comparison of the B-format signals computed directly from the GA plane wave expansion of 614 elements and the B-format signals recorded from the auralization system, which are very similar for the case of the closest neighbour. The overestimation of the energy in the octave bands of 8 kHz and 16 kHz given by the auralization system was also present for the B-format signals suggesting that this is an intrinsic feature of the approach selected to generate the plane wave expansion from the GA data.

Figures 6.47 to 6.48, which correspond to the results for receiver 7, show that the highest disagreement is given in the frequency range determined by the frequency

predicted by equation (4.12) and the crossover frequency (355 Hz). This discrepancy is associated with the initial error given by the discretization of the plane wave expansion, which restricts the area where the translation operator leads to the reference sound field. Finally, Table 6.5 shows that the closest neighbour is on average the more accurate approach from the three methods implemented to resample the GA data. For this room, the worst approach was the use of a spherical harmonic interpolation.

6.4.2.2 Meeting room

The analysis of the B-format signals was carried out at receivers 2 and 7 for the meeting room (see Figure 5.28). These receivers were chosen because they corresponded to the locations where a good and bad agreement was found in terms of the mean errors for the closest neighbour approach (monaural evaluation). The frequency response in narrow band and in 1/3 octave band resolution of the reference and synthesized B-format signals are illustrated in Figures 6.51 to 6.56. A summary of the mean errors according to the approach selected to resample the GA data is presented in Table 6.6.

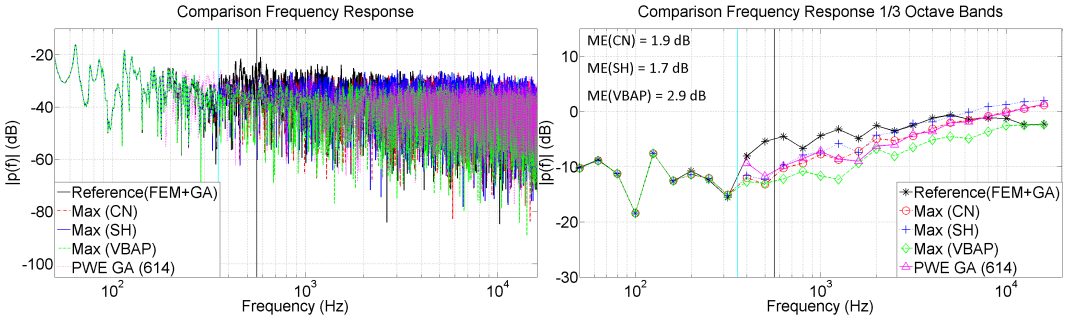


FIGURE 6.51: Comparison of frequency responses. Translated PWE (W) and reference (W) at the translated position 2, meeting room

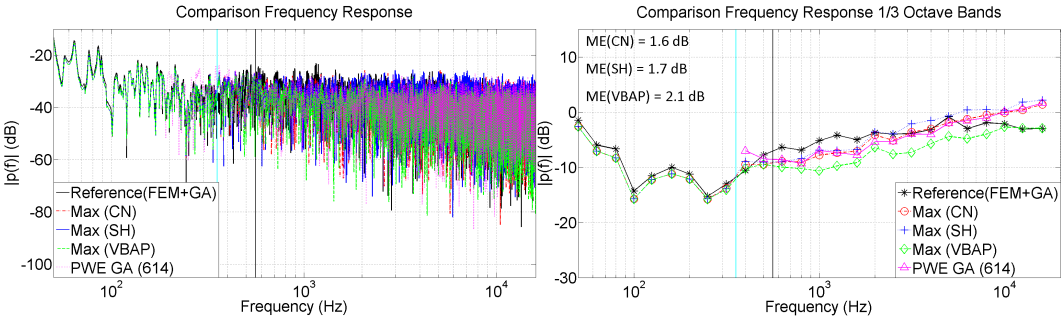


FIGURE 6.52: Comparison of frequency responses. Translated PWE (X) and reference (X) at the translated position 2, meeting room

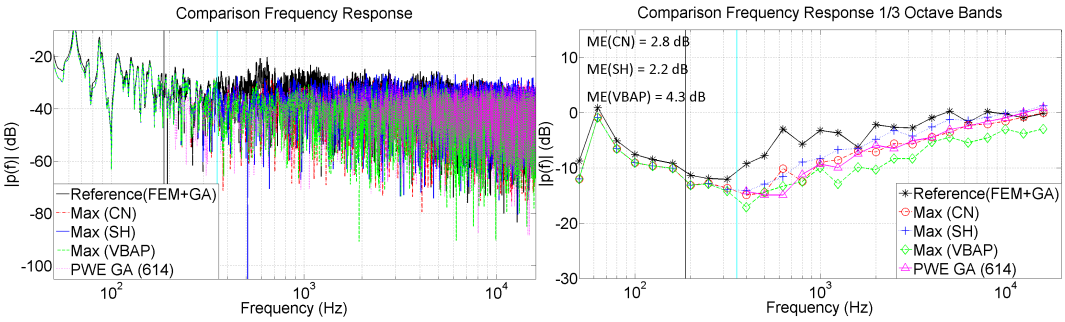


FIGURE 6.53: Comparison of frequency responses. Translated PWE (Y) and reference (Y) at the translated position 2, meeting room

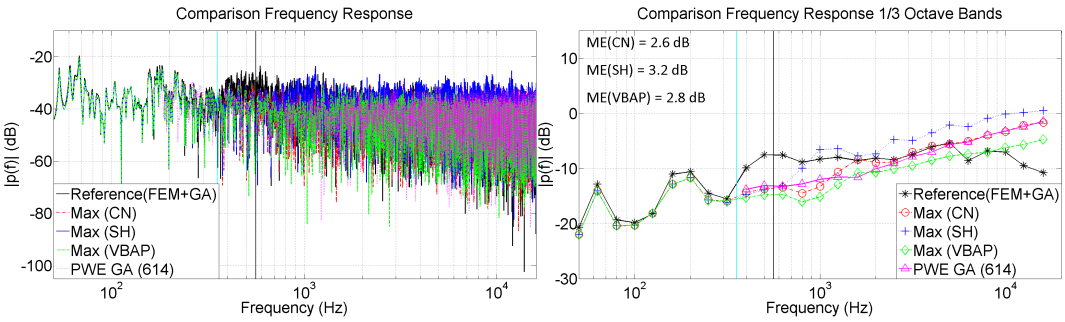


FIGURE 6.54: Comparison of frequency responses. Translated PWE (Z) and reference (Z) at the translated position 2, meeting room

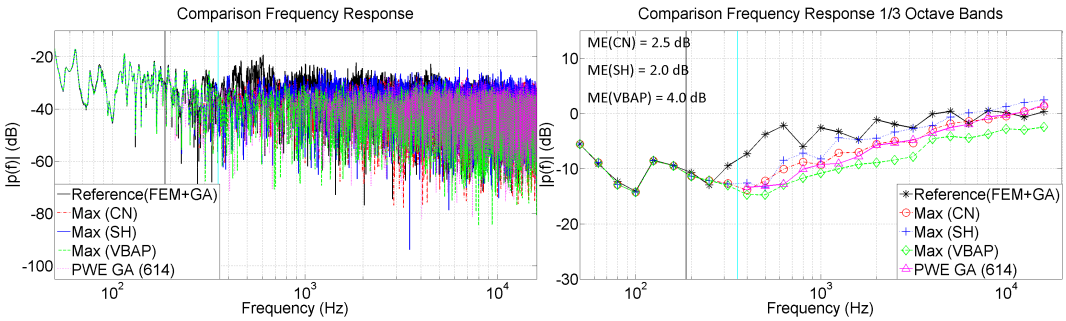


FIGURE 6.55: Comparison of frequency responses. Translated PWE (W) and reference (W) at the translated positionn 7, meeting room

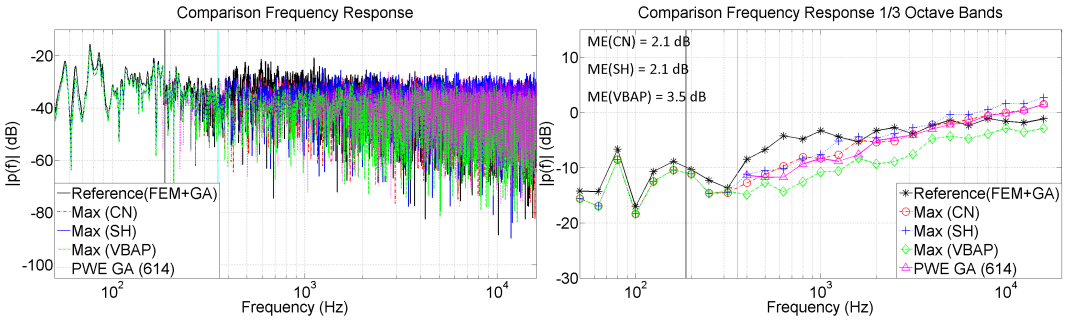


FIGURE 6.56: Comparison of frequency responses. Translated PWE (X) and reference (X) at the translated position 7, meeting room

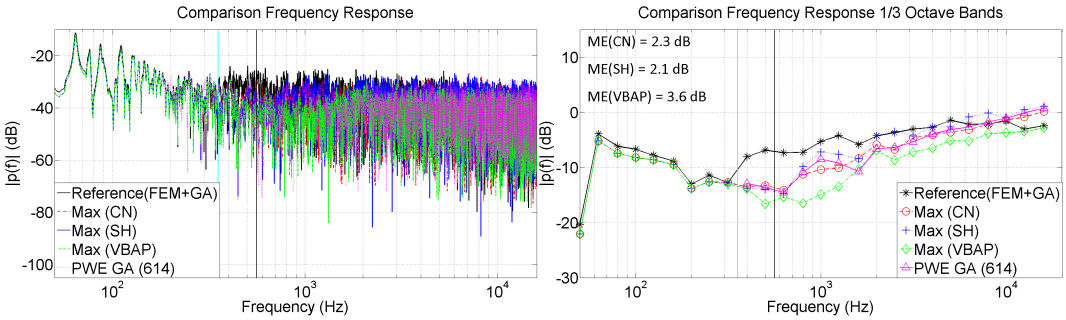


FIGURE 6.57: Comparison of frequency responses. Translated PWE (Y) and reference (Y) at the translated position 7, meeting room

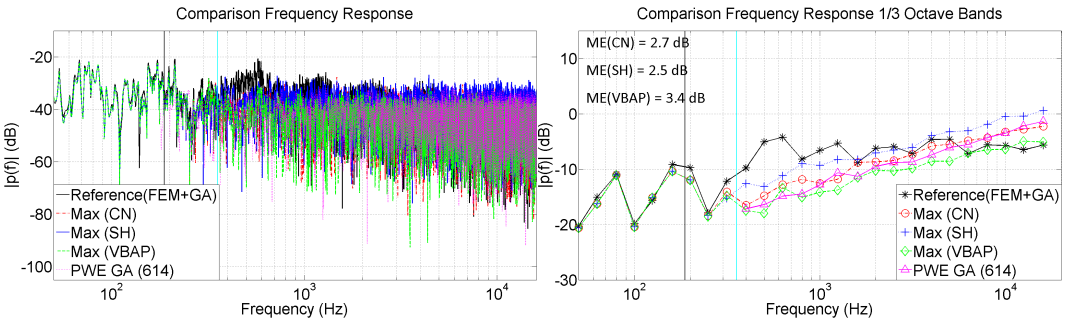


FIGURE 6.58: Comparison of frequency responses. Translated PWE (Z) and reference (Z) at the translated position 7, meeting room

Receiver	**Approach**	**W** (dB)	**Y** (dB)	**Z** (dB)	**X** (dB)	**Average** (dB)
2 (0.5m)	CN	1.9	2.8	2.6	1.6	2.2
	SH	1.7	2.2	3.2	1.7	2.2
	VBAP	2.9	4.3	2.8	2.1	3.0
7 (1.5m)	CN	2.5	2.3	2.7	2.1	2.4
	SH	2.0	2.1	2.5	2.1	2.2
	VBAP	4.0	3.6	3.4	3.5	3.6

TABLE 6.6: Mean errors in the B-format signals for the interactive auralization at different receiver locations, meeting room

The results indicate that the reconstruction of the B-format signals is performed with good accuracy for all the components (W,X, Y and Z) below the crossover frequency (355 Hz). A singnificant disagreement between the synthesized and reference B-format signals was found for both receivers in the octave bands of 500 Hz and 1 kHz. This outcome is consistent with the results obtained in section 5.1.1.2 in which the translation of the plane wave expansion of 614 elements (GA data) was compared to omnidirectional references predicted in Catt-Acoustics. The mismatches with respect to the references in these octave bands are associated with the constraint imposed by the size of the room on the plane wave propagation assumption. At higher frequencies, the level of agreement between the B-format signals mainly depends how well the translation of the plane wave expansion of 614 elements matches the omnidirectional receivers. Regarding the method to resample the GA data, Table 6.6 shows that closest neighbour is the most accurate approach for receiver 2 whilst the spherical harmonic interpolation is in the case of receiver 7.

6.4.2.3 Ightham Mote

Figures 6.59 to 6.64 illustrate the frequency response in narrow band and in 1/3 octave band resolution of the reference and synthesized B-format signals at two different receiver positions for the Ightham Mote. The locations correspond to 04 and 07 based on the configuration established in Figure 6.7. A report of the mean errors for each channel of the B-format signals and receivers is presented in Table 6.7.

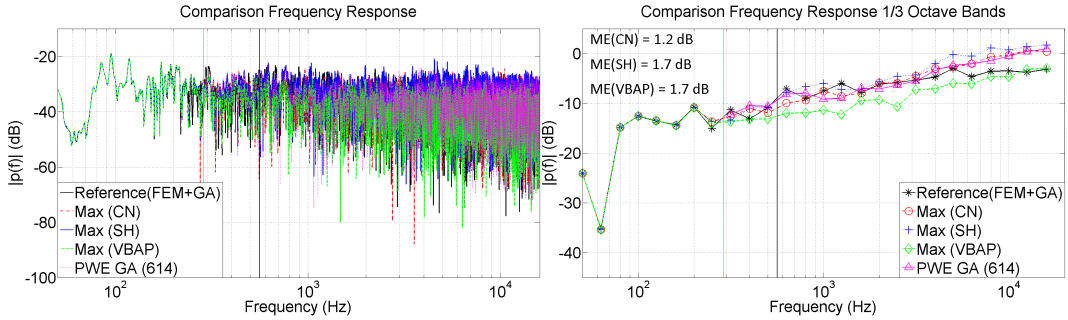


FIGURE 6.59: Comparison of frequency responses. Translated PWE (W) and reference (W) at the translated position 4, Ightham Mote

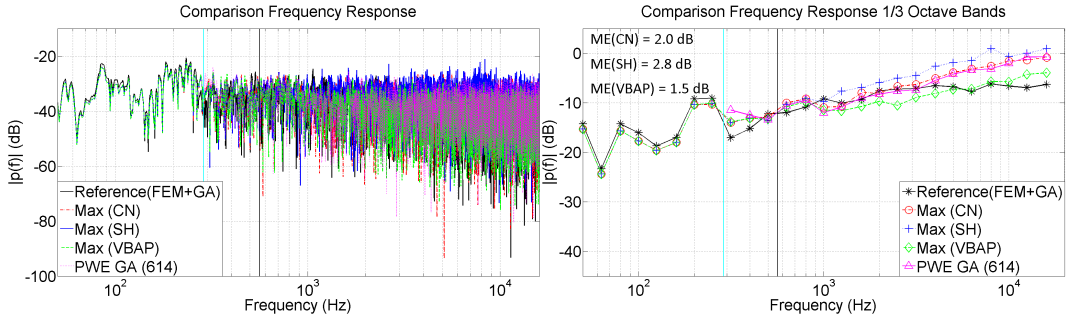


FIGURE 6.60: Comparison of frequency responses. Translated PWE (X) and reference (X) at the translated position 4, Ightham Mote

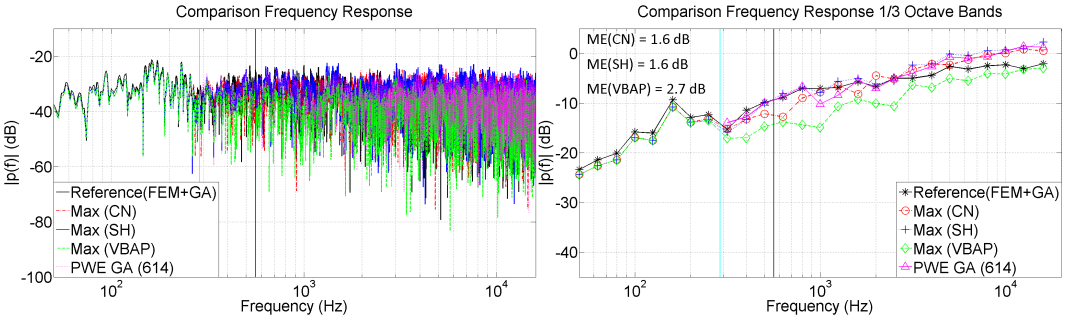


FIGURE 6.61: Comparison of frequency responses. Translated PWE (Y) and reference (Y) at the translated position 4, Ightham Mote

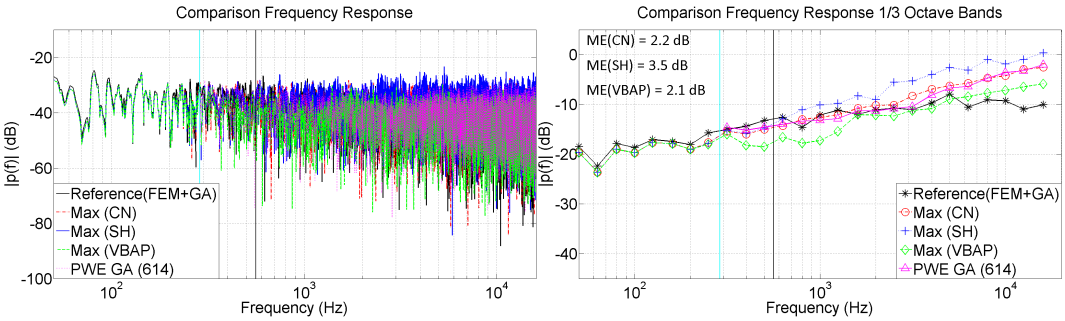


FIGURE 6.62: Comparison of frequency responses. Translated PWE (Z) and reference (Z) at the translated position 4, Ightham Mote

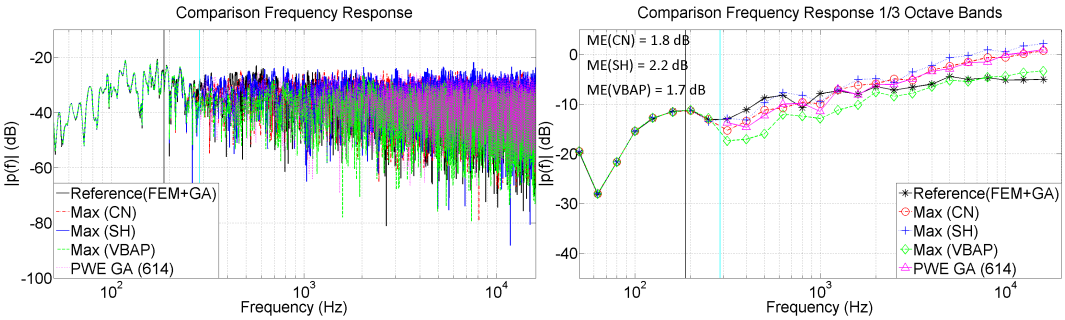


FIGURE 6.63: Comparison of frequency responses. Translated PWE (W) and reference (W) at the translated position 7, Ightham Mote

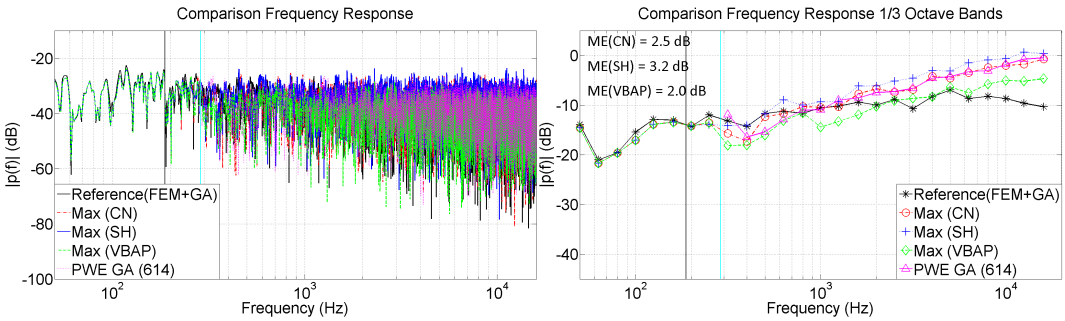


FIGURE 6.64: Comparison of frequency responses. Translated PWE (X) and reference (X) at the translated position 7, Ightham Mote

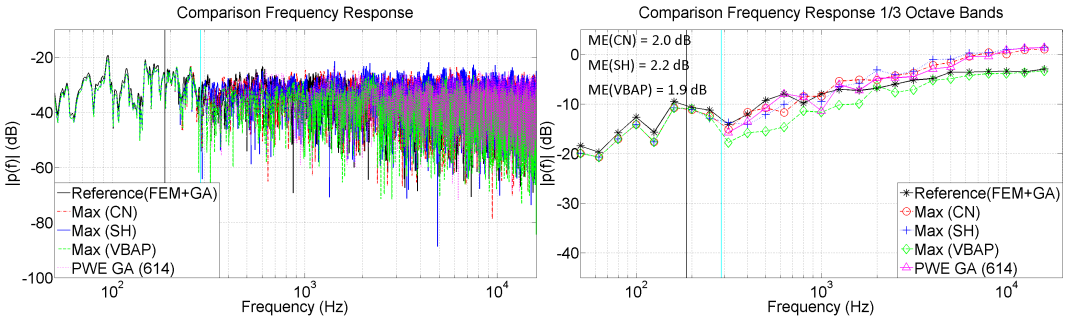


FIGURE 6.65: Comparison of frequency responses. Translated PWE (Y) and reference (Y) at the translated position 7, Ightham Mote

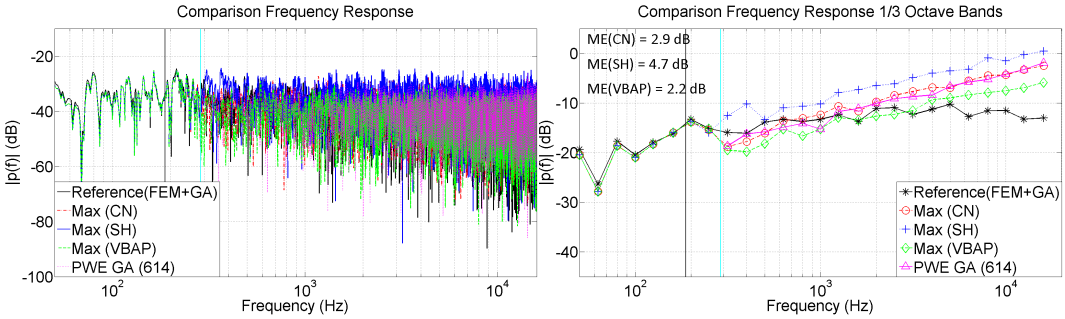


FIGURE 6.66: Comparison of frequency responses. Translated PWE (Z) and reference (Z) at the translated position 7, Ightham Mote

Receiver	**Approach**	**W** (dB)	**Y** (dB)	**Z** (dB)	**X** (dB)	**Average** (dB)
4 (0.5m)	CN	1.2	1.6	2.2	2.0	1.8
	SH	1.7	1.6	3.5	2.8	2.4
	VBAP	1.7	2.7	2.1	1.5	2.0
7 (1.5m)	CN	1.8	2.0	2.9	2.5	2.3
	SH	2.2	2.2	4.7	3.2	3.1
	VBAP	1.7	1.9	2.2	2.0	2.0

TABLE 6.7: Mean errors in the B-format signal for the interactive auralization at different receiver locations, Ightham Mote

The figures indicate that the sound field reconstruction of the B-format signals in the octave bands of 500 Hz and 1 kHz is more accurately performed than for the meeting room. This result supports the concept that the disagreements obtained for the meeting room are associated with the plane wave model. At frequencies below the crossover frequency (288 Hz), the synthesis of the all signals (W ,X, Y and Z) show good agreement with the reference. At high frequencies, the results are consistent with the findings reported in section 5.1.1.2 and 6.4.1.3 where the accuracy in this frequency range depends on the level of agreement between the plane wave expansion (GA data) of 614 elements and the omnidirectional references. Finally, Table 6.7 reveals that the closest neighbour is the most accurate approach in terms of the averaged mean error for receiver 4 whilst VBAP is for receiver 7.

6.4.2.4 Office room

The predicted frequency response in narrow band and in 1/3 octave band resolution of the B-format channels are presented below for the receiver positions 02 and 03 of the office rrom (see Figure 6.38). Table 6.8 summarizes the mean errors according to the method used to resample the GA data.

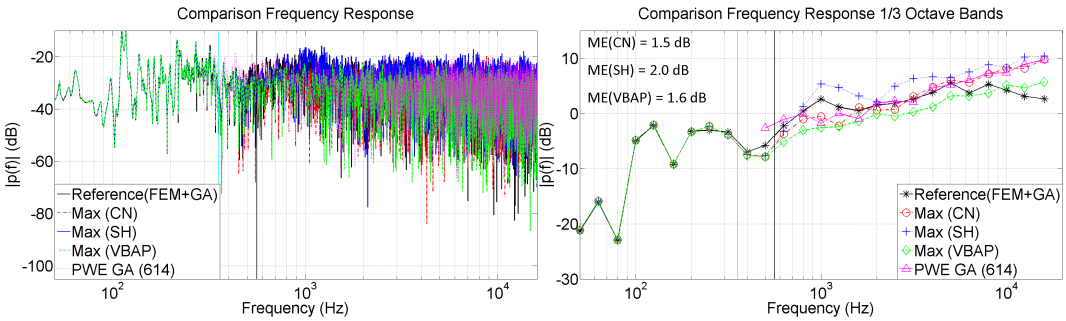


FIGURE 6.67: Comparison of frequency responses. Translated PWE (W) and reference (W) at the translated position 2, office room

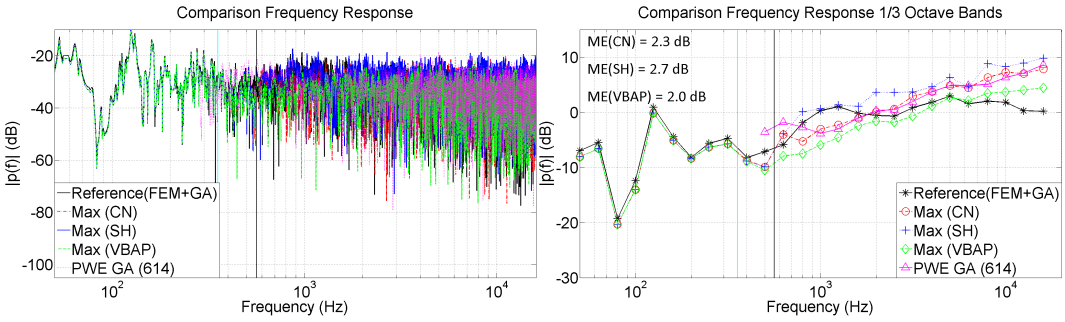


FIGURE 6.68: Comparison of frequency responses. Translated PWE (X) and reference (X) at the translated position 2, office room

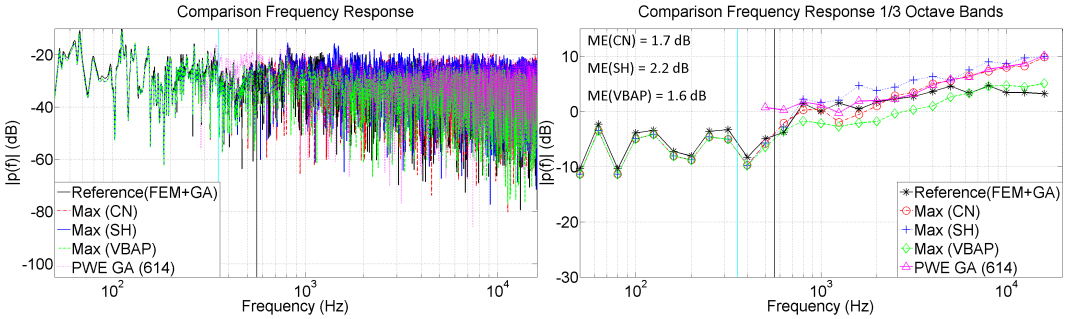


FIGURE 6.69: Comparison of frequency responses. Translated PWE (Y) and reference (Y) at the translated position 2, office room

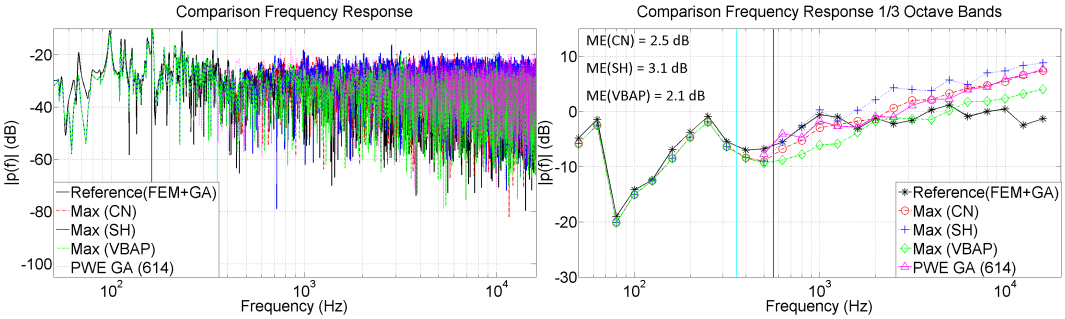


FIGURE 6.70: Comparison of frequency responses. Translated PWE (Z) and reference (Z) at the translated position 2, office room

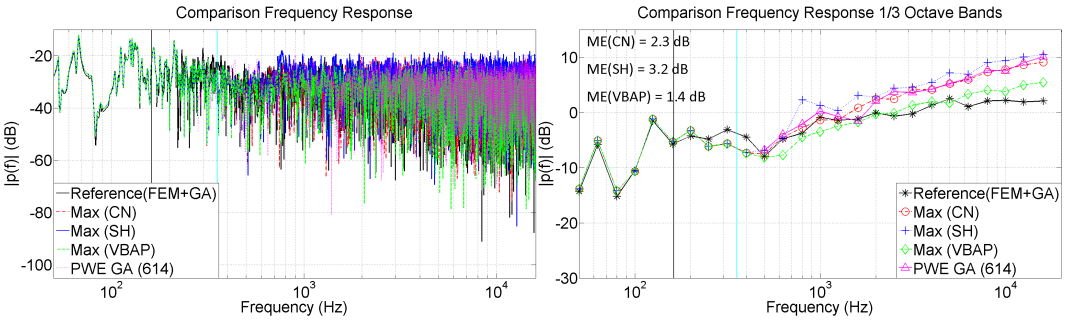


FIGURE 6.71: Comparison of frequency responses. Translated PWE (W) and reference (W) at the translated position 3, office room

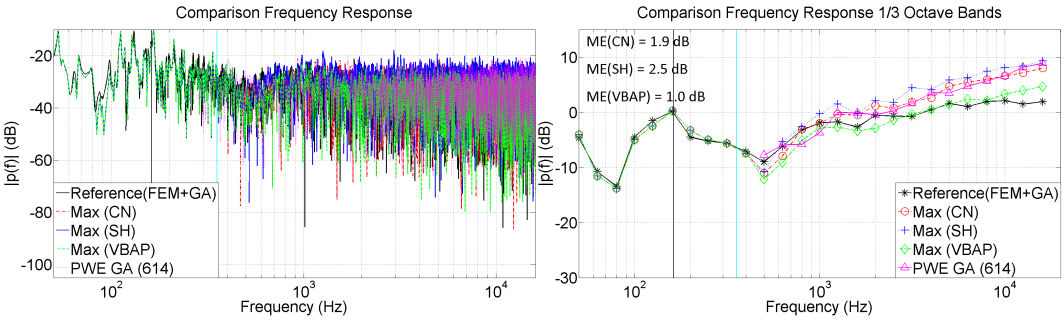


FIGURE 6.72: Comparison of frequency responses. Translated PWE (X) and reference (X) at the translated position 3, office room

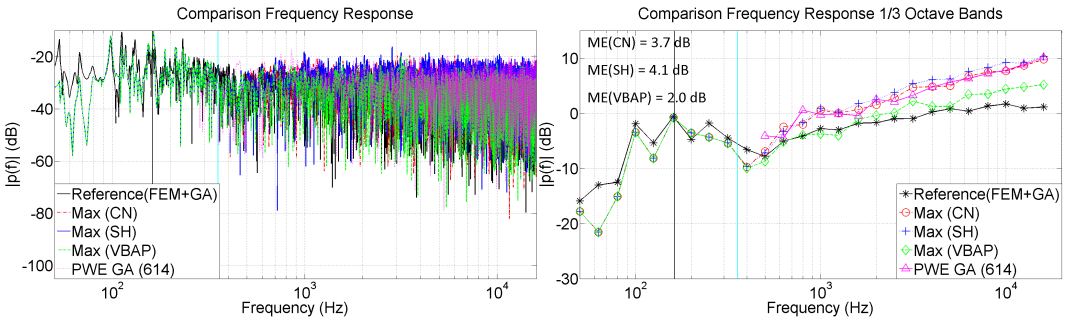


FIGURE 6.73: Comparison of frequency responses. Translated PWE (Y) and reference (Y) at the translated position 3, office room

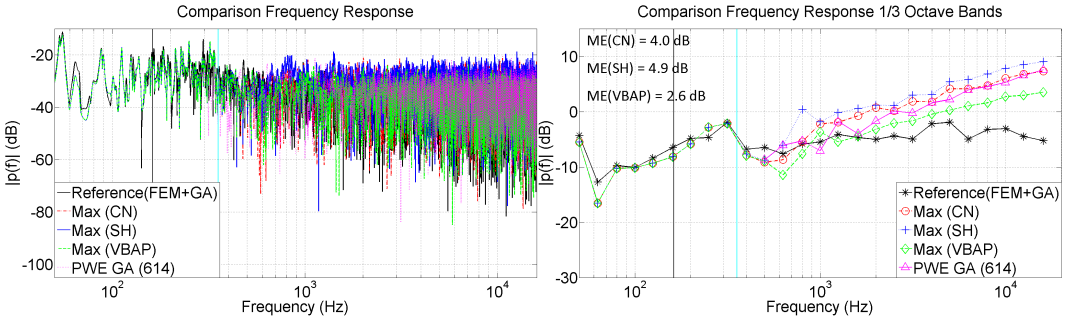


FIGURE 6.74: Comparison of frequency responses. Translated PWE (Z) and reference (Z) at the translated position 3, office room

Receiver	**Approach**	**W** (dB)	**Y** (dB)	**Z** (dB)	**X** (dB)	**Average** (dB)
2 (0.5m)	CN	1.5	1.7	2.5	2.3	2.0
	SH	2.0	2.2	3.1	2.7	2.5
	VBAP	1.6	1.6	2.1	2.0	1.8
3 (1.7m)	CN	2.3	3.7	4.0	1.9	3.0
	SH	3.2	4.1	4.9	2.5	3.7
	VBAP	1.4	2.0	2.6	1.0	1.8

TABLE 6.8: Mean errors in the B-format signal for the interactive auralization at different receiver locations, office room

The results indicate that the W signals are accurately reconstructed up to the frequency determined by equation (4.12). This relation holds as long as the crossover frequency is higher than the frequency predicted by equation (4.12); otherwise, the limit is given by the crossover frequency. A good agreement in terms of the frequency response in 1/3 octave band resolution was found for the other signals, namely Y, Z

and X. Compared to the meeting room, the mismatch in the octave band of 500 Hz is not as large, which is due to the use FE data in this frequency range. At high frequencies, the disagreement in the octave bands of 8 kHz and 16 kHz was also present for all signals and receivers. As expected, the outcomes show that the sound field reconstruction is more accurate at receiver 02, which is closer to the central point of the expansion. Regarding the approaches to resample the GA data, the lowest average mean error is achieved by using VBAP.

The global analysis of the results for the 4 previous reference cases indicates that the auralization system is able to reconstruct with a good level of accuracy the first order of spherical harmonic signals. The average mean error never exceeds 2.4 dB within a radius of 1.5 m for the closest neighbour approach, which was the most accurate method to resample the GA data (PWE of 614 elements) in 3 of the 4 enclosures. As expected, a higher agreement with the references was found at low frequencies where the synthesis of the acoustic field is carried out using FE data. At high frequencies, a good agreement in terms of the frequency response in 1/3 octave band resolution was obtained. Nevertheless, the level of this matching is mainly determined by the initial accuracy given by the translated plane wave expansion of 614 elements compared to the omnidirectional references (section 5.1.1.2). Regarding the B-format signals, the highest accuracy was achieved for W in all rooms. For the remaining signals, namely X, Y and Z, the accuracy depends on the specific enclosure. Finally, the results also confirm the limitations given by the chosen sound field model (plane waves) when an attempt is made to reconstruct a sound field of small spaces at mid frequencies using GA data.

6.5 Discussion

An implementation of the methods discussed in previous chapters was carried out to create an interactive auralization system. The interactive features correspond to the translation and rotation of the sound field. The auralization system was developed using the commercial packages Max and Unity. A translation operator is applied in the plane wave domain, which allows for the translation of the acoustic field. A limitation in the number of spherical harmonic coefficients that can be implemented to perform the rotation was found for the real-time implementation of the auralization system. This is due to the number of plane waves and the high order used in the spherical harmonic transformation, which requires a large amount of audio channels

to be processed in real-time. A strategy was proposed to reduce the effects given by a lower order based on the application of the translation operator before the rotation stage. This signal processing flow ensures that the translation is always performed correctly in the region determined by the initial accuracy and the output of the rotation algorithm must be accurate only in an area equivalent to the listener's head.

Finally, the analysis of the auralization system in terms of the accuracy to reconstruct the acoustic pressure and the first order of spherical harmonic coefficients indicates that the synthesis of the sound field is performed with good accuracy at low frequencies in the region defined by equation (4.12). At high frequencies, a good agreement in terms of the frequency response in 1/3 octave band resolution was found. Nevertheless, an overestimation of the energy in the octave bands of 8 kHz and 16 kHz was present for all the enclosures. The reason for this mismatch is unclear because the wavelengths at this frequency range are short enough to assume a plane wave propagation. It seems to be an intrinsic feature given by the methodology used to generate the plane wave expansion from GA data, but further work is required to gain a better understanding of this artefact.

Chapter 7

Conclusions

A complete framework for the generation of interactive auralization based on a plane wave expansion has been presented. This acoustic representation not only allows for interactive features such as the translation and rotation of sound fields, but is also compatible with a several sound reproduction techniques such as binaural, Ambisonics, WFS and VBAP, enhancing the auralization technique. The directional impulse responses corresponding to this plane wave representation are predicted by means of a combination of the finite element method and geometrical acoustics. The finite element method was implemented to calculate the low frequency content of the directional impulse responses whilst geometrical acoustics was used to estimate the middle and high frequency content for which its assumptions yield reasonable results.

A discussion of the modelling parameters has been presented for both simulation methods by using several reference cases. Although the combination of methods for the numerical solution of the wave equation and geometrical acoustics is not a novel topic in the scientific literature, the consolidation of the methodology and the choice of simulation parameters are still open topics for research in more general applications. In this thesis, an analysis of the use of reverberation time measurements to determine the effective specific acoustic impedance for finite element predictions has been carried out. This approach leads to a practical methodology to characterize the boundary conditions for FE simulations in existing rooms. An analysis of the results supports the suitability of this method in simple enclosures. However, in complex spaces such as the Ightham Mote, further investigation on how to tackle different modelling parameters is required. This is mainly required for FE simulations but in

some cases extends also for GA.

The implementation of the finite element method yielded an improvement in the acoustic prediction at low frequencies compared to geometrical acoustics for the meeting and office room. In this case, the modal responses of these two enclosures were more accurately estimated. Furthermore, the study of the assumption of a purely resistive specific acoustic impedance to characterize the boundary conditions in FE simulations indicated that the main implication is a shift in the predicted frequency response. This outcome was consistent for a simple analytical example and for different numerical test cases.

From the predicted acoustic data, different approaches to generate a plane wave expansion according to the simulation method (FEM-GA) have been proposed. The suitability of inverse methods to estimate a set of plane waves whose complex amplitudes reconstruct a target field has been proven. This technique is useful to extract directional information from meshed data that corresponds to the output from FE simulations. However, the discretization of the integral representation into a finite number of plane waves constrains the spatial accuracy of the sound field reconstruction. In other words, the discretization of the integral equation leads to a locally accurate representation of the sound field. The extent of the region in which the synthesis is accurate depends on the number of plane waves and the frequency of the field. The relation between frequency, number of plane waves and spatial location derived by Kennedy et al. [142] has been tested for several references cases. The outcomes indicated that this prediction is accurate as long as the acoustic field satisfies the homogeneous Helmholtz equation, i.e. there are no acoustic sources, scattering objects or reflective boundaries inside of the region of reconstruction.

The representation of acoustic fields that correspond to spherical wave propagation has also been considered. The curvature of the wave can be correctly represented but at the expense of a considerably higher amount of energy (compared to a plane wave field) and a spread spatial distribution in the plane wave density. However, an analysis of the acoustic fields reveals that a significant amount of that energy is actually used to reconstruct the sound field outside of the region where an accurate reconstruction is expected (radius given by equation (4.12)).

The use of Tikhonov regularization in the formulation of the inverse problem has three main effects on the sound field representation. The first is that the energy of the plane wave density used for the synthesis of the sound fields is considerably lower than for the non-regularized solution. The second consequence is that the energy distribution of the plane wave density is much more directionally concentrated. For simple fields such as a free field or a single reflective boundary, the energy of the plane wave density is directionally concentrated into the incoming direction of the wave(s). The last effect is a reduction of the area where the sound field reconstruction is accurate compared to the non-regularized solution. Nevertheless, the amplitude of the reconstructed field outside of the radius predicted by equation (4.12) is lower when regularization is applied. The implementation of regularization is convenient for an interactive auralization system because apart from it being a more efficient solution, the translation operator can lead to a zone with high acoustic pressure if no regularization is applied. The perceptual effects of a more concentrated plane wave density in the direction of the incoming propagation is an interesting topic for future research.

The equivalence of a discretized plane wave expnasion and a truncated spherical harmonic expansion has been studied for different analytical cases. Such equivalence is valid as long as the number of plane waves is equal to the number of spherical harmonics coefficients, namely $L = (N + 1)^2$. Then, the correspondence between these two sound field representations is given inside the area predicted by equation (4.12). Outside this region, the lack of higher orders in the spherical harmonic expansion leads to a different reconstructed sound field. However, a decoding of the spherical harmonic coefficients into a plane waves recreates "these higher orders" leading to a similar sound field synthesis to the plane wave expansion over the whole domain.

Regarding the plane wave expansion from the geometrical acoustic data, a time-spatial record of the reflections can be directly computed when a simulation is performed. However, the use of a commercial package restricted the amount of information that could be accessed. From the alternatives provided by Catt-Acoustics, the use of a high spatial resolution microphone to trace the reflections was chosen. The plane wave expansion is generated by exporting directional impulse responses according to a selected spatial discretization over a unit sphere. The main assumption is that the reflection paths are in fact plane waves propagating inside the space. The robustness of this postulation has been tested for several enclosures of different volumes. As expected, a higher accuracy is found in larger spaces or at distances

where the far field can be considered, which in turn, is frequency dependent.

A translation operator has been used to predict the sound field at different locations in the enclosure. The results indicate that this approach leads to a good agreement in terms of the frequency response in 1/3 octave band resolution when it is compared to omnidirectional impulse responses computed using Catt-Acoustics. In the worst scenario, a mean error of 3.1 dB was of found at a distance of 1.5 m. It is worth mentioning that the stochastic implementation of the scattering coefficients can lead to a mean error of 0.7 dB between trials of the simulations without any variation in the modelling parameters. In order to combine the FE and GA results, three different approaches have been proposed to resample the plane wave expansion obtained from geometrical acoustics into a number of plane waves that is equal to the plane wave expansion created from the FE data. Experiments indicate that the closest neighbour method is the most accurate approach to resample the GA data.

After a common basis in terms of the number of elements L in the plane wave expansion has been created for both simulation methods, a combination of the FE and GA data is carried out in the frequency domain using 8th order Butterworth filters. The low and high pass filters were designed to obtain a uniform magnitude and phase in the crossover point avoiding colouration in the frequency response. Based on a unified plane wave expansion, the tools required to generate an interactive auralization system have been proposed. The interactive features that have been generated in the current research correspond to the translation and rotation of the listener within the room. Modifications of the acoustic source in terms of its location and directivity pattern or variations in the acoustic properties of the boundaries are not considered. This is because the calculation of the directional impulse responses (plane wave expansion) is not performed in real-time. The equivalence between the rotation and translation of the listener as the respective rotation and translation of the sound field has been used to create the interactive features.

A theoretical formulation to perform a translation of acoustic fields has been presented. This approach allows for the translation of a given sound field over the whole domain. However, in the case of the low frequency content of the directional impulse responses (FE data), the error given by the discretization of the plane wave expansion restricts the area where the translation accurately synthesizes the target field.

At high frequencies, a good agreement in terms of the frequency response in 1/3 octave band resolution was achieved when the translation operator is applied within a radius of 1.5 m from the central point of the expansion.

Two approaches have been proposed to generate a rotation of the sound field. An implementation of Vector Base Amplitude Panning (VBAP) as an interpolation algorithm to predict the amplitudes of a set of plane waves, which rotates the acoustic field, has been discussed. Alternatively, the use of a spherical harmonic transformation has been also considered. An analytical formulation for the rotation in the azimuth plane has been derived for this domain. The plane waves that reconstruct the set of spherical harmonic coefficients are computed with a process based on a mode-matching approach. A comparison of these two methods reveals that the spherical harmonic transformation is a more accurate technique to rotate acoustic fields.

An interactive auralization system has been created using Max as an acoustic rendering and Unity as a graphical interface in which an avatar moves inside of a virtual enclosure. The signal processing is carried out based on a unified plane wave expansion and by the implementation of algorithms to translate and rotate the acoustic fields. In order to improve the accuracy of the synthesis of the acoustic fields, the rotation is always performed after the translation. This means that the output of the rotation operator only has to be accurate in a radius equivalent to the half of listener's head. The implementation of the auralization system reveals limitations when encoding a high number of plane waves into a high order of spherical harmonics due to computational resources. In the current research encoding up to a 5th order of spherical harmonics was possible to process in real-time using a plane wave expansion of 64 plane waves.

The evaluation of the auralization system has been carried out in the frequency domain using as reference numerical simulations of the enclosures. The assumption of this approach is that the FE-GA simulations provide an adequate description of the acoustics of the spaces. It is worth mentioning that the scientific contribution of this thesis is not to prove that the combination of methods for the numerical solution of the wave equation and geometrical acoustics is a suitable approach to predict room impulse responses. This has been already addressed by several researchers. The novelty in this thesis is the generation and manipulation of a plane wave expansion from

FE-GA simulations which allow for interactive features in a real-time auralization system.

The use of numerical information as a reference solution was selected due to a lack of sufficient measured (omnidirectional and spatial) impulse responses at different positions of the enclosures. Multiple field points were selected for several reference cases in which the acoustic pressure and the B-format signals (first order) were estimated from the finite element and geometrical acoustics simulations. This information is compared to the real-time acoustic rendering produced by the auralization system at these field points. The mean error between the reference and predicted signals has been chosen as a metric for the analysis. The results indicate that the auralization system is able to recreate the acoustic pressure and the first order of spherical harmonic coefficients with good accuracy within a radius of 1.5 m. In this case the average mean error never exceeds the 3.3 dB for the closest neighbour approach which was the most accurate method for resampling the GA data in three of the four enclosures considered.

The methodology based on plane waves provides a convenient representation, which allows for the interaction of different approaches (VBAP, spherical harmonics) and several reproduction techniques. This framework enhances the practicality and applicability of the auralization technique. The potential to reconstruct sound fields which have been predicted by means of methods for the numerical solution of the wave equation expands the branch of problems in which the auralization technique can be used as a tool for the analysis. The suitability of generating an interactive auralization improves the applications the proposed method. Nevertheless, two main fundamental limitations constrain the sound field reconstruction using a plane wave expansion. The first corresponds to the plane wave model, which it is constrained in small rooms where the far field assumption is not fullfilled. The second is the discretization of the integral equation, which leads to a locally accurate reconstruction of the acoustic field.

Finally, it is important to mention that the analysis of the performance of the auralization system has been addressed based on the accuracy of the reconstructed acoustic fields, which has been carried out with objective metrics. No perceptual evaluations of the auralization system have been conducted. The use of listening

tests can be an alternative approach to validate and improve the methods presented in this thesis. This is planned to be carried out in the near future.

Chapter 8

Further Work

An interesting topic for future research is the disagreement found in the octave bands of 8 kHz and 16 kHz between the predicted and reference sound fields. This particular result is associated with the method for the creation of the plane wave expansion from the GA data. However, further research is needed to establish the physical reasons behind this result. In addition, the use of a commercial package restricted the analysis and validation of the geometrical acoustics data. Alternative methods to generate the plane wave expansion such as the computation of high order spherical harmonic signals (e.g. equivalent to the number of plane waves) or the tracing of the reflection paths in real-time when the simulation is performed seem to be more robust approaches to processing the GA data. This is a topic for future research, which could lead to more accurate results at high frequencies.

Due to the local description of the sound field provided by the use of the plane wave expansion, a significant improvement of the proposed method would be the generation of a more global representation of the acoustics of the enclosure. This can be done by the synthesis of plane wave expansions which are centred at different positions within the room. An algorithm that interpolates the amplitude of the plane waves can be implemented allowing a more robust translation and rotation of the acoustic field. The suitability of this interpolation approach is the next stage to be investigated.

Finally, the implementation of the proposed method in other areas of acoustics such as noise control is an interesting topic for future research. One example is the auralization of the changes in the acoustic field due to the installation of an acoustic barrier. Investigations about the applicability of the method in this type of problem will be addressed in the future.

Appendix A

Absorption and scattering coefficients used in GA simulations

A.1 Meeting room

Material	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz
Lights	0.02	0.02	0.03	0.02	0.03	0.03
Foam	0.02	0.03	0.04	0.04	0.05	0.06
Glass	0.33	0.24	0.17	0.09	0.05	0.03
Ceiling	0.04	0.03	0.03	0.04	0.05	0.06
Wood	0.04	0.04	0.07	0.04	0.04	0.05
Wood door	0.13	0.10	0.06	0.06	0.07	0.07
Wood table	0.02	0.02	0.02	0.01	0.02	0.02
Acrylic	0.01	0.01	0.01	0.01	0.01	0.02
Brick	0.01	0.02	0.02	0.01	0.02	0.02
Carpet	0.02	0.06	0.14	0.26	0.43	0.49
Plaster	0.25	0.20	0.10	0.07	0.05	0.06

Table A.1: Absorption coefficients used in GA simuations, meeting room

Material	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz
Lights	0.3	0.3	0.3	0.3	0.3	0.3
Foam	0.1	0.1	0.1	0.1	0.1	0.1
Glass	0.2	0.2	0.2	0.2	0.2	0.2
Ceiling	0.2	0.2	0.2	0.2	0.2	0.2
Wood	0.2	0.2	0.2	0.2	0.2	0.2
Wood door	0.2	0.2	0.2	0.2	0.2	0.2
Wood table	0.2	0.2	0.2	0.2	0.2	0.2
Acrylic	0.1	0.1	0.1	0.1	0.1	0.1
Brick	0.2	0.2	0.2	0.2	0.2	0.2
Carpet	0.1	0.1	0.1	0.1	0.1	0.1
Plaster	0.2	0.2	0.2	0.2	0.2	0.2

Table A.2: Scattering coefficients used in GA simuations, meeting room

A.2 Ightham Mote

Material	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz
Paint-2	0.01	0.01	0.01	0.01	0.02	0.02
Lamp	0.01	0.01	0.01	0.01	0.02	0.02
Paint	0.08	0.08	0.14	0.18	0.20	0.23
Wood-E	0.14	0.1	0.06	0.08	0.1	0.1
Wood-H	0.6	0.55	0.26	0.16	0.16	0.1
Brick	0.09	0.14	0.17	0.17	0.17	0.21
Carpet	0.09	0.08	0.21	0.26	0.27	0.37
Tilefloor	0.02	0.02	0.02	0.02	0.03	0.03
Tapestry	0.17	0.25	0.41	0.06	0.82	0.82
Wood-W	0.14	0.1	0.06	0.08	0.1	0.1
Glass	0.35	0.25	0.18	0.12	0.07	0.04

Table A.3: Absorption coefficients used in GA simuations, Ightham Mote

Material	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz
Paint-2	0.1	0.1	0.1	0.1	0.1	0.1
Lamp	0.1	0.1	0.1	0.1	0.1	0.1
Paint	0.13	0.2	0.2	0.3	0.3	0.4
Wood-E	0.1	0.1	0.1	0.2	0.2	0.3
Wood-H	0.1	0.1	0.1	0.2	0.2	0.3
Brick	0.13	0.2	0.2	0.3	0.3	0.4
Carpet	0.1	0.1	0.1	0.1	0.1	0.1
Tilefloor	0.1	0.1	0.1	0.2	0.2	0.3
Tapestry	0.1	0.1	0.1	0.1	0.1	0.1
Wood-W	0.3	0.5	0.7	0.7	0.8	0.9
Glass	0.2	0.2	0.2	0.2	0.3	0.3

TABLE A.4: Scattering coefficients used in GA simuations, Ightham Mote

A.3 Office room

Material	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz
Shelf	0.03	0.03	0.07	0.04	0.04	0.05
Blackboard	0.01	0.01	0.01	0.01	0.01	0.01
Window	0.26	0.2	0.17	0.08	0.04	0.03
Wood door	0.1	0.08	0.06	0.06	0.06	0.06
Carpet	0.01	0.04	0.13	0.26	0.38	0.42
Concrete	0.1	0.10	0.02	0.02	0.02	0.02
Plaster	0.08	0.08	0.07	0.04	0.03	0.03
Acrylic	0.01	0.01	0.01	0.01	0.01	0.01

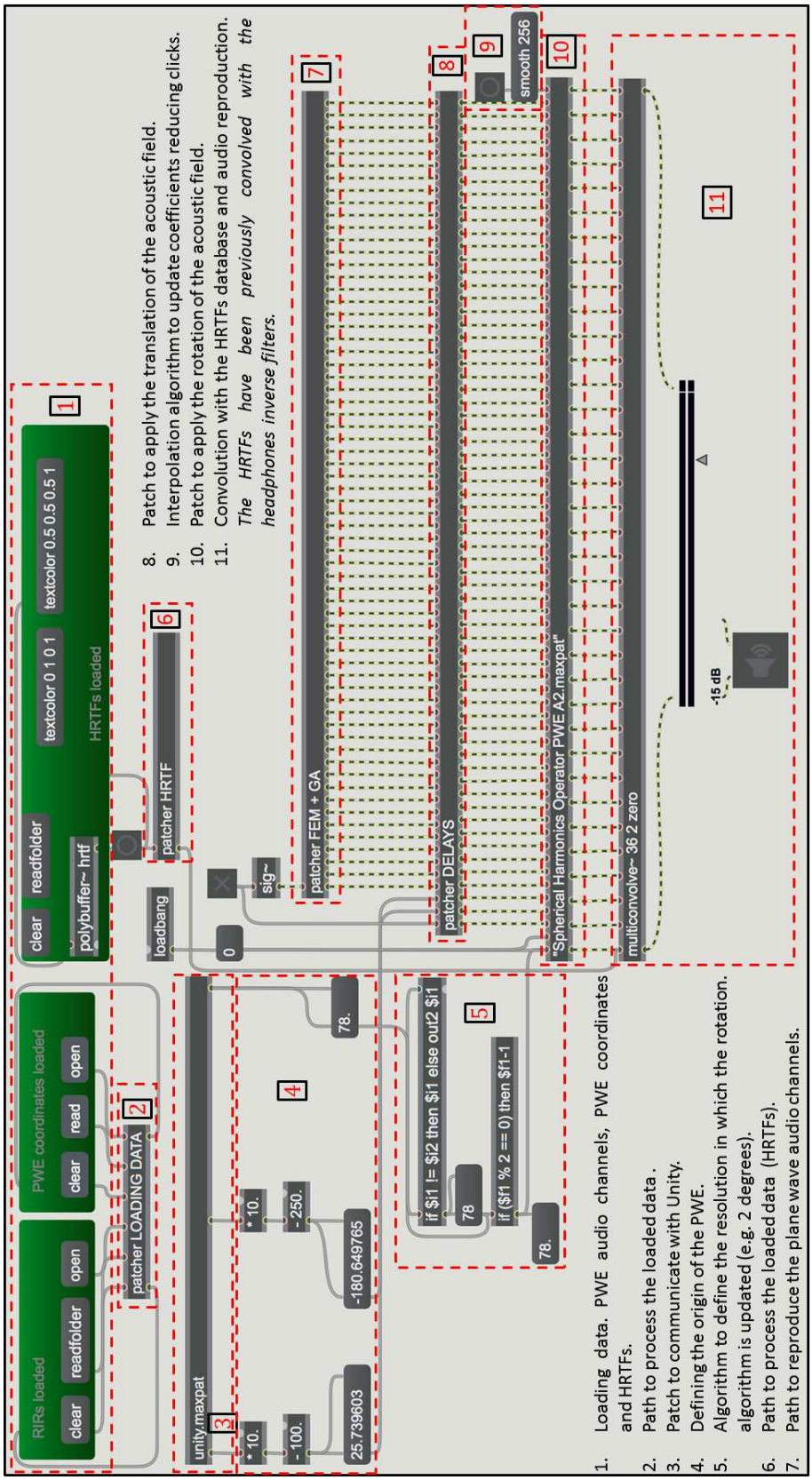
TABLE A.5: Absorption coefficients used in GA simuations, office room

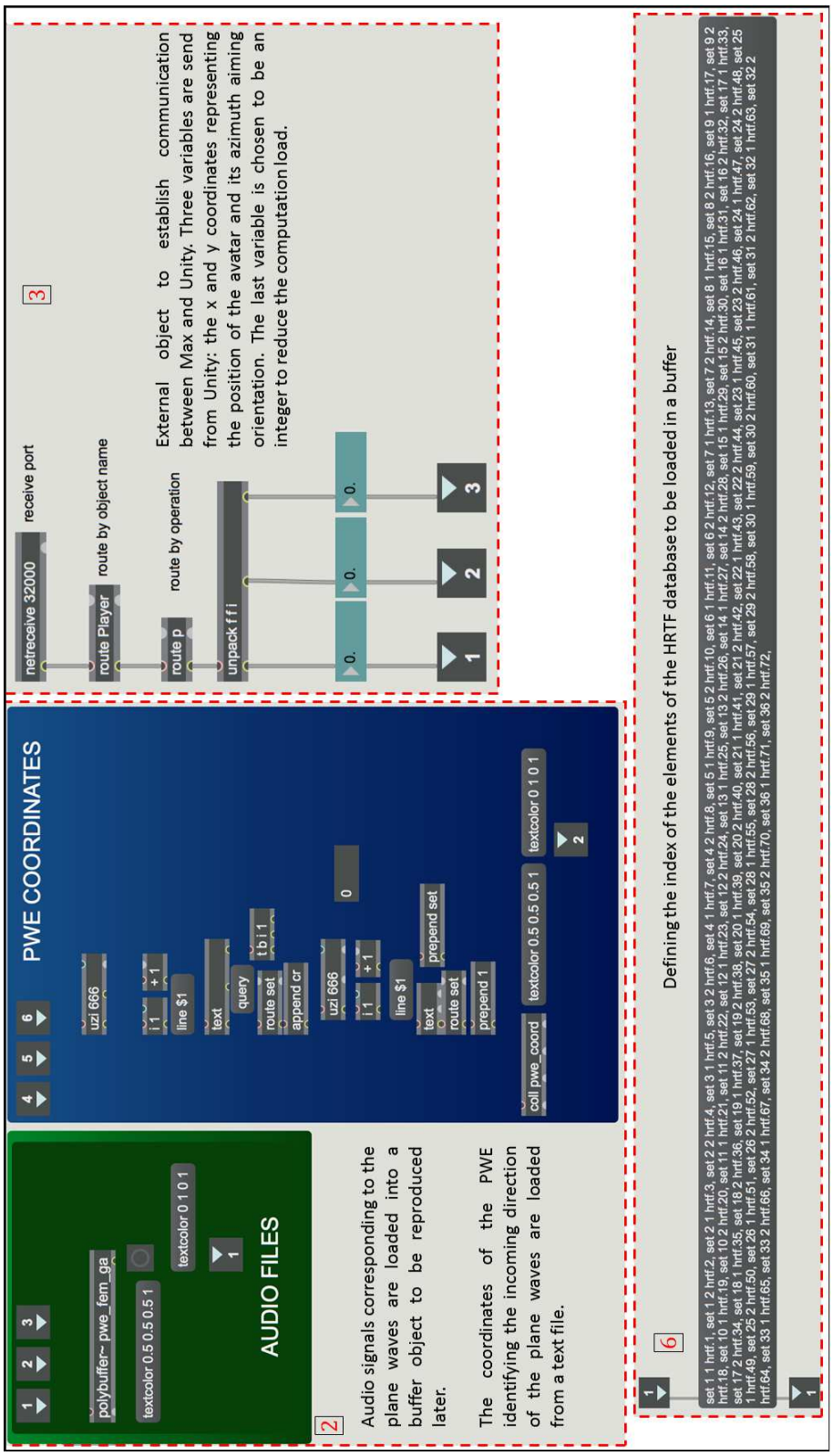
Material	**125 Hz**	**250 Hz**	**500 Hz**	**1000 Hz**	**2000 Hz**	**4000 Hz**
Shelf	0.2	0.2	0.2	0.2	0.2	0.2
Blackboard	0.1	0.1	0.1	0.1	0.1	0.1
Window	0.2	0.2	0.2	0.2	0.2	0.2
Wood door	0.1	0.1	0.1	0.1	0.1	0.1
Carpet	0.1	0.1	0.1	0.1	0.1	0.1
Concrete	0.2	0.2	0.2	0.2	0.2	0.2
Plaster	0.2	0.2	0.2	0.2	0.2	0.2
Acrylic	0.1	0.1	0.1	0.1	0.1	0.1

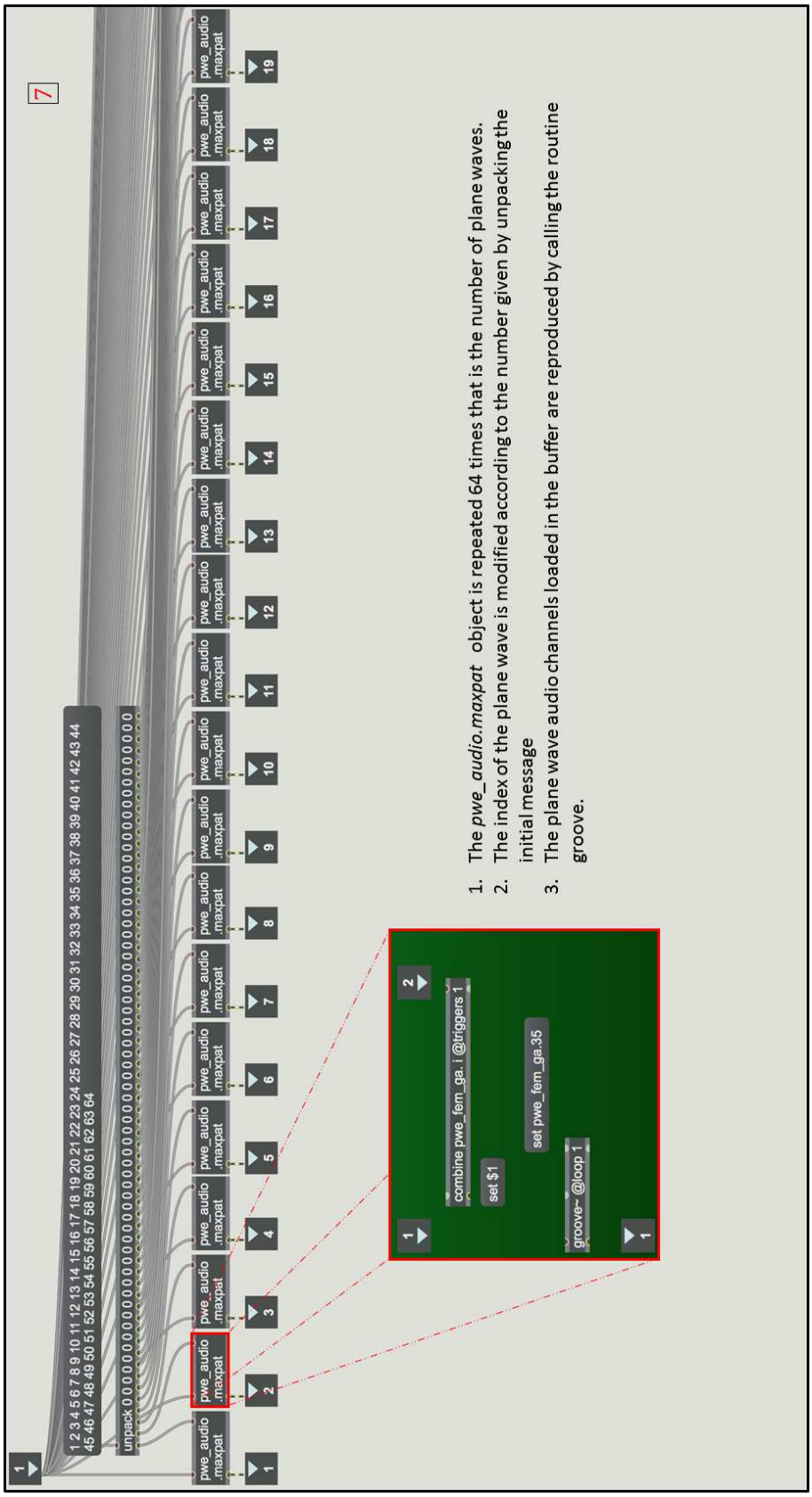
TABLE A.6: Scattering coefficients used in GA simuations, office room

Appendix B

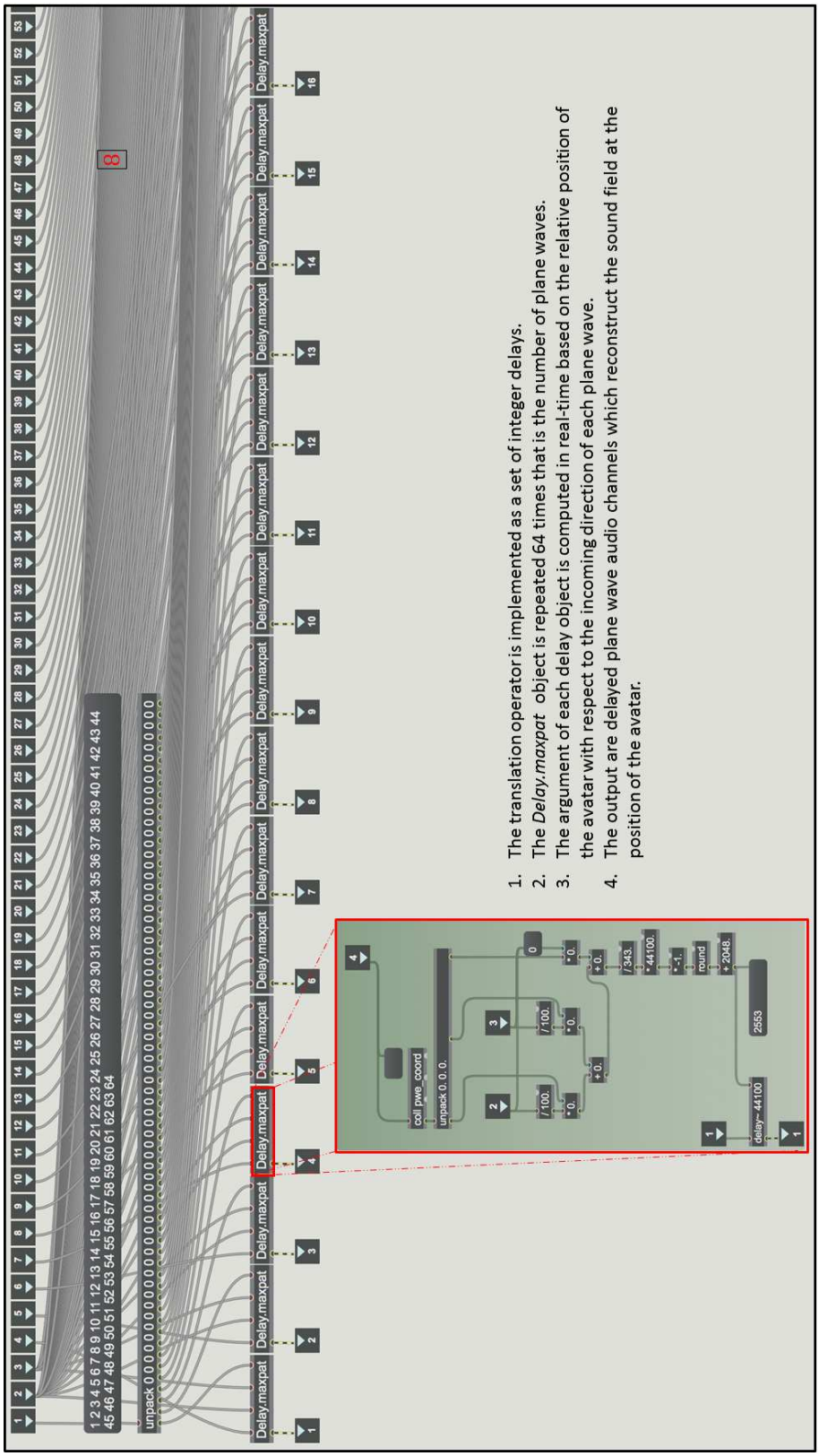
Max patches of the auralization system


RIRs loaded
clear
readfolder
open
PWE coordinates loaded
clear
read
open
patcher LOADING DATA
unity.maxpat
* 10.
- 100.
25.739603
* 10.
- 250.
-180.649765
78.
if $i1 != $i2 then $i1 else out2 $i1
78
if ($i1 % 2 == 0) then $i1-1
78.
clear
readfolder
textcolor 0 1 0 1
textcolor 0.5 0.5 0.5 1
polybuffer~ hrtf
HRTFs loaded
loadbang
patcher HRTF
0
sig~
patcher FEM + GA
patcher DELAYS
"Spherical Harmonics Operator PWE A2.maxpat"
multiconvolve~ 36 2 zero
smooth 256
-15 dB
1
2
3
4
5
6
7
8
9
10
11
1. Loading data. PWE audio channels, PWE coordinates and HRTFs.
2. Path to process the loaded data .
3. Patch to communicate with Unity.
4. Defining the origin of the PWE.
5. Algorithm to define the resolution in which the rotation. algorithm is updated (e.g. 2 degrees).
6. Path to process the loaded data (HRTFs).
7. Path to reproduce the plane wave audio channels.
8. Patch to apply the translation of the acoustic field.
9. Interpolation algorithm to update coefficients reducing clicks.
10. Patch to apply the rotation of the acoustic field.
11. Convolution with the HRTFs database and audio reproduction. *The HRTFs have been previously convolved with the headphones inverse filters.*

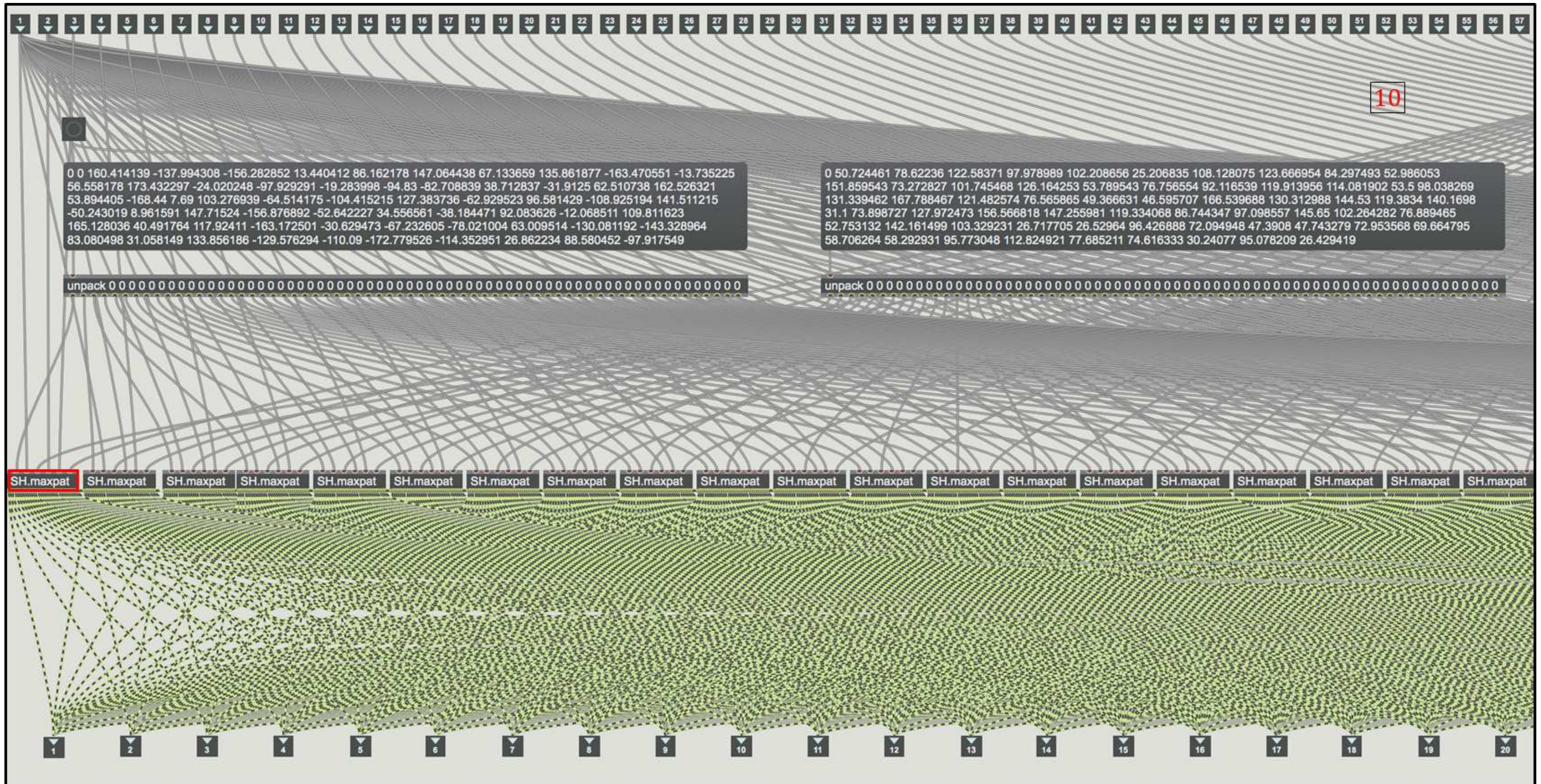

AUDIO FILES
polybuffer~ pwe_fem_ga
textcolor 0.5 0.5 0.5 1
textcolor 0 1 0 1
2
Audio signals corresponding to the plane waves are loaded into a buffer object to be reproduced later.
The coordinates of the PWE identifying the incoming direction of the plane waves are loaded from a text file.
PWE COORDINATES
uzi 666
i 1
+ 1
line $1
text
query
route set
append cr
t b i 1
uzi 666
0
i 1
+ 1
line $1
text
prepend set
route set
prepend 1
coll pwe_coord
textcolor 0.5 0.5 0.5 1
textcolor 0 1 0 1
3
netreceive 32000
receive port
route Player
route by object name
route p
route by operation
unpack f f i
External object to establish communication between Max and Unity. Three variables are send from Unity: the x and y coordinates representing the position of the avatar and its azimuth aiming orientation. The last variable is chosen to be an integer to reduce the computation load.
6
Defining the index of the elements of the HRTF database to be loaded in a buffer



1. The *pwe_audio.maxpat* object is repeated 64 times that is the number of plane waves.
2. The index of the plane wave is modified according to the number given by unpacking the initial message
3. The plane wave audio channels loaded in the buffer are reproduced by calling the routine groove.



1. The translation operator is implemented as a set of integer delays.
2. The *Delay.maxpat* object is repeated 64 times that is the number of plane waves.
3. The argument of each delay object is computed in real-time based on the relative position of the avatar with respect to the incoming direction of each plane wave.
4. The output are delayed plane wave audio channels which reconstruct the sound field at the position of the avatar.



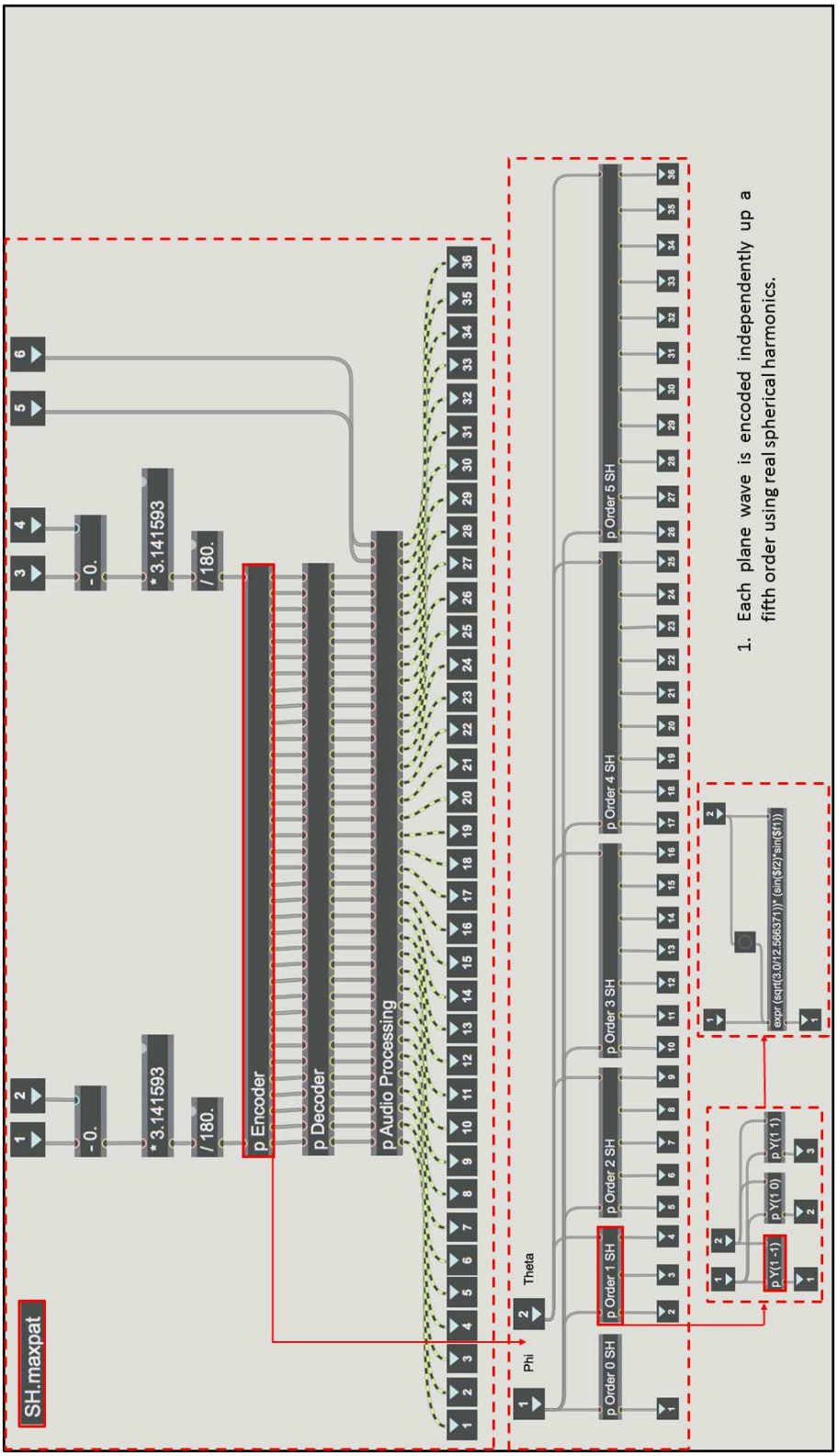
```
0 0 160.414139 -137.994308 -156.282852 13.440412 86.162178 147.064438 67.133659 135.861877 -163.470551 -13.735225
56.558178 173.432297 -24.020248 -97.929291 -19.283998 -94.83 -82.708839 38.712837 -31.9125 62.510738 162.526321
53.894405 -168.44 7.69 103.276939 -64.514175 -104.415215 127.383736 -62.929523 96.581429 -108.925194 141.511215
-50.243019 8.961591 147.71524 -156.876892 -52.642227 34.556561 -38.184471 92.083626 -12.068511 109.811623
165.128036 40.491764 117.92411 -163.172501 -30.629473 -67.232605 -78.021004 63.009514 -130.081192 -143.328964
83.080498 31.058149 133.856186 -129.576294 -110.09 -172.779526 -114.352951 26.862234 88.580452 -97.917549
```

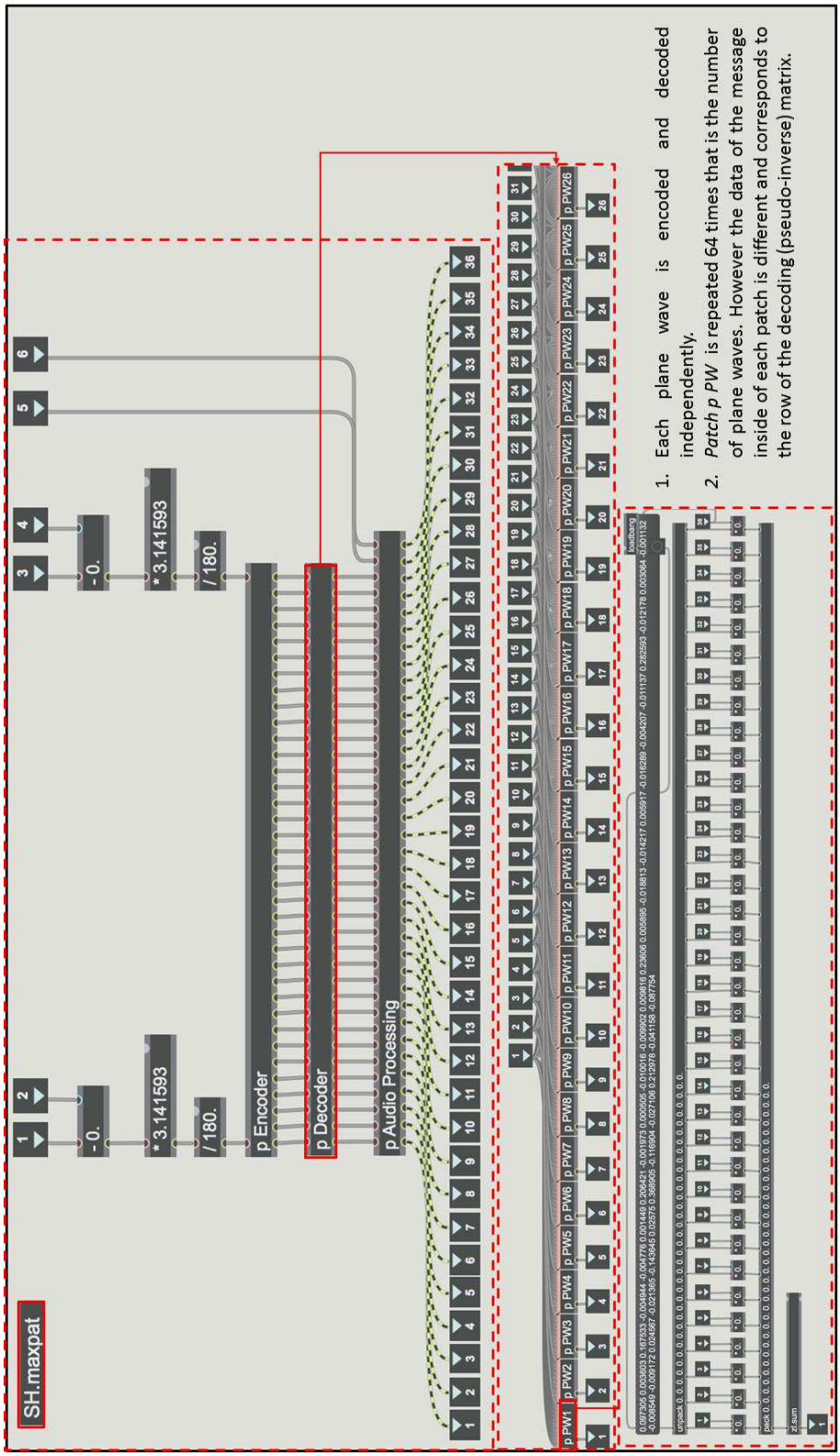
```
unpack 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
```

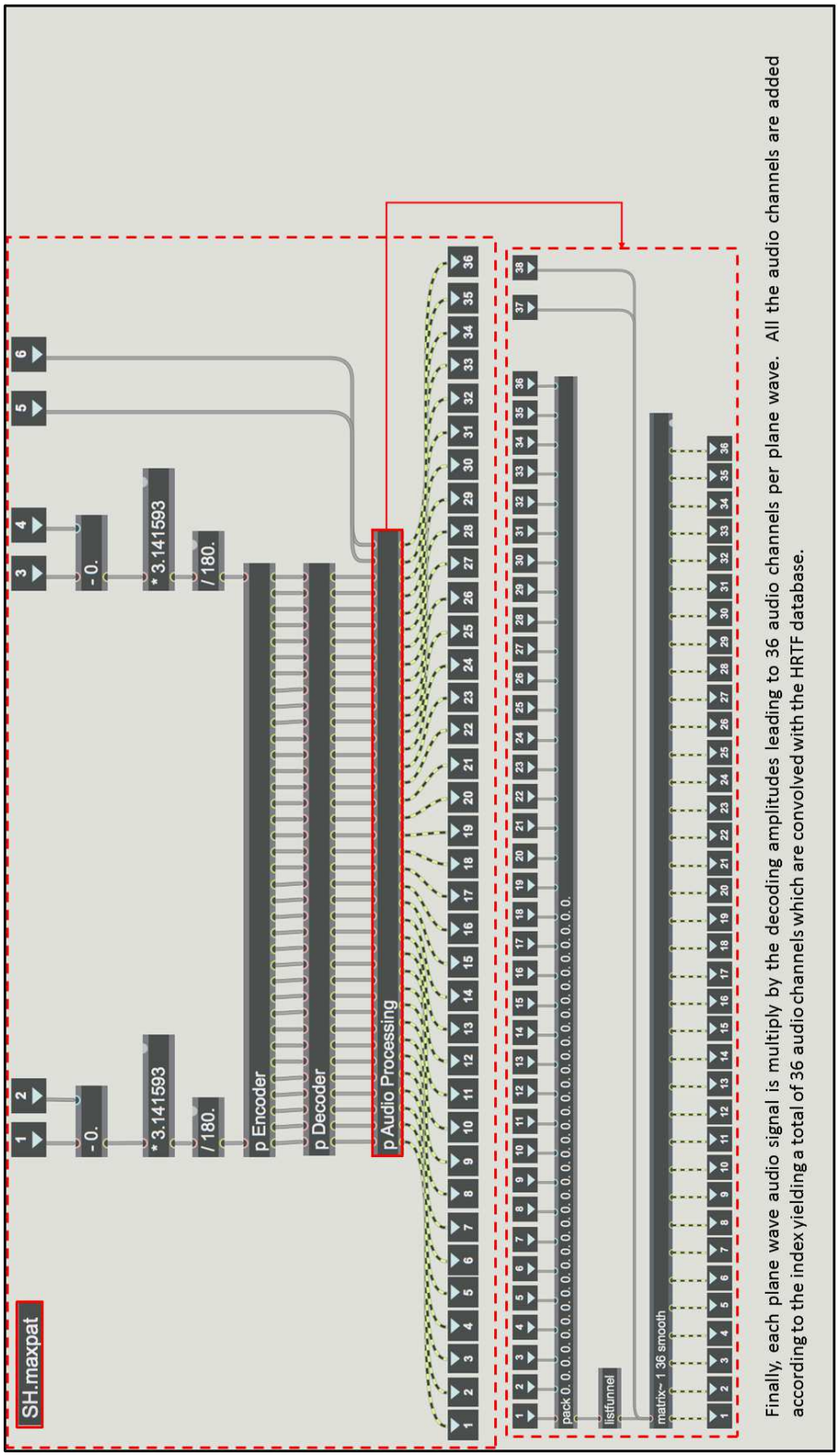
```
0 50.724461 78.62236 122.58371 97.978989 102.208656 25.206835 108.128075 123.666954 84.297493 52.986053
151.859543 73.272827 101.745468 126.164253 53.789543 76.756554 92.116539 119.913956 114.081902 53.5 98.038269
131.339462 167.788467 121.482574 76.565865 49.366631 46.595707 166.539688 130.312988 144.53 119.3834 140.1698
31.1 73.898727 127.972473 156.566818 147.255981 119.334068 86.744347 97.098557 145.65 102.264282 76.889465
52.753132 142.161499 103.329231 26.717705 26.52964 96.426888 72.094948 47.3908 47.743279 72.953568 69.664795
58.706264 58.292931 95.773048 112.824921 77.685211 74.616333 30.24077 95.078209 26.429419
```

```
unpack 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
```

1. The rotation operator is implemented using a spherical harmonics transformation.
2. The *SH.maxpat* object is repeated 64 times that is the number of plane waves.
3. The argument of each *SH* object is computed in real-time based on the difference between the incoming angles of the plane waves and the aiming angle of the avatar.
4. The output are a set of RIRs which reconstruct the translated and rotated sound field at the position of the avatar.


SH.maxpat
- 0.
*3.141593
/180.
p Encoder
p Decoder
p Audio Processing
Theta
Phi
p Order 0 SH
p Order 1 SH
p Order 2 SH
p Order 3 SH
p Order 4 SH
p Order 5 SH
p Y(1 -1)
p Y(1 0)
p Y(1 1)
expr (sqrt(3.0/12.566371))*(sin($f2)*sin($f1))
1. Each plane wave is encoded independently up a fifth order using real spherical harmonics.


SH.maxpat
p Encoder
p Decoder
p Audio Processing
* 3.141593
/ 180.
- 0.
1. Each plane wave is encoded and decoded independently.
2. *Patch p PW* is repeated 64 times that is the number of plane waves. However the data of the message inside of each patch is different and corresponds to the row of the decoding (pseudo-inverse) matrix.


SH.maxpat
- 0.
* 3.141593
/ 180.
p Encoder
p Decoder
p Audio Processing
pack 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
listfunnel
matrix~ 1 36 smooth
Finally, each plane wave audio signal is multiply by the decoding amplitudes leading to 36 audio channels per plane wave. All the audio channels are added according to the index yielding a total of 36 audio channels which are convolved with the HRTF database.

Appendix C

Optimization of the high frequency reconstruction by means of Ambisonics decoding methods

Due to the artifacts created by the truncation in the number of spherical harmonic coefficients, different decoding methods have been proposed to increase the physical or perceptual accuracy of the sound field reconstruction. The implementation of these types of decoder is made by applying a monotonically decreasing weighting function (like a "fade out") to the spherical harmonic coefficients, which describes the acoustic filed. Consequently, each decoder yields a different sound field reconstruction performance. The concept of this weighting function can be explained using an analogy to the Fourier transform of a Dirac Delta function with different window types. Figure C.1 shows signals created by applying different frequency-domain windows to a Dirac delta function. According to the window type selected, the energy of the complex coefficients is weighted in different proportion leading to an altered signal when the inverse Fourier transform is applied.

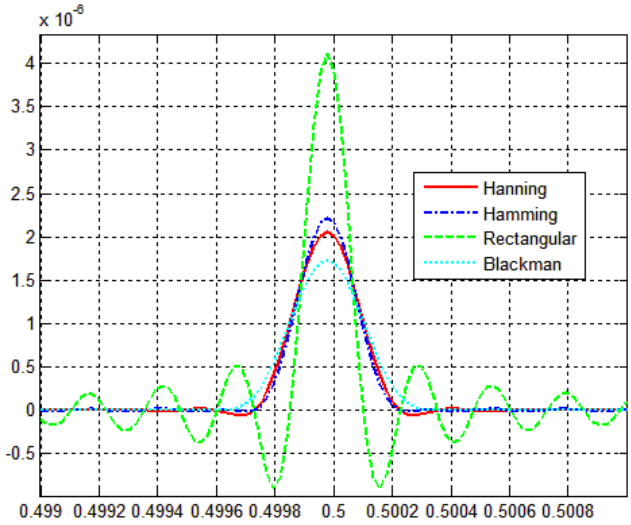


FIGURE C.1: Reconstructed Delta Dirac signal

One approach to decode Ambisonics signals is Max-rE, which aims to maximize the energy vector optimizing the sound reproduction at high frequencies [152]. The energy vector is defined as [80]:

$$rE \cdot \mathbf{u_E} = \frac{\sum_{l=1}^{L} g_l^2 \widehat{\mathbf{y}}_l}{\sum_{l=1}^{L} g_l^2}, \tag{C.1}$$

where L represents a superposition of plane waves incoming from the directions $\widehat{\mathbf{y}}_l$ at the reference point ($r = 0$). Therefore, the energy vector is the sum of the unitary vectors $\widehat{\mathbf{y}}$ weighted by their respective amplitude gains squared (g^2) and divided by the sum of their amplitude gains squared. The value of rE compounds between 0 and 1 and the unitary vector u_E is defined as $u_E = (cos\theta_E, sin\theta_E)$. A different method is the In-Phase decoder, which recreates the condition that the loudspeakers feed the signals in phase decreasing the localization artifacts [80]. Finally, a unitary gain is applied to all orders up to the truncation order N for the Basic decoding method, which means that every order of spherical harmonics has the same contribution to the reconstructed sound field.

The performance of three decoding approaches (Basic, Max-rE and In-Phase) has been studied and evaluated in the light of the results of experimental measurements. The latter were performed using a spherical array composed of 40 uniformly distributed loudspeakers and a 5th order Ambisonics system. The measurements were

conducted in the anechoic chamber of the ISVR using a translating microphone array composed by 29 transducers (Ref. Brüel & Kjær 4189-L001) across 40 positions. The total number of measured points corresponded to 1160 with a spatial resolution of 0.05 m leading to an approximate spatial alias frequency of 3.4 kHz. An error analysis is presented based on the difference between the target (far field spherical wave) and the synthesized acoustic pressure and intensity field.

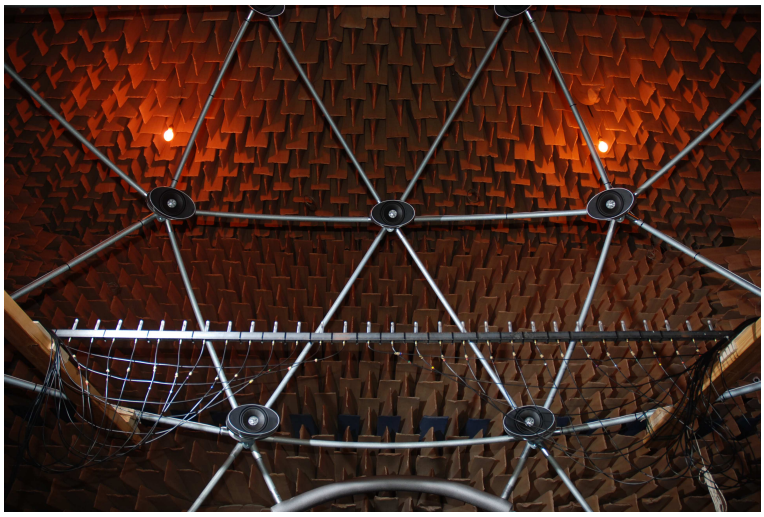


FIGURE C.2: Measurement of the acoustic field produced by a 5th order Ambisonics system

The algorithm to test the performance of the Ambisonics decoding methods was developed using the software package Max. Figure C.3 shows the architecture of the decoder with its respective modules. The first part corresponds to the encoding stage using up to 5th order of spherical harmonics. Then, the resulting signals were weighted by a g_n function according to the chosen type of decoder (Basic, Max-rE or In-Phase). The values for the g_n functions were calculated using the methodology suggested by Jerome Daniel [80]. Finally, at the last stage, the signals are decoded using a decoding matrix obtained by a mode-matching approach and reproduced them by the loudspeaker array.

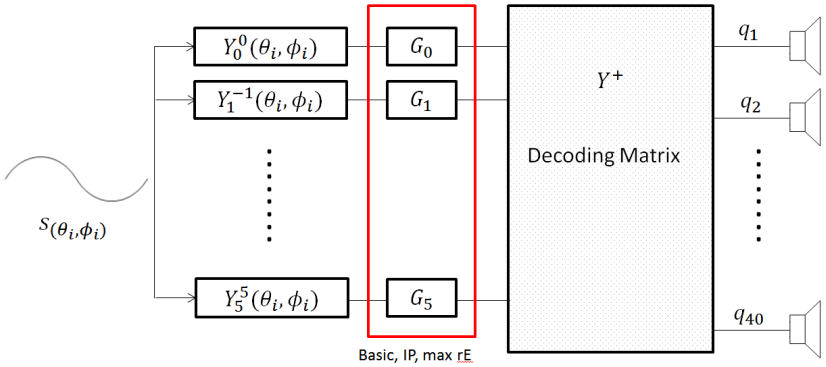


FIGURE C.3: Ambisonics decoder

The sound pressure field was directly computed from the measurements. However, in the case of the acoustic intensity, the values were determined by taking the real part of the product between the sound pressure $p(\mathbf{x})$ and the complex conjugate of the particle velocity $\mathbf{u}(\mathbf{x})$ (see equation C.2). The particle velocity was calculated based on the Euler's equation (equation C.3) by approximating the gradient of the pressure as the difference between neighbouring sound pressure measurement positions (equation C.4).

$$\mathbf{I}(\mathbf{x},\omega) = \frac{1}{2} Re\left\{p(\mathbf{x})\mathbf{u}(\mathbf{x})^*\right\}, \tag{C.2}$$

$$\mathbf{u}(\mathbf{x}) = -\frac{\nabla p(\mathbf{x})}{j\omega\rho_0}, \tag{C.3}$$

$$\mathbf{u}(\mathbf{x}) \approx -\frac{1}{j\omega\rho_0}\frac{[p(\mathbf{x}+\mathbf{dx}) - p(\mathbf{x})]}{d\mathbf{x}}, \tag{C.4}$$

where $\nabla p(\mathbf{x})$ is the gradient of the pressure and ρ_0 is the static density of the air. The measured acoustic pressure and acoustic intensity flow field at 250 Hz, 1 kHz and 2 kHz are presented in Figures C.4, C.5 and C.6, respectively. The black circle represents the region of validity which is determined from equation (4.12). The measurement procedure involved the recording of the sound field generated by each type of decoder using the microphone array. The excitation signal corresponded to a virtual point source (white noise) located at ($\theta = 90$, $\phi = 45$) and 2.3 m far away.

The comparison between Figures C.4, C.5 and C.6 clearly identifies the constraint to reproduce high frequencies over a wide listener area. The radius of validity provides an insight on the area where the sound field reconstruction is accurate. However, it was found that this assumption is not always valid and depends strongly on the decoder. Regarding the acoustic intensity flow, the Figures C.4, C.5 and C.6 show a good agreement between the reference radius and the direction of the intensity flow. However, at 1 kHz and 2 kHz only by the Basic and Max-rE decoders match the radius.

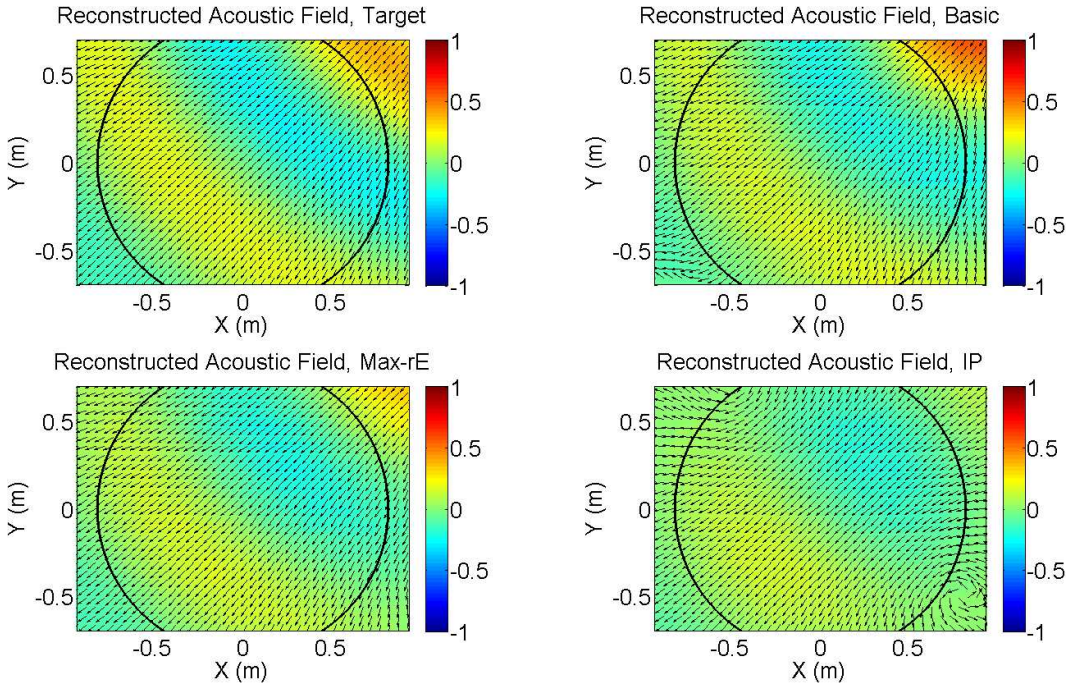


FIGURE C.4: Acoustic field reconstruction Ambisonics decoding methods, 250 Hz

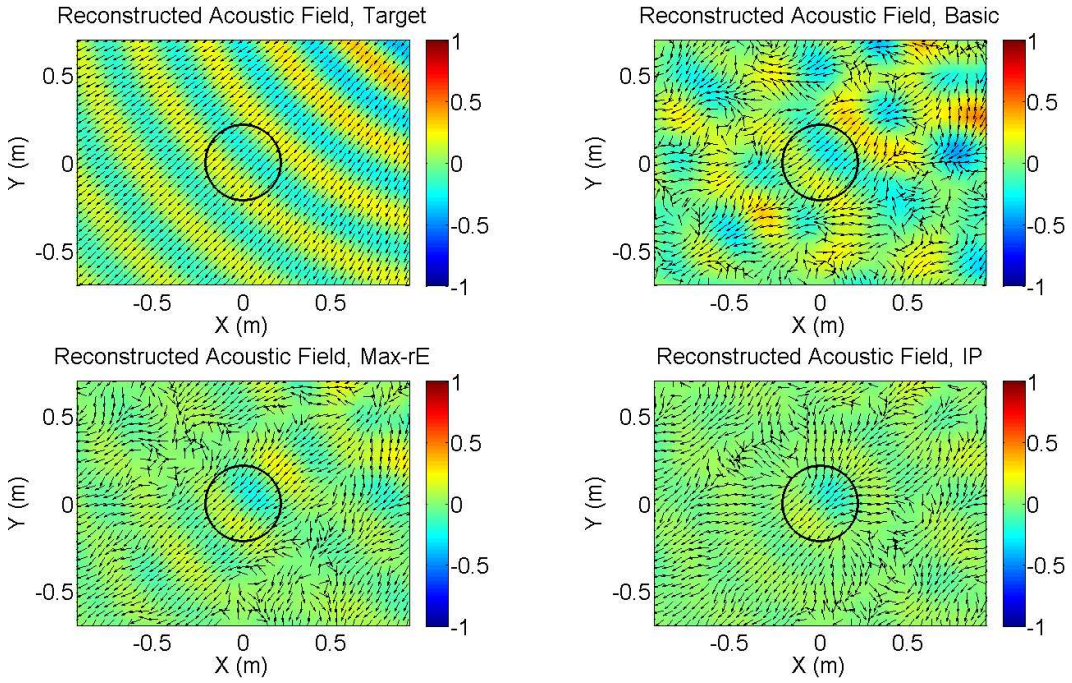


FIGURE C.5: Acoustic field reconstruction Ambisonics decoding methods, 1 kHz

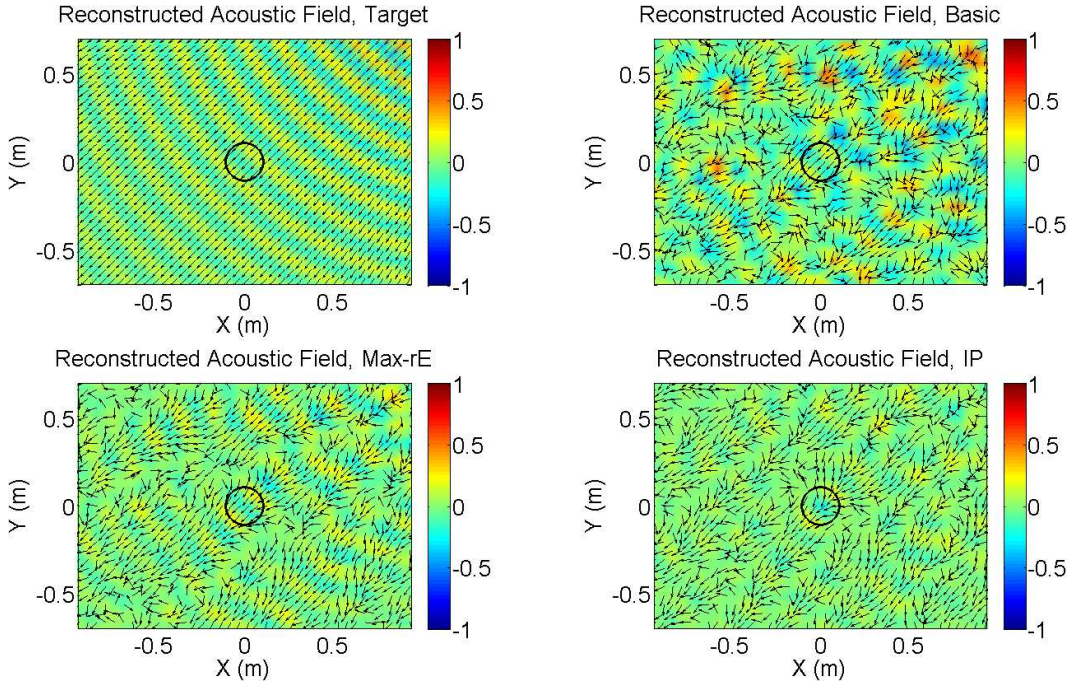


FIGURE C.6: Acoustic field reconstruction Ambisonics decoding methods, 2 kHz

An error analysis was conducted based on the acoustic pressure data using the amplitude and phase errors defined in section 4.1.1.2. Figures C.7, to C.12 show these

errors for the frequencies of 250 Hz, 1 kHz and 2 kHz, respectively.

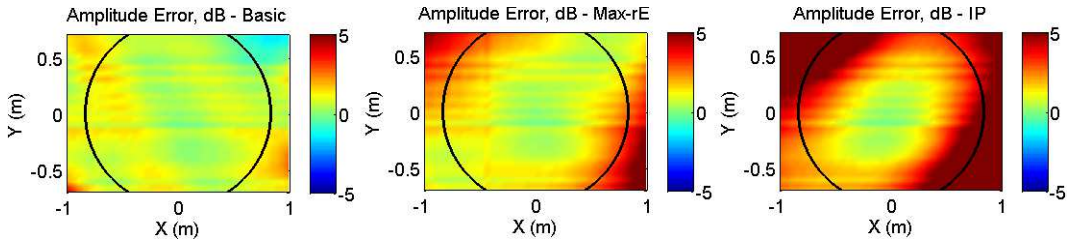


FIGURE C.7: Amplitude error, 250 Hz

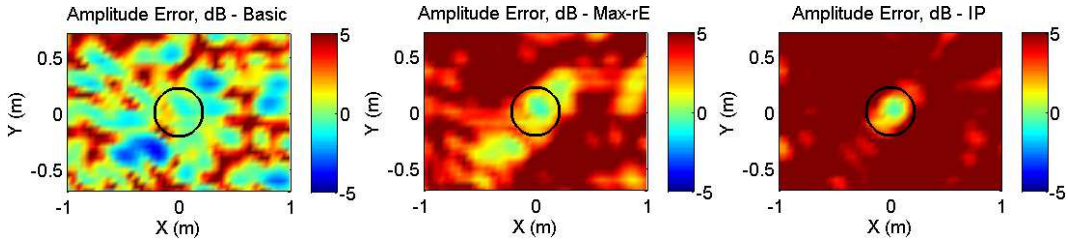


FIGURE C.8: Amplitude error, 1 kHz

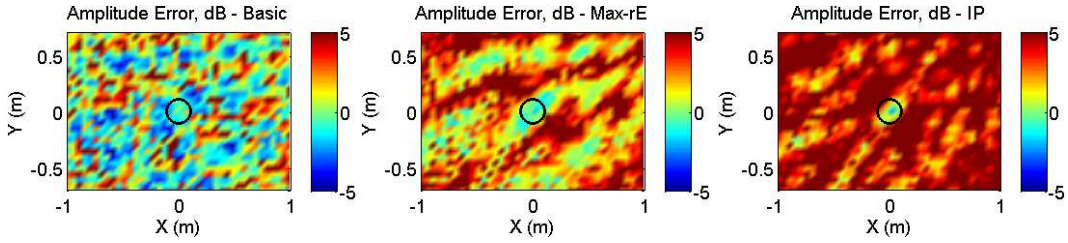


FIGURE C.9: Amplitude error, 2 kHz

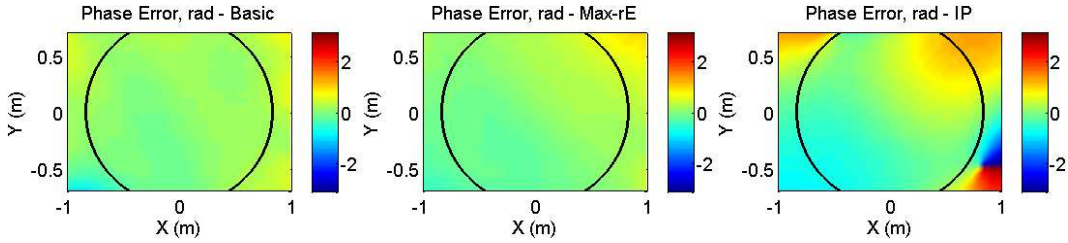


FIGURE C.10: Phase error, 250 Hz

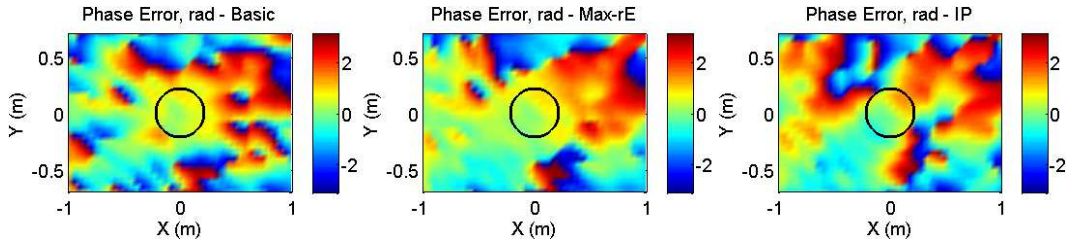


FIGURE C.11: Phase error, 1 kHz

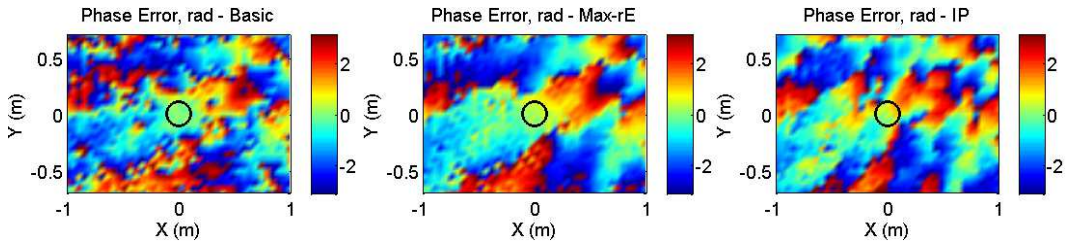


FIGURE C.12: Phase error, 2 kHz

At 250 Hz, the errors reveal that a good agreement between the target field and the synthesized field is found for the Basic decoder. For the Max-rE and In-Phase decoders, the reconstruction is accurate at the center of the listener area, but over a region with a smaller radius than the predicted by the equation (4.12). At 2 kHz, the Basic decoder does not reconstruct the sound field as is expected, even within the radius of validity. The In-Phase decoder presents a better performance compared to the Basic decoder, but the Max-rE decoder offers the best performance at this frequency. Furthermore, an improvement on the accuracy of the sound field reconstruction was found in some zones outside the radius using Max-rE at 2 kHz, which suggests that this decoder may lead to an optimization of the reconstruction at high frequencies. The reduction in the area of accuracy at low frequencies due to the implementation of this decoding method in the auralization system is not relevant because the listener radius is fully covered.

The results suggest that the implementation of the Max-rE decoder can be a suitable approach to optimize the sound field reconstruction at high frequencies in the auralization system. However, this decoder is based on the assumption of an uncorrelated summation of signals, which may not be the case for the plane wave expansion. Further investigations about the suitability of implementing different decoding methods in the auralization system are required to evaluate their effectiveness.

Bibliography

[1] M Kleiner, B Dalenbäck, and P Svensson. Auralization - An Overview. *Journal of the Audio Engineering*, 41(11):861–875, 1993.

[2] A Southern, J Wells, and D Murphy. Rendering Walk-Through Auralisations Using Wave-Based Acoustical Models. In *17th European Signal Processing Conference*, pages 715–719, Glasgow, 2009. IEE.

[3] M Vorländer. *Auralization*. Springer, Berlin, 1st edition, 2010.

[4] H Kuttruff. *Room acoustics*. Spon Press, Oxon, 5th edition, 2009.

[5] M Vorländer. Computer Simulations in Room Acoustics: Concepts and Uncertainties. *Journal of the Acoustical Society of America*, 133(3):1203–1213, 2013.

[6] B Dalenbäck, M Kleiner, and P Svensson. A Macroscopic View of Diffuse Reflection. *Journal of the Audio Engineering*, 42(10):793–807, 1994.

[7] P Svensson, R Fred, and J Vanderkooy. An Analytic Secondary Source Model of Edge Diffraction Impulse Responses. *Journal of the Acoustical Society of America*, 106(5):2331–2344, 1999.

[8] V Pulkki, T Lokki, and L Savioja. Implementation and Visualization of Edge Diffraction with Image-Source Method. In *112th Convention of Audio Engineering Society*, pages 1–13, Munich, 2002. AES.

[9] J Astley. Numerical Acoustical Modeling (Finite Element Modeling). In M Crocker, editor, *Handbook of Noise and Vibration Control*, chapter 7, pages 101–115. John Wiley & Sons, New Jersey, 1st edition, 2007.

[10] T Okuzono, T Otsurub, R Tomiku, and N Okamoto. Fundamental accuracy of time domain finite element method for sound-field analysis of rooms. *Applied Acoustics*, 71(10):940–946, 2010.

[11] M Aretz, R Nöthen, M Vorländer, and D Schröder. Combined Broadband Impulse Response Using FEM and Hibrid Ray-Based Methods. In *EAA Symposium in Auralization*, pages 1–6, Espoo, 2009. EEA.

[12] L Tafur, T Takeuchi, and K Holland. Creation of Virtual Sound Environments Using Geometrical Acoustics and Finite Elements Method. In *Reproduced Sound 2012 Auralisation*, pages 1–10, Brighton, 2012. IOA.

[13] B Støfringsdal and P Svensson. Conversion of Discretely Sampled Sound Field Data to Auralization Formats. *Journal of the Audio Engineering Society*, 54 (5):380–400, 2006.

[14] E Williams. *Fourier Acoustics*. Academic Press, London, 1st edition, 1999.

[15] M Crocker. Theory of Sound-Predictions and Measurement. In M Crocker, editor, *Handbook of Noise and Vibration Control*, chapter 2, page 41. John Wiley & Sons, New Jersey, 1st edition, 2007.

[16] F Fazi, M Noisternig, and O Warusfel. Representation of Sound Fields for Audio Recording and Reproduction. In *Acoustics 2012*, pages 1–6, Nantes, 2012.

[17] B Rafaely. Plane Wave Decomposition of the Sound Field on a Sphere by Spherical Convolution. Technical report, Institute of Sound and Vibration Reserach, ISVR, Southampton, 2003.

[18] N Xiang and J Blauert. Binaural Scale Modelling for Auralisation and Prediction of Acoustics in Auditoria. *Applied Acoustics*, 38(2-4):267–290, 1993.

[19] M Schroeder and B Atal. Computer Simulation of Sound Transmission in Rooms. In *Proceedings of the IEEE*, pages 536–537, New York, 1963. IEEE.

[20] H Lehnert and J Blauert. Principles of Binaural Room Simulation. *Applied Acoustics*, 36(3-4):259–291, 1992.

[21] B Dalenbäck, M Kleiner, and P Svensson. Auralization, Virtually Everywhere. In *100th Convention of Audio Engineering Society*, pages 1–20, Copenhagen, 1996. AES.

[22] B Dalenbäck, M Kleiner, and P Svensson. Audibility of Changes in Geometric Shape Source Directivity and Absorptive Treatment-Experiments in Auralization. *Journal of the Audio Engineering*, 41(11):905–913, 1993.

[23] L Wang and M Vigeant. Evaluations of Output from Room Acoustic Computer Modeling and Auralization due to Different Sound Source Directionalities. *Applied Acoustics*, 69(12):1281–1293, 2008.

[24] S Hoare, A Southern, and D Murphy. Study of the Effect of Source Directivity on the Perception of Sound in a Virtual Free-Field. In *128th Convention of Audio Engineering Society*, pages 1–8, London, 2010. AES.

[25] International Organization for Standardization. ISO 10140-5, Acoustics- Laboratory measurement of sound insulation of building elements Part 5: Requirements for test facilities and equipment , 1998.

[26] D Martin. Directivity and the Acoustic Spectra of Brass Wind Instrument. *Journal of the Acoustical Society of America*, 13(3):309–313, 1942.

[27] R Jacques, B Albrecht, H Schade, D Vries, and F Melchior. Multichannel Source Directivity Recording in Anechoic Chamber and in a Studio. In *ForumAcusticum 2005*, pages 479–484, Budapest, 2005. European Acoustics Association.

[28] D Fernandez, T Takeuchi, S Morales, and K Holland. Measuring Musical Instruments Directivity Patterns with Scanning Techniques. In *19th International Congress of Sound and Vibration*, pages 1–8, Vilnius, 2012. IIAV.

[29] International Organization for Standardization. ISO 3741, Acoustics- Determination of Sound Power Levels and Sound Energy Levels of Noise Sources Using Sound Pressure, 2010.

[30] International Organization for Standardization. ISO 9614, Acoustics- Determination of Sound Power Levels of Noise Sources Using Sound Intensity Part 1, 2009.

[31] Catt-Acoustics. Catt-Acoustic Homepage, 2015. URL `http://www.catt.se/`.

[32] Odeon. Odeon Room Acoustics Software Homepage, 2015. URL `http://www.odeon.dk/`.

[33] AMFG. EASE Enhanced Acoustic Simulator For Engineers Homepage, 2015. URL `http://ease.afmg.eu`.

[34] V Hansen and G Munch. Making Recordings for Simulation Tests in the Archimedes Project. *Journal of the Audio Engineering Society*, 39(10):768–774, 1991.

[35] R Bader. Radiation Characteristics of Multiple and Single Sound Hole Vihuelas and a Classical Guitar. *The Journal of the Acoustical Society of America*, 131 (1):819–28, 2012.

[36] University of York. Open AIR Library, 2015. URL `http://www.openairlib.net`.

[37] T Lokki, J Pätynen, and V Pulkky. Recording of anechoic symphonic music. In *Acoustics '08 Paris*, pages 6431–6436, Paris, 2008. EAA.

[38] F Otondo and J Rindel. A New Method for the Radiation Representation of Musical Instruments in Auralizations. *Acta Acustica*, 91(5):902–906, 2005.

[39] L Savioja, T Huopaniemi, T Lokki, and R Vaananen. Creating Interactive Virtual Acoustic Environments. *Journal of the Audio Engineering Society*, 47 (9):675–705, 1999.

[40] A Pietrzyk and M Kleiner. The Application of the Finite Element Method to the prediction of Sound Fields of Small Rooms at Low Frequencies. In *102th Convention of Audio Engineering Society*, pages 1–5, Munich, 1997. AES.

[41] H Schmalle, D Noy, S Feistel, G Hauser, W Ahnert, and J Storyk. Accurate Acoustic Modeling of Small Rooms. In *131th Convention of Audio Engineering Society*, pages 1–8, New York, 2011. AES.

[42] E Granier, M Kleiner, B Dalenbäck, and P Svensson. Experimental Auralization of Car Audio Installations. *Journal of the Audio Engineering Society*, 44 (10):835–849, 1996.

[43] A Papastefanou, C Sevastiadis, G Kalliris, and G Papanikolaou. An Amphitheatric Hall Modal Analysis Using the Finite Element Method Compared to in Situ Measurements. In *120th Convention of Audio Engineering Society*, pages 1–8, Paris, 2006. AES.

[44] S Kopuz and N Lalor. Analysis of Interior Acoustic Fields Using the Finite Element Method and the Boundary Element Method. *Applied Acoustics*, 45 (3):193–210, 1995.

[45] D Herrin, T Wu, and A Seybert. Boundary Element Method. In M Crocker, editor, *Handbook of Noise and Vibration Control*, chapter 8, pages 116–127. John Wiley & Sons, New Jersey, 1 edition, 2007.

[46] A Seybert, C Cheng, and T Wu. The Solution of Coupled Interior Exterior Acoustic Problems Using the Boundary Element Method. *Journal of the Acoustical Society of America*, 88(September):1612–1618, 1990.

[47] W Haitao, Z Xiangyang, and C Ling. Calculation of Sound Fields in Small Enclosures Using a Meshless Model. *Applied Acoustics*, 74(3):459–466, 2013.

[48] J Hargreaves. *Time Domain Boundary Element Method for Room Acoustics*. Ph.d. dissertation, University of Salford, 2007.

[49] T Sakuma, S Sakamoto, and T Otsuru. *Computational Simulation in Architectural and Environmental Acoustics*. Springer, Tokio, 1st edition, 2014.

[50] D Botteldooren. Finite-Difference Time-Domain Simulation of Low-Frequency Room Acoustic Problems. *Journal of the Acoustical Society of America*, 98(6): 3302–3308, 1995.

[51] A Southern, D Murphy, and L Savioja. Spatial Encoding of Finite Difference Time Domain Acoustic Models for Auralization. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(9):2420–2432, 2012.

[52] L Savioja and V Välimäki. Reducing the Disperion Error in the Digital Waveguide Mesh Using interpolation and Frequency-Warping Techniques. *IEEE Transactions on Audio, Speech, and Language Processing*, 8(2):184–194, 2000.

[53] K Kowalczyk and M Walstijn. Room Acoustics Simulation Using 3-D Compact Explicit FDTD Schemes. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(1):34–46, 2011.

[54] J Escolano, B Pueo, S Bleda, and J Lopez. An Approach to Discrete-Time Modelling Auralization for Wave Field Synthesis Applications. In *118th Convention of Audio Engineering Society*, pages 1–9, Barcelona, 2005. AES.

[55] A Southern, S Siltanen, and L Savioja. Spatial Room Impulse Responses with a Hybrid Modelling Method. In *130th Convention of Audio Engineering Society*, pages 1–14, London, 2011. AES.

[56] International Organization for Standardization. ISO 354, Acoustics- Measurement of Sound Absorption in a Reverberation Room, 2003.

[57] E Lehmann and A Johansson. Diffuse Reverberation Model for Efficient Image-Source Simulation of Room Impulse Responses. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(6):1429–1439, 2010.

[58] D Schröder and L Tobias. Real-Time Processing of Image Sources Using Binary Space Partitioning. *Journal of the Audio Engineering Society*, 54(7):604–619, 2006.

[59] H Kuttruff. A Simple Iteration Scheme for the Computation of Decay Constants in Enclosures with Diffusely Reflecting Boundaries. *Journal of the Acoustical Society of America*, 98(1):288–293, 1995.

[60] J Allen and D Berkley. Image Method for Efficiently Simulating Small-Room Acoustics. *Journal of the Acoustical Society of America*, 65(4):943–950, 1979.

[61] D Borish. Image Method for Efficiently Simulating Small-Room Acoustics. *Journal of the Acoustical Society of America*, 75(6):1927–1836, 1984.

[62] F Mechel. Improved Mirror Source Method in Room Acoustics. *Journal of Sound and Vibration*, 256(5):873–940, 2002.

[63] M Vorländer. Simulation of the Transient and Steady-State Sound Propagation in Rooms Using a New Combined Ray-Tracing/Image-Source Algorithm. *Journal of the Acoustical Society of America*, 86(1):172–178, 1989.

[64] R Heinz. Binaural Room Simulation Based on an Image Source Model with Addition of Statistical Methods to Include the Diffuse Sound Scattering of Walls and to Predict the Reverberant Tail. *Applied Acoustics*, 38(2-4):145–159, 1993.

[65] A Kulowski. Algorithmic Representation of the Ray Tracing Technique. *Applied Acoustics*, 18(6):449–469, 1985.

[66] T Cox, B Dalenbäck, P Antonio, J Embrechts, J Jeon, and E Mommertz. A Tutorial on Scattering and Diffusion Coefficients for Room Acoustic Surfaces. *Acta Acustica united with Acustica*, 92(1):1–15, 2006.

[67] International Organization for Standardization. ISO 17497, Acoustics- Sound Scattering properties of Surfaces. Measurement of the Random-Incidence Scattering Coefficient in a Reverberation Room Part 1, 2004.

[68] H Lehnert. Systematic Errors of the Ray-Tracing Algorithm. *Applied Acoustics*, 38(2-4):207–221, 1993.

[69] Ze Xiangyang, C Keán, and S Jincai. On the Accuracy of the Ray-Tracing Algorithms Based on Various Sound Receiver Models. *Applied Acoustics*, 64 (4):433–441, 2003.

[70] A Krokstad, S Strom, and S Sorsdal. Calculating the Acoustical Room Response by the Use of a Ray Tracing Technique. *Journal of Sound and Vibration*, 8(1):118–125, 1968.

[71] A Krokstad, S Strom, and S Sorsdal. Fifteen Years ' Experience with Computerized Ray Tracing. *Applied Acoustics*, 16(4):291–312, 1983.

[72] R Bense and P Swarte. The Accurate Prediction of the Speech Transmission Index (STI) of a Sound Reinforcement System By Means of a Computer Model Using the Ray-Tracing Techniques. In *75th Convention of Audio Engineering Society*, pages 1–20, Paris, 1984. AES.

[73] T Lewers. A Combined Beam Tracing and Radiant Exchange Computer Model of Room Acoustics. *Applied Acoustics*, 38(2-4):161–178, 1993.

[74] Ramsete. RAMSETE Homepage, 2013. URL `http://www.ramsete.com/`.

[75] L Savioja and P Svensson. Overview of Geometrical Room Acoustic Modeling Techniques. *Journal of the Acoustical Society of America*, 138(2):708–730, 2015.

[76] M Gerzon. Periphony: With-Height Sound Reproduction. *Journal of the Audio Engineering Society*, 21(1):2–10, 1973.

[77] D Ward and T Abhayapala. Reproduction of a Plane-Wave Sound Field Using an Array of Loudspeakers. *IEEE Transactions on Audio, Speech, and Language Processing*, 9(6):697–707, 2001.

[78] M Poletti. Three-Dimensional Surround Sound Systems Based on Spherical Harmonics. *Journal of the Audio Engineering Society*, 53(11):1004–1025, 2005.

[79] M Gerzon. General Metatheory of Auditory Localization. In *92nd Convention of Audio Engineering Society*, pages 1–64, Viena, 1992. AES.

[80] D Jerôme, J Rault, and J Polack. Ambisonics Encoding of Other Audio Formats for Multiple Listening Conditions. In *105th Convention of Audio Engineering Society*, volume 4795, pages 1–29, San Francisco, 1998. AES.

[81] D Arteaga. An Ambisonics Decoder for Irregular 3D Loudspeaker Arrays. In *134th Convention of Audio Engineering Society*, pages 1–10, Rome, 2013. AES.

[82] S Favrot and J Buchholz. LoRA: A Loudspeaker-Based Room Auralization System. *Acta Acustica united with Acustica*, 96(2):364–375, 2010.

[83] T Falke, J Batke, and T Gorne. Remixing a Historic Film in Higher Order Ambisonics 3D Audio Workow and Technical Solutions. In *136th Convention of Audio Engineering Society*, pages 1–4, Berlin, 2014. AES.

[84] F Martellotta. On the Use of Microphone Arrays to Visualize Spatial Sound Field Information. *Applied Acoustics*, 74(8):987–1000, 2013.

[85] M Noisternig, A Sontacchi, T Musil, and R Höldrich. A 3D Ambisonic Based Binaural Sound Reproduction System. In *AES 24th International Conference on Multichannel Audio*, pages 1–5, Banff, 2003. AES.

[86] R Nishimura and K Sonoda. B-Format for Binaural Listening of High Order Ambisonics. In *21st International Congress on Acoustics*, pages 1–9, Montreal, 2013. ASA.

[87] G. Enzner, M Weinert, S Abeling, J Batke, and P Jax. Advanced System Options for Binaural Rendering of Ambisonic Format. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 251–255, Vancouver, 2013. IEEE.

[88] H Møller. Fundamentals of Binaural Technology. *Applied Acoustics*, 36 (December 1991):171–218, 1992.

[89] H Mø ller, M Sorensen, C Jensen, and D Hammershoi. Binaural Technique : Do We Need Individual Recordings ? *Journal of the Audio Engineering Society*, 44(6):451–469, 1996.

[90] E Langendijk and A Bronkhorst. Fidelity of Three-Dimensional-Sound Reproduction Using a Virtual Auditory Display. *Journal of the Acoustical Society of America*, 107(May 1999):528–537, 2000.

[91] J Breebaart, F Nater, and A Kohlrausch. Spectral and Spatial Parameter Resolution Requirements for Parametric , Filter-Bank-Based HRTF Processing. *Journal of the Audio Engineering Society*, 58(3):126–140, 2010.

[92] K Hartung, J Braasch, and S Sterbing. Comparison of Different Methods for the Interpolation of Head-Related Transfer Functions. In *16th International Conference: Spatial Sound Reproduction of Audio Engineering Society*, pages 319–329, Rovaniemi, 1999. AES.

[93] T Ajdler, C Faller, L Sbaiz, and Vetterli. M. Interpolation of Head Related Transfer Functions Considering Acoustics. In *118th Convention of Audio Engineering Society*, pages 1–7, Barcelona, 2005. AES.

[94] X Zhong, F Zhang, and B Xie. On the Spatial Symmetry of Head-Related Transfer Functions. *Applied Acoustics*, 74(6):856–864, 2013.

[95] Z Schärer and A Lindau. Evaluation of Equalization Methods for Binaural Signals. In *126th Convention of Audio Engineering Society*, pages 1–17, Munich, 2009. AES.

[96] A Lindau and F Brinkmann. Perceptual Evaluation of Headphone Compensation in Binaural Synthesis Based on non-individual Recordings. *Journal of the Audio Engineering Society*, 60(1):54–62, 2012.

[97] W Martens. Individualized and Generalized Earphone Correction Filters for Spatial Sound Repeduction. In *2003 International Conference on Auditory Display*, pages 263–266, Boston, 2003.

[98] C Sander, F Wefers, and D Leckschat. Scalable Binaural Systhesis on Mobile Devices. In *133rd Convention of Audio Engineering Society*, pages 1–8, San Francisco, 2012. AES.

[99] Staff Technical Writter. Binaural Technology for Mobile Applications. *Journal of the Audio Engineering Society*, 54(10):990–995, 2006.

[100] A Moore and A Tew. An Initial Validation of Individualized Crosstalk Cancellation Filters for Binaural Perceptual Experiments. *Journal of the Audio Engineering Society*, 58(1/2):36–45, 2010.

[101] T Takeuchi and P Nelson. Optimal Source Distribution for Binaural Synthesis over Loudspeakers. *Journal of the Acoustical Society of America*, 112(6):2786–2797, 2002.

[102] Christos Tsakostas and Andreas Floros. Optimized Binaural Modeling for Inmersive Audio Applications. In *122nd Convention of Audio Engineering Society*, pages 1–7, Viena, 2007. AES.

[103] T Lentz. Dynamic Crosstalk Cancellation for Binaural Synthesis in Virtual Reality Environments. *Journal of the Audio Engineering Society*, 54(4):283–294, 2006.

[104] H Lehnert and J Blauert. Virtual Auditory Environment. In *5th International Conference Advanced Robotics*, pages 211–216, Pisa, 1991. IEEE.

[105] D Furlong, M Doyle, E Kelly, C MacCabe, and R MacLaverry. Interactive Virtual Acoustics Synthesis for Architectural Acoustics Design. In *93rd Convention of Audio Engineering Society*, pages 1–10, San Francisco, 1992. AES.

[106] J Jot. Real-Time Spatial Processing of Sounds for Music, Multimedia and Interactive Human-Computer Interfaces. *Multimedia Systems*, 7(1):55–69, 1999.

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

[108] J Sandvad. Dynamic Aspects of Auditory Virtual Environments. In *100th Convention of Audio Engineering Society*, volume 4226, pages 1–14, Copenhagen, 1996. AES.

[109] M Noisternig, B Katz, S Siltanen, and L Savioja. Framework for Real-Time Auralization in Architectural Acoustics. *Acta Acustica united with Acustica*, 94(6):1000–1015, 2008.

[110] A Chandak, C Lauterbach, M Taylor, Z Ren, and D Manocha. AD-Frustum: Adaptive Frustum Tracing for Interactive Sound Propagation. *IEEE transactions on Visualization and Computer Graphics*, 14(6):1707–1714, 2008.

[111] M Taylor. RESound : Interactive Sound Rendering for Dynamic. In *17th International ACM Conference on Multimedia 2009*, pages 271–280, Beijing, 2009. ACM.

[112] M Crocker. Fundamentals of Acoustics, Noise and Vibration. In M Crocker, editor, *Handbook of Noise and Vibration Control*, chapter 2, page 60. John Wiley & Sons, New Jersey, 1st edition, 2007.

[113] A Reilly and D McGrath. Real-Time Auralisation with Head Tracking. In *5th Australian Regional Convention of Audio Engineering Society*, pages 1–5, Sydney, 1995. AES.

[114] F Freeland, L Wagner, and P Diniz. Efficient HRTF Interpolation in 3D Moving Sound. In *22nd International Conference: Virtual, Synthetic, and Entertainment Audio of Audio Engineering Society*, pages 1–9, Espoo, 2002. AES.

[115] M Matsumoto, S Yamanaka, M Tohyama, and H Nomura. Effect of Arrival Time Correction on the Accuracy of Binaural Impulse Response Interpolation Interpolation Methods of Binaural Response. *Journal of the Audio Engineering Society*, 52(1):56–61, 2004.

[116] M Matsumoto and M Tohyama. Algorithms for Moving Sound Images. In *114th Convention of Audio Engineering Society*, pages 1–6, Amsterdam, 2003. AES.

[117] M Queiroz and G Montesiao. Efficient Binaural Rendering of Moving Sound Sources Using HRTF Interpolation Efficient Binaural Rendering of Moving Sound Sources Using HRTF Interpolation. *Journal of New Music Research*, 40(3):37–41, 2011.

[118] V Pulkki. Virtual sound source positioning using vector base amplitude panning. *Journal of the Audio Engineering Society*, 45(6):456–466, 1997.

[119] N Hahn and S Spors. Physical Properties of Modal Beamforming in the Context of Data-based Sound Reproduction. In *139th Convention of Audio Engineering Society*, pages 1–13, New York, 2015. AES.

[120] F Schultz and S Spors. Data-based Binaural Synthesis Including Rotational and Translatory Head-Movements. In *52nd International Conference of Audio Engineering Society*, pages 1–11, Guildford, 2013. AES.

[121] A Walther and C Faller. Linear Simulation of Spaced Microphone Arrays Using B-Format Recordings. In *128th Convention of Audio Engineering Society*, pages 1–7, London, 2010. AES.

[122] International Organization for Standardization. ISO 3382 part 2, Acoustics-Measurement of Room Acoustic Parameters, Reverberation Time in Ordinary Rooms, 2009.

[123] A Pierce. *Acoustics - An introduction to Its Physical Principles and Applications* . McGraw-Hill, USA, 1st edition, 1981.

[124] International Organization for Standardization. ISO 3382 part1, Acoustics-Measurement of Room Acoustic Parameters, Performance spaces, 2009.

[125] I Bork. Report on the 3rd Round Robin on Room Acoustical Computer Simulation Part II : Calculations. *Acta Acustica united with Acustica*, 91(4):753–763, 2005.

[126] S Siltanen, T Lokki, L Savioja, and C Christensen. Geometry Reduction in Room Acoustics Modeling. *Acta Acustica united with Acustica*, 94(3):410–418, 2008.

[127] A Foteinou, D Murphy, and A Masinton. Investigation of Factors Influencing Acoustic Characteristics in Geometric Acoustics Based Auralization. In *13th Int. Conference on Digital Audio Effects*, pages 8–11, Graz, 2010. IEM.

[128] M Lisa, J Rindel, and C Christensen. Predicting the Acoustics of Ancient Open-Air Theatres: the Importance of Calculation Methods and Geometrical Details. In *Baltic-Nordic Acoustics Meeting*, pages 1–8, Mariehamn, 2004.

[129] K Kuttruff. Auralization of Impulse Responses Modeled on the Basis of Ray-Tracing Results. *Journal of the Audio Engineering Society*, 41(11):876–880, 1993.

[130] A Gade, M Lisa, C Lynge, J Rindel, and D Lyngby. Roman Theatre Acoustics; Comparison of Acoustic Measurement and Simulation Results from the Aspendos Theatre, Turkey. In *18th International Congress of Acoustics*, pages 1–4, Kyoto, 2004.

[131] M Aretz. *Combine Wave And Ray Based Room Acoustic Simulations of Small Rooms*. Ph.d. dissertation, RWTH Aachen University, 2012.

[132] Catt-Acoustics. *Catt-Acoustics V9.0 User's Manual*. Catt-Acoustics, Berlin, 2011.

[133] H Kuttruff. Sound Fields in Fitted Rooms. In *15th International Conference of Audio Engineering Society*, pages 1–8, Copenhagen, 1998. AES.

[134] International Organization for Standardization. ISO 10534-2, Acoustics- Determination of Sound Absorption Coefficient and Impedance in Impedance Tubes, 2001.

[135] Y Takahashi, T Otsuru, and R Tomiku. In Situ Measurements of Surface Impedance and Absorption Coefficients of Porous Materials using Two Microphones and Ambient Noise. *Applied Acoustics*, 66(7):845–865, 2005.

[136] E Tijs and H Bree. Recent Developments Free Field PU Impedance Technique. In *Symposium on the Acoustics of Poro-Elastic Materials-2008*, pages 1–2, Bradford, 2008.

[137] F Fahy. *Foundations of Engineering Acoustics*. Academic Press, Oxford, 1st edition, 2008.

[138] J Suh and P Nelson. Measurement of Transient Response of Rooms and Comparison with Geometrical Acoustic Models. *Journal of the Acoustical Society of America*, 105(4):2034–2317, 1999.

[139] P Nelson and S Yoon. Estimation of Acoustic Source Strength By Inverse Methods: Part I, Conditioning of the Inverse Problem. *Journal of Sound and Vibration*, 233(4):639–664, 2000.

[140] J Fliege. Sampled Sphere.pdf. Technical report, University of Dortmund, Dortmund, 1999.

[141] H Herlufsen, S Gade, and H Zaveri. Analyzers and Signal Generators. In M Crocker, editor, *Handbook of Noise and Vibration Control*, chapter 40, pages 101–115. John Wiley & Sons, New Jersey, 1st edition, 2007.

[142] R Kennedy, P Sadeghi, T Abhayapala, and H Jones. Intrinsic Limits of Dimensionality and Richness in Random Multipath Fields. *IEEE Transactions on Signal Processing*, 55(6):2542–2556, 2007.

[143] Y Kim and P Nelson. Optimal Regularisation for Acoustic Source Reconstruction by Inverse Methods. *Journal of Sound and Vibration*, 275(3-5):463–487, 2004.

[144] S Yoon and P Nelson. Estimation of Acoustic Source Strength By Inverse Methods: Part II, Experimental Investigation of Methods for Choosing Regularization Parameters. *Journal of Sound and Vibration*, 233(4):665–701, 2000.

[145] B Dalenbäck. Syntes av virtuell direktiv mikrofon, 01 2012.

[146] Technology Arts Sciences TH Koln. Spherical Far Field HRIR Compilation of the Neumann KU100, 2016. URL `http://audiogroup.web.th-koeln.de/ku100hrir.html`.

[147] B Masiero and J Fels. Perceptually Robust Headphone Equalization for Binaural Reproduction. In *130th Convention of Audio Engineering Society*, pages 1–7, London, 2011. AES.

[148] M Poletti. Unified description of Ambisonics using real and complex spherical harmonics. In *Ambisonics Symposium 2009*, pages 1–10, Graz, 2009. IAEM.

[149] Department of Music. Virginia Tech-School of Performing Arts , 2015. URL `http://disis.music.vt.edu/main/index.php`.

[150] F Fazi and D Menzies. Estimation of the Stability of a Virtual Sound Source Using a Microphone Array. In *22nd International Congress of Sound and Vibration*, pages 1–8, Florence, 2015. IIAV.

[151] D Menzies and F Fazi. A Theoretical Analysis of Sound Localisation, with Application to Amplitude Panning. In *138th Convention of Audio Engineering Society*, pages 1–5, Warsaw, 2015. AES.

[152] M Gerzon. Psychoacoustic Decoders for Multispeaker Stereo and Surround Sound. In *93rd Convention of Audio Engineering Society*, pages 1–47, San Francisco, 1992. AES.