FROM SOURCE TO BRAIN: MODELLING SOUND PROPAGATION AND LOCALISATION IN ROOMS

Jonathan Sheaffer

School of Computing, Science and Engineering
University of Salford, Salford UK

Submitted in Partial Fulfilment of the Requirements of the Degree
of Doctor of Philosophy, March 2013

Contents

List of Figures

List of Tables

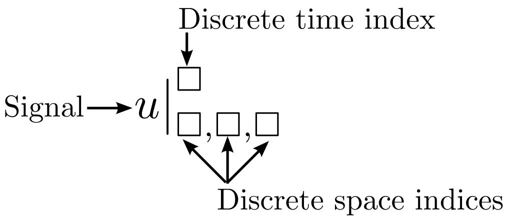
Nomenclature

$\dot{\alpha}$	Absorption coefficient
$\delta(\cdot)$	Dirac function
$\gamma_{\dot{k}}$	Normalised cross-correlation function at band $\dot{k}$
$\nabla\cdot$	Divergence operator
∇	Gradient operator
∇^2	Laplace operator
$\hat{\alpha}_{\dot{k}}$	Instantaneous ILD at band $\dot{k}$
$\hat{\tau}_{\dot{k}}$	Instantaneous ITD at band $\dot{k}$
$\hat{c}_{\dot{k}}$	Instantaneous interaural coherence at band $\dot{k}$
ξ_w	Specific acoustic impedance (normalised)
α	Damping factor
$*$	Convolution operator
β	Bilinear operator
λ	Courant number
μ	Slope of IC threshold curve
ν	Surface velocity
ω	Angular Frequency
ω_0	Undamped angular resonance frequency
Ψ	Model response azimuth
ψ	Arbitrary source function
ρ_0	Ambient density of air
Θ	Model confidence
θ	Azimuth of incident wave

$A(\dot{k}, \theta')$	Cue probability pattern (known azimuth)
$A(\dot{k}, \theta)$	Cue probability pattern (target azimuth)
a_h	Radius of head or sphere approximating the head
F	Driving force
j	Imaginary unit, $\sqrt{-1}$
K	Elasticity constant
M	Mass constant
Q	Quality factor
q	Rate of fluid emergence
R	Damping constant
s	Laplace variable
z	Complex frequency variable
c	Velocity of sound in air
CPP	Cue Probability Pattern
CUDA	Compute Unified Device Architecture
ERB	Equivalent Rectangular Bandwidth
f	Frequency
GTFB	Gammatone Filter Bank
IC	Inferior Colliculus
IC	Interaural Coherence
ICI	Inter-Click Interval
ILD	Interaural Level Difference
ITD	Interaural Time Difference
IWB	Interpolated Wideband
JND	Just Noticeable Difference
k	Acoustic wavenumber
LRS	Locally Reacting Surface
LSO	Lateral Superior Olive
MAA	Minimum Audible Angle
MSO	Medial Superior Olive
PCS	Physically Constrained Source
PDF	Probability Density Function

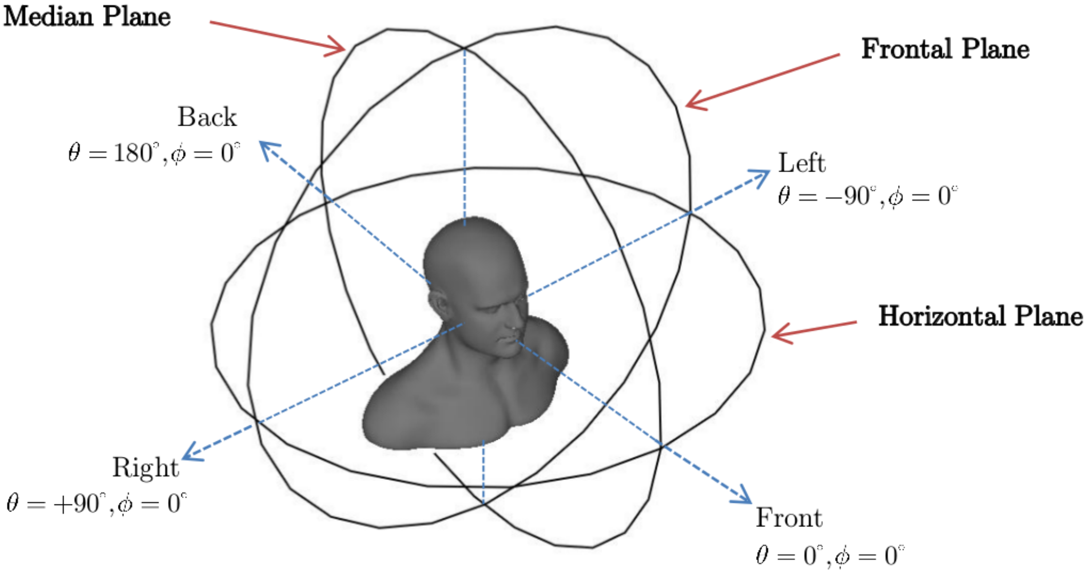
r	Radial distance
SOC	Superior Olivary Complex
SRL	Standard Rectilinear
c_0	Interaural coherence threshold
f_s	Sample rate
$\dot{k}$	Index number of auditory filter
$\hat{\mathbf{n}}$	Normal component of vector
$p(\mathbf{x}, t)$	Acoustic pressure
$P_{\dot{k}}(\hat{\alpha})$	Probability density function for ILD
$P_{\dot{k}}(\hat{\tau})$	Probability density function for ITD
$\hat{q}$	Volume velocity
$\hat{r}$	Reflection coefficient
s_g	Grid signal (processed source function)
s_p	Excitation signal
T	Temporal period
$\mathbf{u}(\mathbf{x}, t)$	Acoustic particle velocity
X	Spatial period
$x^l_{\dot{k}}$	Output of left ear peripheral processor at band $\dot{k}$
$x^r_{\dot{k}}$	Output of right ear peripheral processor at band $\dot{k}$
Y	Specific acoustic admittance
z_0	Characteristic impedance of air
z_w	Surface impedance material

Time domain functions are written in lower case letters, e.g. $p(x, t)$, whereas frequency domain functions are written in upper case letters, e.g. $P(x, \omega)$. Discrete signals are notated in multidimensional form. For example, an arbitrary signal u which depends on both time and space, is notated as follows:



In the discrete frequency domain, normalised frequencies are given over the range $0 \leq f \leq 0.5$, where 0 is DC and 0.5 is the Nyquist frequency.

When working in a spherical coordinate system, azimuth is denoted θ, elevation is denoted ϕ and radial distance is denoted r, as follows:



Acknowledgements

There are many people who helped and inspired this work, both academically and personally. First and foremost, I would like to thank my supervisor, Bruno Fazenda, for his ongoing availability, guidance and advice, especially given the challenging geographic circumstances behind this endeavour. Bruno is one of the most open-minded academics I know, and has always been supportive of my passion to explore new avenues. His ability to motivate me, see the big picture and steer me in the right direction is very much reflected in this work.

I have had the pleasure to collaborate with some of the brightest minds in the field of finite difference modelling. My long discussions with Maarten van Walstijn have resulted in a fruitful collaboration which is evident in a thesis chapter on source modelling. I would like to thank Maarten for a fantastic lesson on academic rigour, and for always being patient and forthcoming. At earlier stages of this work, Damian Murphy provided important insights on practical aspects of finite difference modelling, which illuminated some fundamental issues and resulted in a conference paper. I would also like to thank my co-supervisor, Jamie Angus, for some interesting discussions on GPU processing, which were very helpful in realising the WaveCloud project.

Throughout the evolution of this dissertation, I received some useful advice from the thesis committee and I would like to thank them for taking the time to review this work at different stages. Particularly, I would like to extend my gratitude to the thesis examiners, Michael Vorländer and Trevor Cox, for an interesting discussion and for their useful advice. Other people who provided assistance and/or inspiration are, in no particular order, Konrad Kowalczyk, Mark

Avis, Jeroen Breebaart, Alex Southern, Eldad Klaiman, Yuvi Kahana, Floyd Toole, Nir London, John O'Hare, Hanoch Lavee, Matan Zehavi, Assaf Tal, Ohad Dekel and my colleagues at Salford's acoustics research centre. I apologise if I forgot anyone.

Last but not least, I would like to thank my family without whom this journey could not have been possible: to my children, Omer and Daphne, for occasionally steering my attention away from acoustics, and to my parents, Zachary and Arza, for their never-ending support. Over the past 3.5 years, my beloved spouse, Roni, has supported me in every way possible and has always been there for me, no matter the cost. For that, I could never be grateful enough.

Abstract

Human localisation of sound in enclosed spaces is a cross-disciplinary research topic, with important applications in auditory science, room acoustics, spatial audio and telecommunications. By combining an accelerated model of 3D sound propagation in rooms with a perceptual model of spatial processing, this thesis provides an integrated framework for studying sound localisation in enclosed spaces on the horizontal plane, with particular emphasis on room acoustics applications. The room model is based on the finite difference time domain (FDTD) method, which has been extended to include physically-constrained sources and binaural receivers based on laser-scanned listener geometries. The underlying algorithms have been optimised to run on parallel graphics hardware, thus allowing for a high spatial resolution, and accordingly, a significant decrease of numerical dispersion evident in the FDTD method. The perceptual stage of the model features a signal processing chain emulating the physiology of the auditory periphery, binaural cue selection based on interaural coherence, and a final decision maker based on supervised learning. The entire model is shown to be capable of imitating human sound localisation in different listening situations, including free field conditions and at the presence of sound occlusion, diffraction and reflection. Results are validated against subjective data found in the literature, and the model's applications to the fields of room acoustics and spatial audio are demonstrated and discussed.

—If there's a book you really want to read, but it hasn't been written yet, then you must write it.

Toni Morrison

1

Introduction

Whether we are aware of it or not, models have been an integral part of our lives since ancient times. We build physical models to understand the operation of mechanical systems. We look up to people who we view as role models, and draw from their behaviour. We use fashion models as human manikins to advertise and promote commercial products. More recently, with the availability of computing technology we are also able to generate virtual models of complex physical phenomena.

Consider, for example, a person enjoying a peaceful evening at home, when a sudden sound of shattering glass catches his or her attention. The mechanical motion of the breaking glass causes air particles to vibrate in a certain order and an acoustic disturbance is generated. When this disturbance propagates, it gets reflected off the different surfaces in the room, causing a composite sound field to be built. At the listener's ears, this vast amount of information is transformed into neural firings which get decoded by the listener's brain. Within a fraction of a second, the listener processes this information and is able to tell, with remarkable accuracy, where the sound is coming from. We take this process for granted. It is an evolutionary skill essential to our survival. Yet if we wanted to model this phenomena, how would we go about it?

1.1 Background

To cover the entire process of sound localisation in rooms, one would not only need to model the perception of spatial sound, but also the effects of the sound source interacting with the room itself. It is therefore acknowledged that the room itself may play an important role in the way we localise sound. Human sound localisation is well understood in idealised environments (Gaskell, 1983; Hartmann, 1983; Rakerd and Hartmann, 1985, 1986; Hartmann and Rakerd, 1989; Blauert, 1997; Litovsky et al., 1999; Goupell et al., 2012), and such listening situations are straightforward to measure or model. Accordingly, many sound propagation models employed in auditory research are fairly simplified, as they are only required to account for basic acoustic phenomena, such as superposition between sources, or reflections in a rectilinear enclosure. However, in the fields of room acoustics and spatial audio, the effects of more advanced acoustic parameters on sound localisation (e.g. room materials, geometry, and source directivity) present interesting research questions. Thus, the ability to model sound localisation whilst being in control of parameters in both the acoustic and psychoacoustic domains, would be beneficial to scientists studying room acoustics, as well as to engineers who wish to evaluate the effectiveness of their designs.

Studying room acoustics using auditory models is an emerging area of research (Blauert et al., 2009; Blauert and Braasch, 2011; Blauert, 2013; van Dorp Schuitman et al., 2013), yet it appears that there is no single model that follows the entire process, from source to brain. In this thesis, the problem of sound localisation in rooms is treated as five individual processes which are integrated into a single model, as schematically shown in Figure 1.1.

Macroscopically, the entire model can be seen as an integration of a room model and a perceptual model of sound localisation, operating in the acoustic

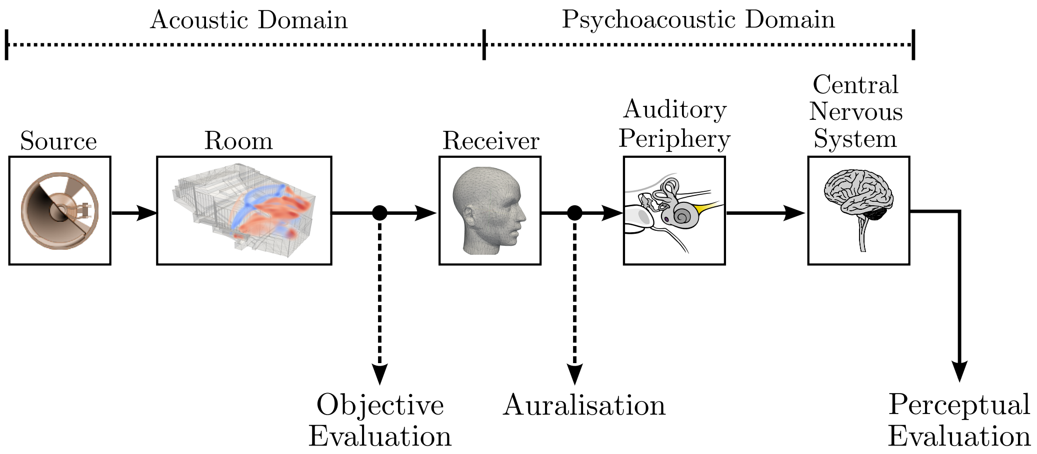


Figure 1.1: Schematic representation of a complete sound localisation model and its applications.

domain and psychoacoustic domain, respectively. The room model accounts for sound generation (source model) and propagation through the room, whereas the perceptual model accounts for processing of the information by the peripheral auditory system and the central nervous system. The two domains are linked by the receiver model, whose transfer function converts the acoustic pressure field into a binaural response. The output of the entire integrated model can be used to evaluate the perceptual effects of changing parameters in both domains, such as source directivity, room geometry, boundary conditions, head size and shape, and the dynamics of the peripheral auditory system. Additionally, the different components of the model can be used individually. The output of the room model can be used to objectively evaluate the acoustics, either by visually inspecting the room's transfer function, or by extrapolating objective room parameters such as T_{60}, EDT, C_{80} and STI (Kuttruff, 2000), to name a few. Otherwise, the outputs of the receiver model can be used to create an audible rendering of the sound source within the acoustic space, a concept known as auralisation (Kleiner et al., 1993; Vorländer, 2008).

Processes in the acoustic domain, namely the source, room and receiver, are modelled using the *Finite Difference Time Domain* (FDTD) method, which has recently become more suitable for room acoustics simulation (Murphy and Beeson, 2007; Kowalczyk and van Walstijn, 2011). In room acoustics, the FDTD method is considered to be less established than geometrical methods (Krokstad et al., 1968; Allen and Berkley, 1979; Vorländer, 1989) or other wave methods (Ciskowski and Brebbia, 1991; Ihlenburg, 1998), and as such, one may wonder why it has been chosen in context of this project. There are a few arguments supporting this choice:

1. Many acoustic phenomena involving wave behaviour such as superposition and diffraction are more difficult to handle in geometrical methods in comparison to FDTD. These features are especially important when studying problems involving small rooms, and sound reproduction paradigms.

2. Being relatively new in its applications to room acoustics, the FDTD method presents interesting avenues of research. To the author's knowledge, FDTD has never been used in such a wide context, and accordingly, it presents an exciting challenge.

3. The FDTD method is straightforward to implement. As such, a customised computer program can be generated rather than relying on third party solutions. This allows the author to have full control over the entire modelling process.

4. The FDTD method can be used to model time-variant systems, such as walkthrough auralisation (Southern et al., 2009), moving sound sources and even moving media (Ostashev et al., 2005). Although these features are not directly studied in this thesis, they exemplify potential future applications of the integrated model.

Using the FDTD method, it is now possible to model rooms with efficient numerical schemes (Savioja and Valimaki, 2000; Kowalczyk and van Walstijn, 2011; Bilbao, 2012) and to account for frequency dependent boundaries (Kowalczyk and van Walstijn, 2008, 2011), diffusing boundaries (Kowalczyk et al., 2011) and directional sound sources (Escolano et al., 2007). Yet, there are still some unresolved issues which need to be addressed when employing the FDTD method in such a broad context:

1. **Sources**. Existing methods for exciting FDTD grids have fundamental problems, such as source scattering, numerical inconsistency and undesired

frequency response effects, with only partial solutions found in the literature (Schneider et al., 1998a).

2. **High Resolution Modelling**. Similar to other wave methods, the computational cost of the FDTD method is high relative to geometrical methods, and increases dramatically with frequency. Thus, in order to accurately model the entire audible spectrum even in a small room, one needs extensive computing power.

3. **Binaural Receivers**. In order to interface between the acoustic domain (source and room) and psychoacoustic domain (ear and brain) or for auralisation purposes, one needs to obtain binaural room responses. Insofar, the issue of directly rendering binaural room responses using FDTD is only partially addressed in the literature (Murphy and Beeson, 2007)

In the psychoacoustic domain, there are various approaches for modelling the auditory periphery and prominent mechanisms in the central nervous system which govern sound localisation. Models of the peripheral auditory system (the middle-ear, inner-ear and associated neural transduction mechanisms) are well established in the field of hearing science (Giguère and Woodland, 1994; Meddis and Lopez-Poveda, 2010), and can be used in a pre-processing stage of sound localisation models normally involving binaural and/or monaural processing. Some of these models attempt to simulate neural processing mechanisms in the central nervous system, e.g. (Goodman and Brette, 2010), some are biologically inspired (Jeffress, 1948; Lindemann, 1986a; Breebaart et al., 2001), whilst others are analytically motivated (Macpherson, 1991; Faller and Merimaa, 2004b). It is important to point out that these models aim to imitate the way humans localise sound, unlike more general signal processing models whose goal is to perform localisation as accurately as possible, e.g. (Valin et al., 2007; MacDonald, 2008).

It appears that models which are *physiologically plausible* are designed to focus on the specifics of the auditory system, rather than on the general process of sound localisation. Conversely, models which are *psychoacoustically plausible* do not attempt to correctly model biological processes, but to yield results which correlate well with those obtained by means of subjective testing. Accordingly, the main challenge in the psychoacoustic domain is to identify, improve and generalise a binaural processor capable of handling acoustic settings which involve multiple sound sources and/or sound reflections. Whilst exact emulation of the auditory system is not necessary, such a model should at least yield psychoacoustically plausible results.

1.2 Objectives and Scope

Drawing from these challenges, the aim of the thesis is to develop and validate a generalised model for sound localisation in rooms, which follows the entire process of sound generation, propagation and auditory perception. The objectives of the thesis are:

1. To contribute to room acoustics modelling by improving the FDTD method and enhancing its usability for simulation of realistic listening situations in rooms.
2. To contribute to psychoacoustics modelling by developing a complete and generalised model of sound localisation in enclosed spaces.
3. To exemplify how an integrated modelling framework, as shown in Figure 1.1, can be applied to problems in room acoustics and spatial audio.

Within the scope of this work, modelling in the psychoacoustic domain involves only binaural processing, meaning that perceptual investigations are limited to localisation on the horizontal plane. Thus, the underlying assumption is that sound sources are only presented on the same plane as the receiver. It is emphasised that both azimuthal and elevation reflections may contribute to binaural localisation, and accordingly, modelling in the acoustic domain is performed in a full 3D space.

Considering the above objectives and the defined scope of work, the modelling framework suggested in this thesis would be particularly suited for studying the psychoacoustics of sound reproduction in rooms. It might also be used for the psychoacoustic evaluation of architectural acoustics design, and with some adjustments, as more generalised room acoustics simulation software.

1.3 Thesis Structure

The thesis is structured as follows:

Chapter 2 provides background information on the physics and perception of spatial sound, with focus on sound localisation in enclosed spaces. The fundamental laws of acoustics are briefly reviewed, followed by an introduction to the physiology of directional hearing. These topics provide the foundations to the acoustic and psychoacoustic models developed in the thesis. Lastly, the psychophysical basis of sound localisation is explained, with emphasis given to interaural cues and the perception of directional sound in rooms.

Chapter 3 provides an overview of room acoustics modelling techniques, with emphasis on the FDTD method which is widely used in this thesis. Additionally, the chapter provides a literature review of sound localisation methods and their suitability to the goals of this thesis (or lack thereof).

It is noted that Chapters 2 and 3 summarise relevant information that can be found in the literature, whereas subsequent chapters constitute the author's original contributions.

Chapter 4 deals with modelling sound sources for excitation of FDTD grids. With numerical methods in mind, the fundamental physics of sound generation and propagation are revisited, and the anatomy of sources in the FDTD method is discussed. The merits and shortcomings of existing excitation methods is reviewed, and a generalised approach for modelling sources in FDTD is offered. Finally, a novel method of exciting FDTD grids is described and compared to existing methods.

Chapter 5 outlines the WaveCloud project, which is the computational engine used to obtain room impulse responses in this thesis. The chapter includes a combination of smaller contributions which together complements the practical applications of the FDTD method. Focus is given to an accelerated implementation of FDTD on cost-efficient general purpose graphics hardware, to modelling binaural receivers and to a method for modelling frequency dependent sources and boundaries.

Chapter 6 introduces an integrated method for sound localisation in rooms. It includes modelling components for the peripheral auditory system and binaural processing which are predicated on the work of other authors, as well as components for central processing and decision making which are novel contributions. The chapter portrays the anatomy of the sound localisation model and discusses its plausibility. Lastly, Chapter 7 validates the model by comparing simulated results to subjective data found in the literature. Additionally, the chapter exemplifies how the model can be applied in a range of listening situations involving multiple sources, sound reflections, occluded sound sources and a reverberant field.

1.4 Contributions

The main contributions of the thesis can be divided into three categories, namely FDTD source modelling (covered in Chapter 4), practical aspects of FDTD modelling (covered in Chapter 5) and modelling sound localisation in rooms (covered in Chapters 6 and 7).

FDTD Source modelling:

- Physical and numerical constraints in context of FDTD source modelling have been clearly identified. These include scaling, differentiation and superposition of excitation functions.
- Existing source models have been systematically reviewed in light of this new knowledge.
- A novel cascaded-filters approach to generalise all existing source models has been proposed and evaluated.
- A novel source model, termed the *Physically Constrained Source*, was introduced. This source model is shown to be the only model that is numerically

consistent, and by definition, converges with a closed form solution to the wave-equation.

Practical aspects of room acoustics modelling using FDTD:

- A generalised FDTD framework for room acoustics simulation is introduced.
- An accelerated FDTD algorithm employing general purpose graphics processors is implemented and benchmarked. This algorithm allows for realistic rooms to be modelled entirely using the FDTD method.
- An approach to render binaural impulse responses in FDTD methods is suggested, by directly embedding laser scans of human subjects in the grid. The approach is evaluated and is shown to correctly reproduce interaural cues.
- A multi-band method to solve frequency dependent boundaries and sources in FDTD is introduced and evaluated.

Modelling sound localisation in rooms:

- The binaural processor suggested in (Faller and Merimaa, 2004b) is extended with a central processor and a decision maker. This constitutes a complete sound localisation model, which yields a single localisation judgement, similar to a listener in a forced-choice listening test.
- The concept of a frequency-dependent interaural coherence threshold is introduced. This contributes to the generality of the sound localisation model, in the sense that a single frequency dependent threshold can be used to evaluate a wide range of listening situations.
- An integrated model of sound localisation in the horizontal plane is developed. This the first model which simulates the entire process of sound generation, propagation and localisation, in a physically accurate and psychoacoustically plausible manner.

1.5 Related Publications

Some of the work presented in this thesis has been published in the following conference proceedings (* denotes a fully peer-reviewed publication):

- J. Sheaffer and B.M. Fazenda. FDTD/K-DWM simulation of 3D room acoustics on general purpose graphics hardware using compute unified device architecture (CUDA). In *Proceedings of the Institute of Acoustics*, volume 32. Institute of Acoustics, 2010
- J. Sheaffer, B.M. Fazenda, D.T. Murphy, and J.A.S. Angus. A simple multiband approach for solving frequency dependent problems in numerical time domain methods. In *Proceedings of Forum Acusticum*, pp. 269–274. S. Hirzel, 2011
- *J. Sheaffer, M. van Walstijn, and B.M. Fazenda. A physically-constrained source model for FDTD acoustic simulation. In *Proceedings of the 15th International Conference on Digital Audio Effects (DAFx-12)*, 2012
- J. Sheaffer, C. Webb, and B.M. Fazenda. Modelling binaural receivers in finite difference simulation of room acoustics. In *Accepted to Proceedings of the 21st International Congress on Acoustics (ICA)*, 2013

Other scholarly activities attended by the author are (abstract only):

- J. Sheaffer, B.M. Fazenda, and J.A.S. Angus. Computational modelling techniques for small room acoustics (A). In *Proceedings of the 1st CSE Doctoral School Research Conference*, University of Salford, November 2010. University of Salford
- J. Sheaffer and B.M. Fazenda. An integrated model of sound localisation in rooms (A). In *Acoustics and Audio Conference 2012*, Ben Gurion University, September 2012. IAA

Additionally, a multimedia presentation demonstrating the entire model is available online at: `http://www.jonsh.net/blog/?p=311`

—*Everything should be kept as simple as possible; but not simpler.*

Albert Einstein

2

The Physics and Perception of Spatial Sound

This chapter discusses theoretical topics in spatial sound, with an emphasis on sound localisation in enclosed spaces. It covers the physics of sound generation (section 2.1) and propagation in rooms (section 2.2), the physiology of spatial hearing (section 2.3) and the perception of spatial sound (section 2.4). Acknowledging that readers of this work may come from different backgrounds, the purpose of this chapter is to establish an agreed nomenclature and to provide an underlying basis for the wide range of theoretical subjects required for understanding the thesis. As it is not feasible to thoroughly cover all of these, some topics are only superficially introduced and readers are advised to consult the bibliography if more detailed explanations are needed.

2.1 Sound Propagation

This section outlines the fundamental theory of sound in free field. Unless otherwise stated, the this section is based on (Kuttruff, 2006), which provides the foremost reference for more detailed information.

2.1.1 The Linear Wave Equation

The propagation of sound in fluids can be described by the Navier-Stokes equations, whose nonlinear terms often make them difficult to solve. However, if an inviscid-adiabatic flow of energy is assumed, then they can be greatly simplified. As thermo-viscous losses are not a direct concern in this thesis, a sensible starting point would be to consider sound to be governed by the linearised Euler equations, namely the conservation of mass and conservation of momentum, shown in Eq. 2.1 and 2.2 respectively.

$$\frac{\partial p(\mathbf{x},t)}{\partial t} = -\rho_0 c^2 \nabla \cdot \mathbf{u}(\mathbf{x},t) \tag{2.1}$$

$$\rho_0 \frac{\partial \mathbf{u}(\mathbf{x},t)}{\partial t} = -\nabla p(\mathbf{x},t) \tag{2.2}$$

where $p(\mathbf{x},t)$ and $\mathbf{u}(\mathbf{x},t)$ are used to describe the sound pressure and particle velocity fields at $\mathbf{x} = (x,y,z) \in \mathbb{R}^3$, ρ_0 is the ambient density of air and c is the velocity of sound in air. Unless otherwise stated, the quantities 1.21kgm^{-3} and 343.5ms^{-1} are used respectively throughout the thesis. By differentiating Eq. 2.1 with respect to time and taking the divergence of Eq. 2.2, it is possible to eradicate the particle velocity vector and obtain the homogeneous wave equation,

$$\frac{1}{c^2}\frac{\partial^2 p(\mathbf{x},t)}{\partial t^2} - \nabla^2 p(\mathbf{x},t) = 0 \tag{2.3}$$

Note that for brevity, equations 2.2, 2.1 and 2.3 are given here without any source terms. A discussion on the Euler equations with sources and the derivation of the wave equation in inhomogeneous form are discussed in Chapter 4 as they are specifically needed for that chapter. In the frequency domain, wave behaviour can be described by the Helmholtz equation, which is the result of taking the

Fourier Transform of Eq. 2.3, and is given by

$$\nabla^2 P(\mathbf{x},\omega) + k^2 P(\mathbf{x},\omega) = 0 \tag{2.4}$$

where $k = \omega/c$ is the acoustic wavenumber, and $\omega = 2\pi f$ is the angular frequency.

2.1.2 Plane Waves

Although plane waves are of less concern in this work, the solution of Eq. 2.3 in 1D is key to simplifying many types of wave behaviour, as essentially at the far field of a source the wavefront can be treated as planar. In 1D, the Laplace operator, ∇^2, is simply a second order space derivative in the x-direction and the inhomogeneous wave equation becomes

$$\frac{1}{c^2}\frac{\partial^2 p(x,t)}{\partial t^2} = \frac{\partial^2 p(x,t)}{\partial x^2} \tag{2.5}$$

The general solution of Eq. 2.5, often referred to as the d'Alambert Solution, is given by

$$p(x,t) = \overrightarrow{p}(x-ct) + \overleftarrow{p}(x+ct) \tag{2.6}$$

which simply describes the superposition of two plane waves travelling in opposite directions[1]. If time harmonic waves of angular frequency ω are considered, then the general solution can be written in complex form as

$$p(x,t) = \underbrace{Ae^{j(\omega t+kx)}}_{-x\,direction} + \underbrace{Be^{j(\omega t-kx)}}_{+x\,direction} \tag{2.7}$$

where A and B are the pressure magnitudes of the two travelling waves. For *progressive plane waves*, the reverse-propagating component is neglected, resulting in

$$p(x,t) = Ae^{j(\omega t-kx)} \tag{2.8}$$

---

[1] $\overrightarrow{p}$ and $\overleftarrow{p}$ denote pressure waves travelling at opposite directions.

which is used as a fundamental building-block to describe many acoustic phenomena.

2.1.3 Spherical Waves

Of more importance to us is the propagation of sound in a three-dimensional space. At the absense of any limiting boundary conditions, the wave equation is straightforward to solve for the case of a spherically propagating wavefront. In such case, the Laplace operator in Eq. 2.3 can be described in an angle-independent spherical form as

$$\nabla^2 \equiv \frac{\partial^2}{\partial r^2} + \frac{2}{r}\frac{\partial}{\partial r} \tag{2.9}$$

where r denotes the radial distance from the point of origin. This means that for a spherical wave, the equation is dependent only on one space variable (r), and the general solution for a 1D wave equation can be applied, yielding

$$p(r,t) = \frac{1}{r} f\left(r - ct\right) \tag{2.10}$$

If it is assumed that the source is a Dirac function $\delta(t)$, then the solution to the wave equation represents the spatial transfer function between two positions in space $\mathbf{x}$ and $\mathbf{y}$, which is also referred to as the free field Green's function,

$$g(\mathbf{x};\mathbf{y}) = \frac{1}{4\pi R}\delta\left(t - \frac{R}{c}\right) \tag{2.11}$$

where $R = |\mathbf{x} - \mathbf{y}|$ is the distance between the two positions. The physical meaning of Eq. 2.11 is that the pressure at the observation point is simply a scaled and delayed version of the wave at the emitting point, whereas the shape of the wave, does not change. Of course, the shape of the source is dependent on the excitation signal and the sound generating mechanism, as will be further discussed in Chapter 4.

2.2 Sound in Enclosed Spaces

In this section, propagation of sound in enclosed spaces is discussed. Emphasis is given to boundary conditions and resulting sound reflections, which provide an important point of interest in the study of sound localisation. Unless otherwise stated, this section is also based on (Kuttruff, 2000), which provides the foremost reference for more detailed information.

2.2.1 Acoustic Impedance

To study the interaction of sound with different room surfaces, it is useful to discuss the *acoustic impedance* of the medium in which the wave propagates. As impedance is normally a frequency-dependent quantity, for mathematical convenience, the following analysis is performed in the frequency-domain. The acoustic impedance is defined as the ratio of the acoustic pressure to the particle velocity, which can be described as

$$Z_w(\omega) = \frac{P(\omega)}{U(\omega)} \tag{2.12}$$

where $U(\omega)$ is considered to be the normal component of the particle velocity. The notation Z_w is used to distinguish the surface or wall impedance from the impedance of air, Z_0. It is often convenient to normalise the surface impedance according to the characteristic impedance of air, yielding

$$\xi_w(\omega) = \frac{Z_w(\omega)}{Z_0} = \frac{Z_w(\omega)}{\rho_0 c} \tag{2.13}$$

where ξ_w is called the *specific acoustic impedance* of the surface.

2.2.2 Boundary Conditions

In order to develop boundary conditions for the wave equation, let us consider the conservation of momentum. Integrating Eq. 2.2 with respect to time yields

$$\mathbf{u}(\mathbf{x}, t) = \int -\frac{1}{\rho_0} \nabla p(\mathbf{x}, t) dt \tag{2.14}$$

Considering now a wave propagating in the x direction, normal to a surface of impedance $\xi(\omega)$, Eq. 2.14 can be written in the frequency-domain as

$$U(x,\omega) = -\frac{1}{j\omega\rho_0}\frac{\partial P(x,\omega)}{\partial x} \tag{2.15}$$

Next, substituting Eq. 2.13 into 2.15 we obtain the boundary condition

$$c\xi_w(\omega)\frac{\partial P(x,\omega)}{\partial x} + j\omega P(x,\omega) = 0 \tag{2.16}$$

whose inverse Fourier transform yields the boundary condition (for a right surface) in the time-domain,

$$\frac{\partial p(\mathbf{x},t)}{\partial t} = -c\xi_w\frac{\partial p(x,t)}{\partial x} \tag{2.17}$$

Note that in the time-domain, the specific impedance ξ_w is shown as a scalar parameter. In cases where the surface impedance is frequency-dependent, ξ_w can be approximated in the discrete domain with a *digital impedance filter*, as shall be further discussed in Chapter 3.

2.2.3 Reflection and Absorption

To explain the interaction of sound waves with surfaces, let us consider a plane wave travelling in the $x+$ direction towards a wall. The incident pressure normal to the wall, p_i, is given by

$$p_i(x,t) = A_0 e^{j(\omega t - kx)} \tag{2.18}$$

where A_0 is the amplitude of the incident pressure. When the wave reaches the wall, some alterations in the wave structure and behaviour may occur. First, its amplitude and phase may change according to the acoustic properties of the wall, and second, its direction of propagation is reversed. Therefore, the reflected pressure normal to the wall, p_r, is given by

$$p_r(x,t) = A_0 \hat{r} e^{j(\omega t + kx)} \tag{2.19}$$

where $\hat{r}$ is the *reflection coefficient* which shall be shortly related to the impedance of the wall. The total sound pressure is the superposition between the incident and reflected sound fields, which for a boundary placed at $x = 0$ is equal to

$$p(0, t) = (1 + \hat{r})A_0 e^{j\omega t} \tag{2.20}$$

To find the relationship between the reflection coefficient $\hat{r}$ and the boundary impedance, one should also consider the particle velocity component of the incident and reflected sound fields, as impedance is in essence the ratio of pressure to velocity. For brevity, the entire mathematical process is not shown here, however interested readers may find it in any standard acoustics textbook. Accordingly, it can be shown that the reflection coefficient is

$$\hat{r} = \frac{\xi_w - 1}{\xi_w + 1} = \frac{Z_w - \rho_0 c}{Z_w + \rho_0 c} \tag{2.21}$$

The normal range of the reflection coefficient is $-1 \leq |\hat{r}| \leq 1$, which may be used to characterise any type of boundary. When $|\hat{r}| = -1$, the boundary is referred to as a *soft wall* or *phase-reversing boundary*, as it reverses the phase of the wavefront whilst fully reflecting the incident energy. For this to occur, the surface impedance must equal $Z_w = 0$. If the wall is (at normal incidence) perfectly absorbing, then $Z_w = \rho_0 c$, $\xi_w = 1$ and therefore $|\hat{r}| = 0$. For the case of a *rigid wall*, often referred to as a *phase-preserving boundary*, the reflection coefficient is $|\hat{r}| = 1$, which requires that $Z_w = \xi_w = \infty$.

Recall that the underlying assumption used in the above analysis is that the wave propagates in the positive x direction, meaning that it interacts with a boundary situated at $x = 0$ at *normal incidence*. If the wave arrives at an *oblique* incident angle θ, then the reflection coefficient becomes angle-dependent and is given by

$$\hat{r}(\theta) = \frac{\xi_w \cos\theta - 1}{\xi_w \cos\theta + 1} = \frac{Z_w \cos\theta - \rho_0 c}{Z_w \cos\theta + \rho_0 c} \tag{2.22}$$

Another important acoustic property of a material is its *absorption coefficient*, which is described as the ratio of the incident energy to the energy *not* reflected by the surface (i.e. transmitted through the material, or dissipated as heat in processes involving viscous drag and/or mechanical motion). The absorption coefficient $\acute{\alpha}$, is related to the reflection factor by

$$\acute{\alpha}(\theta) = 1 - \left|\hat{r}(\theta)\right|^2 = \frac{4\mathfrak{Re}\left\{\xi_w\right\}\cos\theta}{(\mathfrak{Re}\left\{\xi_w\right\}\cos\theta)^2 + 2\mathfrak{Re}\left\{\xi_w\right\}\cos\theta + 1} \tag{2.23}$$

where $\mathfrak{Re}\left\{\xi_w\right\}$ is the real component[2] of the specific impedance ξ_w. In general, the acoustic behaviour of the reflecting surface is strictly dependent on the angle of incidence. There are, however, special cases where sound waves cannot propagate along the surface. In physical terms, this means that the normal component of the particle velocity depends only on the pressure in front of it. In such cases, the surface is considered to be *locally reacting.* Although in many cases walls are non-locally reacting, the concept of locally reacting surfaces provides an important approximation which allows boundary models to be greatly simplified. In this thesis, the term *LRS* is used to denote processes that strictly adhere to the physics of locally reacting surfaces. When the term *non-LRS* is used, the process does not adhere to LRS theory. This does not necessarily mean that it satisfies *any other* physical condition.

2.2.4 The Room Impulse Response

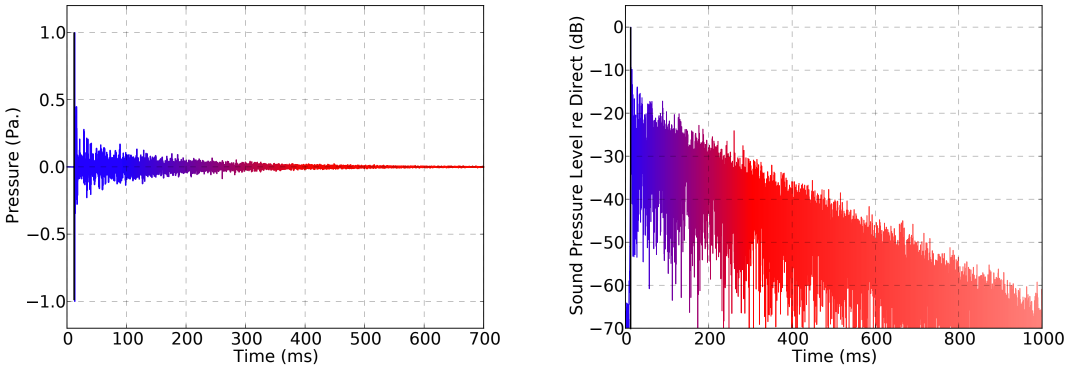
In most practical cases, a room can be regarded to as a *linear time-invariant* system whose spatial transfer function[3] is characterised by an *impulse-response*, or *RIR*. An example of a measured RIR of a 1700m^3 concert hall is shown in Figure 2.1(a). The soundfield represented by the room's transfer function can be decomposed into three principal components. The *direct sound* is simply a scaled and time-shifted version of the presented excitation function. This component is followed by *specular reflections*, often called *early reflections*, which is comprised of discrete reflections representing the early and non-diffuse portion of the soundfield. As time progresses, reflections become more dense in time and

---

[2] An explicit formulation is also possible for complex impedances, however these are of less concern in this work.

[3] That is, where the pressure at the source position is the system's input, and the resulting pressure at the receiving position is the system's output.

space, which causes reverberation to build-up. In the limiting case, when the response is entirely homogeneous and isotropic, then the soundfield is said to be completely diffuse.



(a) Room Impulse Response (b) Envelope Time Curve

Figure 2.1: Acoustic measurement of a small (1700m^3) concert hall. (a) Room Impulse Response and (b) Envelope Time Curve. For illustrative purposes, the direct component of the soundfield is marked in black, the early energy is marked in blue and the gradient from blue to read shows the build-up of the reverberant field.

Since the RIR precisely characterises the room for a specific source-receiver pair, it provides an efficient starting point for studying room acoustics, most noticeably by means of subsequent computational analysis and/or auralisation. It should be noted at this point that as the RIR is only a measure of sound pressure, it does not carry any directional information about the soundfield.

Whilst a great deal of objective information can be computed from a RIR, it is less suitable for tasks involving visual inspection. For such purposes, it is common to extract the envelope of energy decay, which is accomplished by taking the Hilbert transform of the RIR and representing its magnitude on a log-scale (Heyser, 1971; Vanderkooy and Lipshitz, 1990), as shown in Figure 2.1(b). This representation is often referred to as the *Envelope-Time Curve* (ETC) of the room.

2.3 The Physiology of Directional Hearing

In this section, the peripheral processes and neuro-physiology of directional hearing are discussed. Only topics which are directly related to or indirectly inspire models developed in this thesis are discussed. For a rigorous introduction, readers are referred to (Schnupp et al., 2010) which, unless otherwise stated, is the primary reference for this section.

2.3.1 The Peripheral Auditory System

The function of the peripheral auditory system is to convert sound pressure in the acoustical domain to neural firing rates at the auditory nerve, which can be further processed by the central nervous system. A simplified schematic representation of the auditory periphery is shown in Figure 2.2. Although this system is generally described in much more detail, here we focus only on elements which are explicitly related to the computational models employed in this thesis.

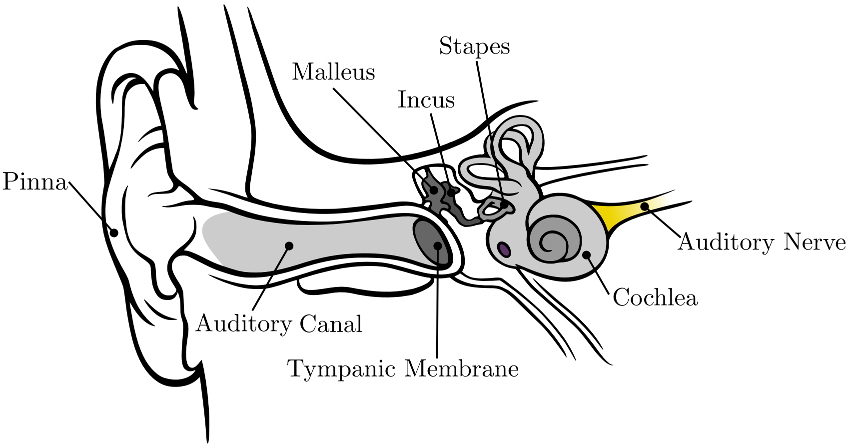


Figure 2.2: Simplified schematic representation of the peripheral auditory pathway, adapted from (Chittka and Brockmann, 2005). Figure modified and used under the creative-commons license.

Waves arriving at the pinna travel down the *auditory canal*, which is an irregularly shaped tube whose average diameter is about 0.65cm horizontally and 0.9cm vertically, and whose length is approximately $2.5 - 3.5$cm. The auditory canal is sealed on its inner side by a conically-shaped tissue called the *tympanic membrane*. This structure can be viewed as an open-closed tube, whose fundamental resonant frequency lies within the 2-5kHz range. In the middle ear, the mechanical vibration of the tympanic membrane induces motion in three inter-connected *ossicles*, namely the *malleus*, *incus* and *stapes*[4]. The stapes then pushes on an ovally-shaped membrane covering the oval window, which is the intersection between the middle ear and the cochlea. As the cochlea is filled with endolymph and perilymph, which are fluids whose acoustical properties resemble those of slightly salted water, the middle ear acts as an impedance matching mechanism, allowing for an effective transmission of acoustic power.

A cross section of cochlea is shown in Figure 2.3. It is an incompressible-fluid-filled spiral structure residing within the temporal bone, along which lies the *basilar membrane* and *Reissner's membrane* who divide it into three sub-spaces, namely the *scala tympani*, *scala media* and *scala vestibuli*.

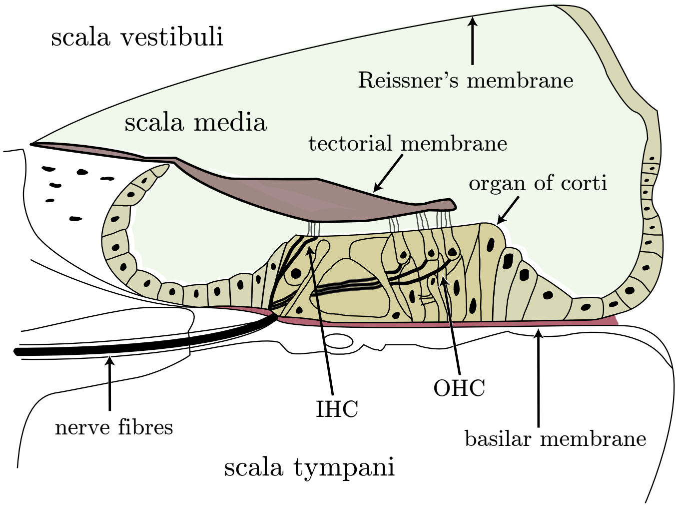


Figure 2.3: Cross-section of the cochlea. IHC and OHC stand for *inner hair cells* and *outer hair cells* respectively. Figure modified and used under the creative-commons license.

When the stapes pushes and pulls on the oval window, the fluid in the cochlea induces motion on the basilar membrane which vibrates at different places according to the spectral content of the signal. This mechanism in essence performs a frequency analysis of the signal, which entails that information is further transmitted in separate frequency channels. From a logical point of view, the process

---

[4] Latin for hammer, anvil and stirrup.

can be seen as if the signal is passed through a network of auditory filters each having its own critical bandwidth. In this work, auditory filters are either referred to by their centre frequencies or by an index number, $\dot{k}$, whose relation to its centre frequency f_c is given by (Glasberg and Moore, 1990)

$$f_c = 228.833 \left(e^{0.11\dot{k}} - 1\right) \tag{2.24}$$

Attached to the basilar membrane is the *organ of corti* whose function is to convert the mechanical vibrations of the membrane to electrical signals that can be further transmitted to the brainstem. The organ of corti features sensory hair cells, which deflect from their resting position when the basilar membrane vibrates and causes the organ on corti to move with it respectively. This deflection is due to the existence of the *tectorial membrane* which pushes against the tips of the hair cells when the organ of corti is moved[5]. When hair cells are deflected, their *tip-links* stretch to allow influx of fluid from the inner part of the organ of corti. Because this fluid is saturated with potassium ions which have a positive electric charge, each corresponding hair cell membrane develops a voltage potential across itself. Thus, the mechanical fluctuations of the basilar membrane are represented as periodic changes in voltage potentials of corresponding hair cells. This behaviour resembles a capacitor in an electric circuit. In a similar manner, charging and discharging of the cell is governed by a certain time constant, which causes the system to respond differently to different stimulating frequencies. At low frequencies the system is able to faithfully reproduce the fine structure of the waveform, as fluctuations are sufficiently slow and allow for a complete depolarisation-repolarisation process to occur. At high frequencies, the system is not able to keep up with stimulus, which results in continuous depolarisation of the cell's membrane potential. This manifests itself as DC-offset in system's output signal.

Unlike traditional neurons, hair cells do not fire an action potential as a re-

---

[5] In fact, this is thought to happen only with *outer hair cells*, whereas motion of the *inner hair cells* is thought to take place because of motion of fluid trapped between the organ of corti and the tectorial membrane.

action to depolarisation. Instead, they have synaptic connections with neurons of the *spiral ganglion*, which form the axons of the auditory nerve. This connection is accomplished by a form of *glutamatergic* neurotransmission, in which the rate of the cell's glutamate release is dependent on the deflection of the hair cell. The greater the deflection, the more depolarised the membrane voltage and the greater the release of glutamate. Since the firing rate of spiral ganglion nerves[6] is dependent on the rate of glutamate release, then deflection of hair cells (and therefore the instantaneous amplitude of the stimulating signal) is encoded as time-varying neural firing rates.

2.3.2 Processing of Localisation Cues in the Auditory Pathway

Head Related Effects

The RIR discussed in section 2.2.4 describes the transfer function between two positions in space. However, if the receiver is human, then in fact two impulse-responses should be considered, describing the transfer functions between the source position and each of the listener's ears. When these transfer functions are measured in free-field they are often referred to as a *Binaural Impulse Response* (in short *BIR*), whereas when they are measured in a room the term *Binaural Room Impulse Response* (in short *BRIR*) is used. Because the two ears are spaced apart, the wavefront's time and level of arrival to each of the ears will change according to the incident angle. This is often described by interaural differences, namely *interaural time differences* (ITD) and *interaural level differences* (ILD), which are the foremost characteristic cues for localisation on the horizontal plane, as shall be further discussed. Moreover, as the head serves as an acoustic obstacle causing incident waves to diffract around it, the transfer function is not only angle-dependent but also frequency-dependent. An example of a BIR for a source placed at 45° is shown in Figure 2.4.

The notion that localisation on the horizontal plane is dominated by ITD and ILD cues is often referred to as *The Duplex Theory of Sound Localisation*,

---

[6] here we refer to Type I nerves.

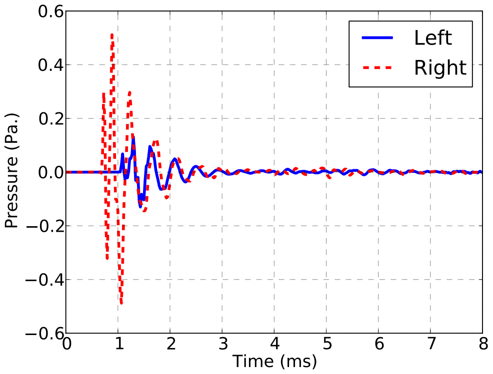


Figure 2.4: Binaural Impulse Response (BIR) for a source positioned at 45° from a Kemar mannikin, obtained from the MIT HRTF catalogue (Gardner et al., 1994).

originally suggested by Rayleigh (1896). Whilst the roots of sound localisation study appear long before Rayleigh, it is common to think of duplex theory as the first scientifically verified theory of sound localisation. For an articulate historical review see (Gulick, 1971; Hickson and Newton, 1981). When a sound source is presented at a certain angle from the listener, there will be a difference in the propagation paths between the source and each of the listener's ears[7], as shown in Figure 2.8. This results in inter-aural differences in time and level of arrival, which provide the foremost cues for localisation on the horizontal plane. According to Duplex Theory, localisation at low frequencies is dominated by *interaural time differences* (ITDs) whereas localisation at high frequencies is dominated by *interaural level differences* (ILDs) due to the shading effects of the head.

Whilst considered a landmark achievement, duplex theory cannot explain how the auditory system resolves front-back confusion and localisation in the median plane. This suggests that other cues are also involved. The role of the pinna in sound localisation have been thoroughly studied in (Angell and Fite, 1901; Batteau, 1967; Fisher and Freedman, 1968; Blauert, 1970) showing that the geometric structure of the pinna causes variance in the signal's response. In other words, the main function of the pinna in sound localisation is to modify the spectral contents of the input signal according to the direction of arrival of the incident wave, effectively resulting in an angle-dependent transfer function. This

---

[7] except for 0° or 180° incidence

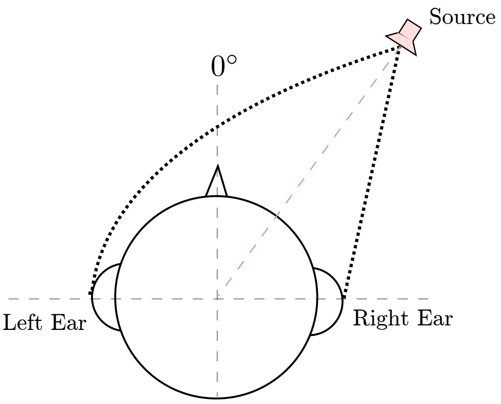


Figure 2.5: Illustration of the different propagation paths from a source to the two ears of a listener, resulting in interaural time and level differences.

is accomplished by superposition of a direct wave and a wave diffracted within the unique structure of the pinna, which results in a different propagation path for different angles of arrivals. The difference in path length between the direct and diffracted components cause a phase difference, which manifests itself in a different *comb-filter* structure for different incidence angles when the two components are summed together. Additional cues include head movements and visual information (Wallach, 1940), however these are less related to this work.

Cue Encoding in the Brainstem

In physiological terms, interaural differences are reflected in differences between signals transmitted through the ipsilateral and contralateral auditory nerves, whereas monaural cues are encoded within the spectral contents of the signals themselves. This leaves the central auditory system with an immense amount of information to process, and not surprisingly, the neural pathway of spatial processing is quite complicated. Since the perceptual models discussed in this thesis are analytically-motivated, a rather simplified overview of the neural processes corresponding to sound localisation would suffice. An illustration of the ascending auditory pathway in the central nervous system is shown in Figure 2.6.

Auditory nerve fibres connect the cochlea to the first processing area in the brainstem, the *cochlear nucleus* (CN), which hosts many distinct types of neurons,

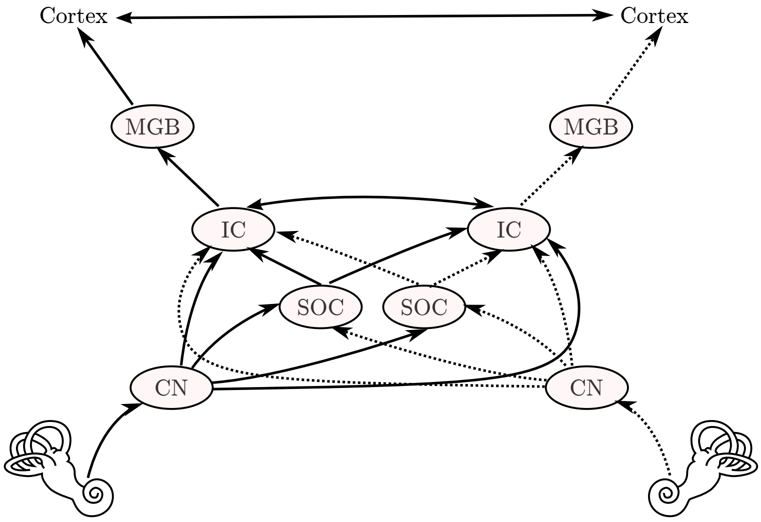


Figure 2.6: Simplified illustration of the ascending auditory pathway. CN, cochlear nucleus; SOC, superior olivary complex; IC, inferior colliculus; MGB, medial geniculate body. Solid lines - ipsilateral pathway; Dotted lines - contralateral pathway. Figure based on (Schnupp et al., 2010)

each playing a different role in the analysis of the signal's acoustic properties, such as spectral content, signal onset, temporal structure and periodicity, to name a few. Information from the CN is then projected to the *superior olivary complex* (SOC) which plays an important role in the processing of binaural cues. Encoding of ILDs occur at the *lateral superior olive* (LSO), whereas encoding of ITDs occur at the *medial superior olive* (MSO). The neural pathways for these cues are shown schematically in Figure 2.7. It is important to recall that information is transmitted from the cochlea in separate frequency channels, meaning that processing in these brainstem areas should also be carried out in different frequency bands. In fact, throughout the auditory pathway, regions of the brain processing different frequencies are topologically located along an axis representing the frequencies they process. This is referred to as a *tonotopic arrangement*.

Neurons of the LSO are excited by signals from the ipsilateral ear but are inhibited by signals from the contralateral ear. Cells of the *anteroventral cochlear nucleus* (AVCN) provide excitation bilaterally, however the connection of contralateral AVCN to the LSO is relayed through the *medial nucleus of the trapezoid body* (MNTB) which make the projection inhibitory. This *excitation-inhibition* (EI) process is the basis of encoding ILD cues in the LSO. For example, if the azimuth of the sound source is 0°, then the intensity at both ears is identical and

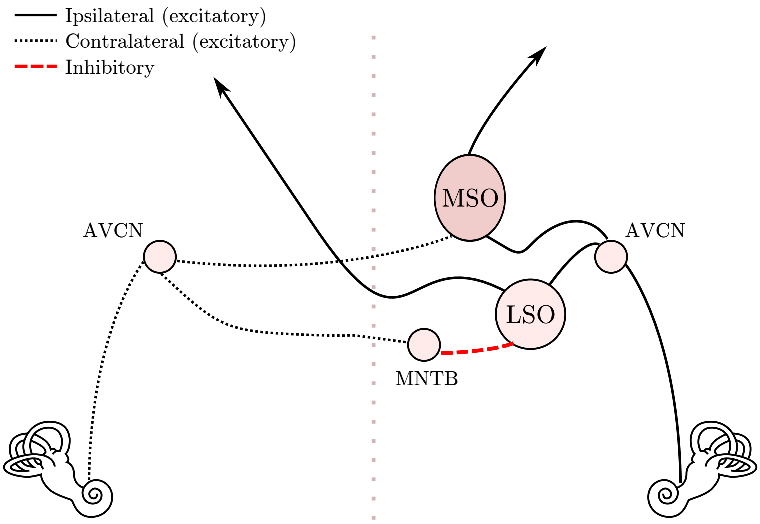


Figure 2.7: Schematic representation of ILD and ITD processing in the auditory brainstem. AVCN, anteroventral cochlear nucleus; MNTB, medial nucleus of the trapezoid body; LSO, lateral superior olive; MSO, medial superior olive. For visual clarity, pathway is shown only for the ipsilateral side. Figure based on (Yin, 2002).

there would be an equal amount of excitation and inhibition, resulting in a low firing rate. As the source is moved towards the ipsilateral ear, the increase in ILD will manifest itself in more excitation and less inhibition. Since excitation and inhibition essentially affect a neuron's firing rate, then ILDs are represented as rate-coding in LSO neurons.

Encoding of ITD cues is somewhat more complicated, and the exact process is still not entirely agreed upon. Perhaps the most widely accepted explanation relies on a landmark model suggested by Jeffress (1948), whose underlying assumption is that MSO neurons act as coincidence detectors. In contrast to the LSO, the MSO receives excitatory inputs bilaterally, meaning that its neurons fire at a maximal rate when ipsilateral and contralateral excitation occurs at the same time. If one considers a single MSO neuron, then the transmission path length from the ipsilateral and contralateral AVCNs depends on the physical location of the neuron. For example, a neuron who is equidistant from both AVCNs will respond well when there is no interaural delay, and a neuron which is physically closer to the ipsilateral side will respond well if the interaural delay is biased against the contralateral side, essentially compensating for the difference in path length and allowing for the two signals to coincide. In other words, these *excitation-excitation* (EE-type) neurons are thought to be tuned to specific ITD values, or *best delays*. This suggests that ITD is represented in the form of a *place code*, in which different neural assemblies (residing in different places in the MSO) are activated as a response to different interaural delay times.

Higher Processing in the Midbrain and Cortex

The MSO and LSO effectively encode binaural cues and as such, provide more refined directional information to upper stages of the auditory system. Still, in order to localise a sound these data should be mapped to corresponding spatial positions. The *inferior colliculus* (IC) is an area of convergence for different pathways, including outputs of the LSO and MSO, and is the first area where binaural and monaural cues are integrated. There is physiological evidence drawn from experiments in barn owls, that IC neurons are tuned to specific combinations of ITDs and ILDs (Peña and Konishi, 2002), effectively establishing a map of the surrounding auditory space[8]. Furthermore, in the IC of such species, information delivered across different frequency channels is integrated.

This, however, is not the end of the auditory pathway involving sound localisation. Visual cues play an important role in sound localisation, as well as feedback from head movements. The IC of mammals is linked to a multisensory structure in the midbrain, called the *superior colliculus* (SC). This area features neural maps of the visual space, and controls reflexes such as eye and head movements, which provide an improved means of auditioning localisation cues. It is understandable that all of these processes occur subcortically, as reflex behaviour is essential for survival. Yet, the auditory cortex (A1) also contributes to sound localisation, in addition to being responsible for our ability to identify sounds. Nevertheless, as higher stages of the analytical model presented in this thesis are not explicitly related to neurophysiological evidence of processes occuring in the cortex, a detailed review of auditory processing in the cortex is outside the scope of this chapter.

---

[8] Note that barn owls use ITDs and ILDs for localisation in both azimuth and elevation

2.4 Psychophysical Basis of Sound Localisation

In the previous section, the physiology of sound localisation was overviewed. However, there are still some phenomena supported by psychophysical evidence, but whose direct neurophysiological correlates are not well understood. Thus, it is important to establish the psychoacoustic basis of sound localisation. Furthermore, the reader is reminded that the goal of this thesis is to design an analytical model which is *psychoacoustically* plausible, but not neccesarily *physiologically* accurate. For convenience, all mathematical analyses in this section are done by assuming that sources are sufficiently apart from the listener, such that plane waves can be used.

2.4.1 A Single Source in Free Field

Considering a listener and a single sound source in free field, as shown in Figure 2.8, in which the path length from the source to the left and right ears is r_L and r_R respectively, then at the absence of head shadowing, the complex pressure at the ears can be calculated from

$$p_L(t) = \frac{A_s}{r_L} e^{j(\omega t - kr_L)} \tag{2.25}$$

$$p_R(t) = \frac{A_s}{r_R} e^{j(\omega t - kr_R)} \tag{2.26}$$

where A_s is the amplitude of the source. In such an oversimplified case, the ILD is the difference between the phasors' magnitudes and the ITD can be calculated from the phase components. However, the presence of the head introduces diffraction which affects both ITD and ILD in a frequency dependent manner. In order to analyse the effects of diffraction, it is common to use a rigid-sphere approximation of the human head geometry, as shown in Figure 2.8.

In (Brungart and Rabinowitz, 1999), the complex sound pressure on the surface of a rigid sphere, $\mathbf{P}_h$ has been formulated using expansion in spherical-

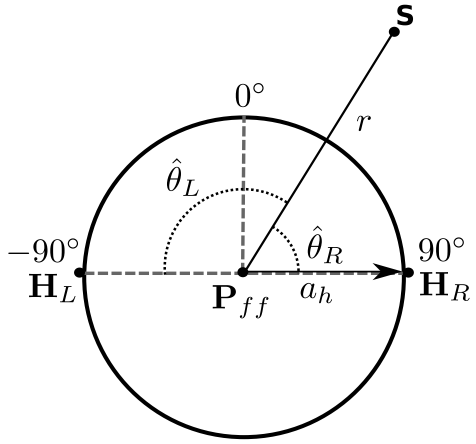


Figure 2.8: Geometry of a rigid sphere approximating a human head. Source is at point S, r is radial distance to the sphere's centre, a_h is the radius of the sphere, and $\hat{\theta}_L$ and $\hat{\theta}_L$ are the angles between the vector from the source to the centre of the sphere and the left and right ear positions, respectively.

harmonics. With some notational changes, their model is given by

$$\mathbf{P}_h(r, a_h, \hat{\theta}, \omega) = \frac{c\rho_0 \hat{q}}{2\pi a_h^2} \sum_{m=0}^{\infty} \left(m + \frac{1}{2} \right) P_m(\cos\hat{\theta}) \frac{H_m(kr)}{H'_m(ka)} e^{-j\pi/2} \tag{2.27}$$

where $\hat{q}$ is the volume velocity of the source, $P_m(\cdot)$ is the Legendre polynomial function, $H_m(\cdot)$ is the spherical Hankel function and $H'_m(\cdot)$ is its first derivative with respect to the argument. The parameter $\hat{\theta}$ is the angle between the two vectors representing rays from the centre of the sphere to the source, and to the sought point on the surface of the sphere. If the sphere is used to approximate a head in which the two ears are located at $\pm\pi/2$, then $\hat{\theta}$ for the left and right ears can be related to the incidence azimuth θ and elevation ϕ by

$$\hat{\theta}_L = \arccos(\sin\theta\cos\phi) \tag{2.28}$$

$$\hat{\theta}_R = \frac{\pi}{2} - \arcsin(\sin\theta\cos\phi) = \pi - \hat{\theta}_L \tag{2.29}$$

It is convenient to think of the head (or the sphere approximating it) as a linear system which is characterised by a transfer-function (in the frequency domain) or by an impulse-response (in the time domain). We therefore use the terms *head-related transfer function* (HRTF) or *head-related impulse response* (HRIR), respectively, in order to explain how the system modifies its input signal. Accordingly, the HRTF at each ear can be mathematically represented as the ratio of pressure at the ear to the free-field pressure at the receiving position (i.e. at the centre of the head, when the head is absent). If the free field pressure is given by

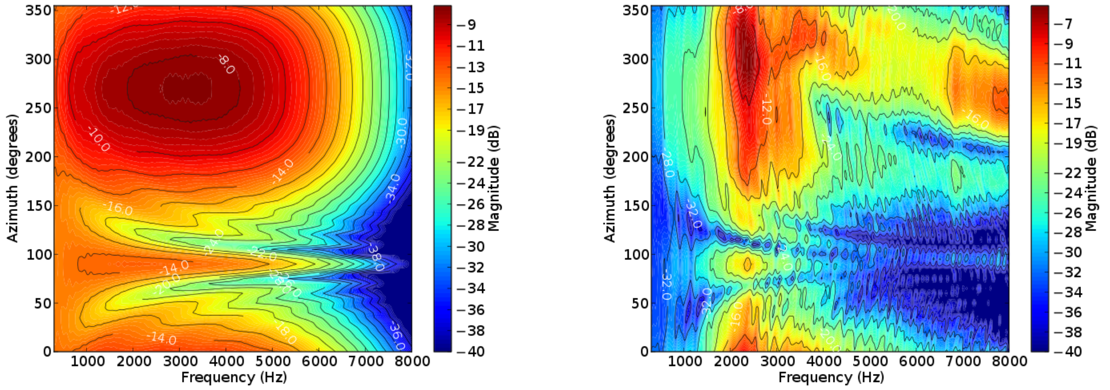
$$\mathbf{P}_{ff}(r,\omega) = \frac{\rho_0 \hat{q}\omega}{\pi r} e^{j(\pi/2 - kr)} \tag{2.30}$$

then the HRTFs for the left and right ears can be obtained from

$$\mathbf{H}_L(r, a_h, \hat{\theta}, \omega) = \frac{\mathbf{P}_h(r, a_h, \hat{\theta}_L, \omega)}{\mathbf{P}_{ff}(r,\omega)} \tag{2.31}$$

$$\mathbf{H}_R(r, a_h, \hat{\theta}, \omega) = \frac{\mathbf{P}_h(r, a_h, \hat{\theta}_R, \omega)}{\mathbf{P}_{ff}(r,\omega)} \tag{2.32}$$

Where $\mathbf{H}_L$ and $\mathbf{H}_R$ denote the HRTFs at the left and right ears respectively. Note that since pressure is expressed in the complex frequency domain, the resulting transfer functions are also complex. The left-ear HRTF magnitude for a rigid sphere and for a measured Kemar mannikin are shown in Figure 2.9.



(a) Rigid Sphere

(b) Kemar Mannikin

Figure 2.9: Left-ear HRTFs for (a) rigid sphere model and (b) measurements of a Kemar mannikin. Measured data obtained from (Gardner et al., 1994).

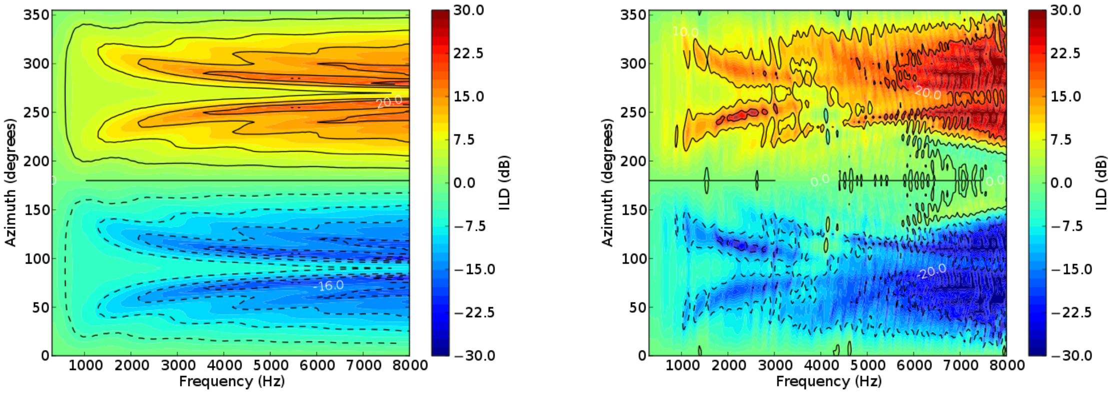
It is not surprising that the rigid sphere HRTF lacks many of the spectral features present in a complete head model, however, of more importance to us are the interaural differences between two bilaterally corresponding HRTFs. The interaural level differences can be calculated by

$$\Delta L(\omega, \theta, \phi) = 20 \log \left| \frac{\mathbf{H}_L(\omega, \theta, \phi)}{\mathbf{H}_R(\omega, \theta, \phi)} \right| \tag{2.33}$$

Similarly, in the frequency domain the interaural time differences denote the amount of time delay of each frequency component, which can be calculated from the phase-delay of the two signals,

$$\Delta T(\omega, \theta, \phi) = -\frac{1}{\omega} \angle \frac{\mathbf{H}_L(\omega, \theta, \phi)}{\mathbf{H}_R(\omega, \theta, \phi)} \tag{2.34}$$

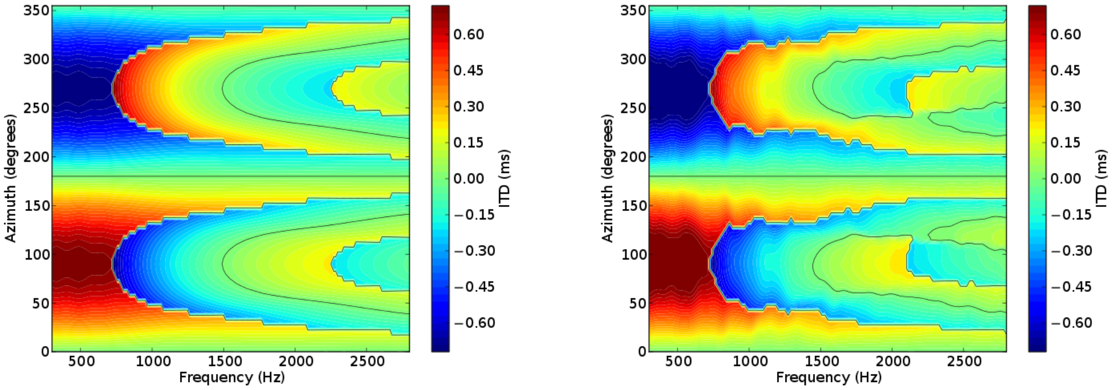
The interaural cues for a rigid sphere and a Kemar mannikin are shown in Figures 2.10 and 2.11, respectively. It can be seen that at low frequencies, the ILD is almost angle independent whereas the ITD increases as the source is moved laterally. On the contrary, at higher frequencies ITDs have very little spatial variance, whereas ILDs are extremely angle dependent. Not surprisingly, these *physical* attributes are in agreement with the *perceptual* properties which characterise the aforementioned duplex theory of sound localisation. The ITD curves show very clearly the transition frequency above which interaural phase becomes ambiguous (note the abrupt change in colour seen in Figure 2.11). It should be noted that this transition frequency is reduced as the source is moved laterally, where at ±90° the extreme phase shift occurs at about 700Hz.



(a) Rigid Sphere

(b) Kemar Mannikin

Figure 2.10: Frequency dependent ILDs for (a) rigid sphere model and (b) measurements of a Kemar mannikin.



(a) Rigid Sphere

(b) Kemar Mannikin

Figure 2.11: Frequency dependent ITDs for (a) rigid sphere model and (b) measurements of a Kemar mannikin.

Even though phase differences become ambiguous above $\sim$ 700Hz, the auditory system may still utilise the ITD mechanism at higher frequencies. Truly, it is not possible to precisely determine time differences from the *fine structure* of the waveform, and this is indeed a limitation when pure tones are used. If the waveform is complex then it is possible to determine time differences by comparing the stimuli's envelope[9]. Therefore, low-frequency time differences relying on the waveform's fine structure are seen as a special case of ITDs, and are referred to as *interaural phase differences* (IPDs).

Interaural cues provide an effective way of localising sound, as each angle of incidence corresponds to a different, frequency-dependent combination of ITD and ILD. There are, however, some limitations to the use of interaural cues. Observation of Figures 2.10(a) and 2.11(a) clearly reveals that both ITDs and ILDs are perfectly symmetrical (but inverted) around 180°. This suggests that if the auditory system relied solely on interaural cues, then discrimination of front to back should not have been impossible. However, note that for the case of a Kemar manikin, in Figures 2.10(b) and 2.11(b), cues are no longer precisely symmetrical, even at frequencies where pinna cues are absent. This is because the human head is not symmetrical across the horizontal plane, and as will be discussed in Chapter 6, it should be possible to discriminate front to back even if only interaural

---

[9] Note the resemblence of this to the behaviour of the neural transduction mechanism discussed in section 6.1

cues are considered. Another interesting symmetry occurs at frequencies below about 3000Hz. For both ITD and ILD, the patterns are symmetrical around lateral angles, at 90° and 180° just below 3kHz. Thus, at the absence of spectral cues, ITD and ILD combinations may cause ambiguity as their uniqueness is compromised. This attribute is the physical basis of the perceptual concept called the *cone of confusion*.

Localisation blur is a metric to quantify the uncertainty in identifying the direction of a sound. This is often measured by the Minimum Audible Angle (MAA), which denotes the JND at a specific angular position (Blauert, 1997). On the frontal axis, localisation-blur ranges from 1° to 10° depending on the type of stimulus and listening conditions. For example, for pure-tone sounds the MAA is 1° at normal incidence and about 7° for sounds at a 75°, thus indicating that localisation blur increases markedly as sources are moved laterally (Moore, 1982). For broadband sources, the mean localisation error is 5° on the median plane, and up to 20° bilaterally (Hofman and Van Opstal, 1998). Makous and Middlebrooks (1990) conducted extensive subjective tests of localisation of broadband sources. Figure 2.12 shows their results for an 'open-loop' test. It can be seen that localisation along the median plane is relatively constant, and that blur gradually increases as the stimuli is moved to the sides.

2.4.2 A Single Source in the Presence of Reflections

Humans have a remarkable ability to localise sounds even under reverberant conditions. Although the direct component of the sound field is followed by numerous reflections, in most cases the auditory system is able to suppress the directional information carried by these reflections. This is owing to different physiological and perceptual mechanisms, often referred to simply as the *precedence effect* (Zurek, 1987; Hartmann, 1997; Blauert, 1997; Litovsky et al., 1999). When the auditory system is presented with two coherent stimuli with arrival time differences (i.e. a lead-lag situation[10]) different mechanisms are utilised to focus on directionally meaningful information. In other words, whilst reflections do carry

---

[10] the leading and lagging sources refer to the first arriving and second arriving wavefronts, respectively

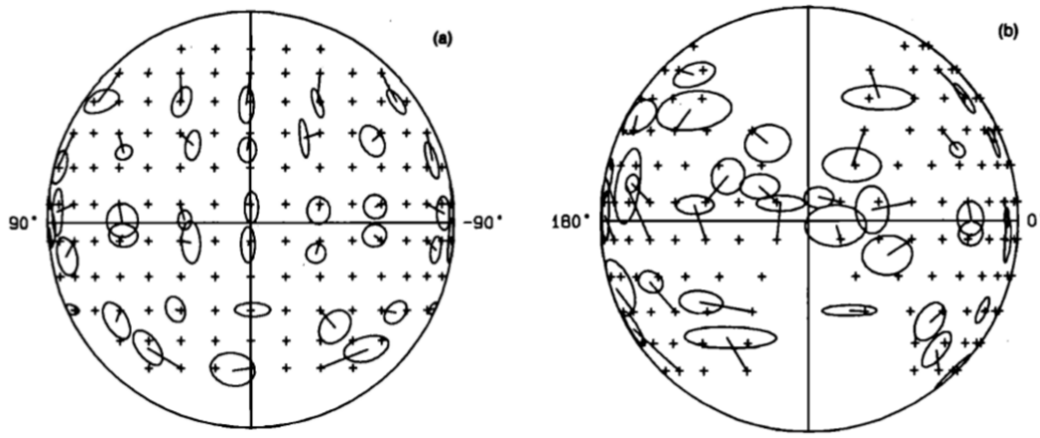


Figure 2.12: Copied figure from (Makous and Middlebrooks, 1990). Localisation of a 150ms broadband source shown from (a) the front and (b) the right side. Cross signs indicate the actual stimulus location and open circles indicate the mean results from six trials, connected to their corresponding stimulus by a solid line. Results are for the 'open-loop' test, meaning the decision is based on instantanous head movement; thus eliminating additional localisation cues that could have been generated by voluntary head movement. Reprinted with permission from the American Institute of Physics. Copyright 1990, Acoustical Society of America.

spatially important information, they may interfere with our ability to correctly localise sound, and should thus be suppressed in a localisation task.

When time differences are small (less than 1ms), the lead and the lag stimuli are perceptually fused and the locus of the auditory event is determined as the average locus of the lead and lag sources. In the literature, this phenomenon is referred to as *summing localisation.* At slightly higher delays, the two sources remain fused (at 1-5ms) or are perceived with some spatial widening (5-50ms), but the locus is determined by the position of the lead stimulus only. This is often referred to as *localisation dominance.* When much higher delays are introduced, stimuli are perceived as completely separate auditory events, or simply an acoustic echo (Litovsky et al., 1999). It should be noted that these values represent crude thresholds, which strongly depend on spectral and temporal characteristics of the stimuli, their direction of arrival, as well as on the way they are presented, be it through loudspeakers in free field, e.g. in (de Boer, 1946; Snow, 1953; Leakey and Cherry, 1957; Wendt, 1963; Olive and Toole, 1989; Bech, 1998); or over headphones, e.g. in (Zurek, 1980; Gaskell, 1983; Yost and Soderquist, 1984; Shinn-Cunningham et al., 1993).

Moreover, the precedence effect is known to have dynamic properties. When a stimulus pair is presented to the auditory system, there are some level and time detection thresholds above which the auditory system is aware of the existence of the lag stimulus. However, if such stimuli are repeated a number of times, the threshold becomes longer, which is suggestive of a buildup process. This effect has been identified by Clifton (1987) and was further investigated by Djelani and Blauert (2001). Interestingly, this buildup effect restarts whenever spatial variables change. This has been originally interpreted as a breakdown of the precedence effect, meaning that when the auditory scene changes the effect ceases to exist and the auditory system employs a new buildup term in order to spatially re-analyse the scene (Clifton et al., 1994).

Another precedence-related phenomena is the Franssen effect (Franssen, 1960), which is also related to the basis of stereophonic reproduction (Franssen, 1964). If two stimuli are presented, one with an impulsive onset gradually replaced with another stimulus at a different position in space, the source locus is identified at the position of the first stimulus, even when it has completely ceased to exist. This illusion works well only in reflective environments, and when the presented stimuli consist of sounds which are known to be difficult to localise, such as mid-frequency pure tones (Hartmann and Rakerd, 1989).

2.4.3 Perception of Reflections and Reverberation

Spatial effects caused by reflections are not unconditional, and do not correspond to a simple audible/not-audible criterion. From a perceptual perspective, they can be categorised by different types of thresholds. Reflections may have a *detection threshold* below which they are completely inaudible, an *image-shift* threshold above which they are audible and fused but contribute to some spatial effects such as image shift, localisation blur or widening of the apparent source occupying area (*apparent source width*, or ASW), and an *echo-threshold* above which two separate auditory events are audible. Figure 2.13 shows subjective detection thresholds for a single lateral reflection as function of stimulus, which were fitted from various sources in the literature. It can be seen that as the stimulus' temporal envelope changes from impulsive (clicks) to slightly impulsive (speech) and finally to a more moderate one (music), the slope of the threshold curves become less steep. This indicates that thresholds are considerably dependent on the presented stimulus.

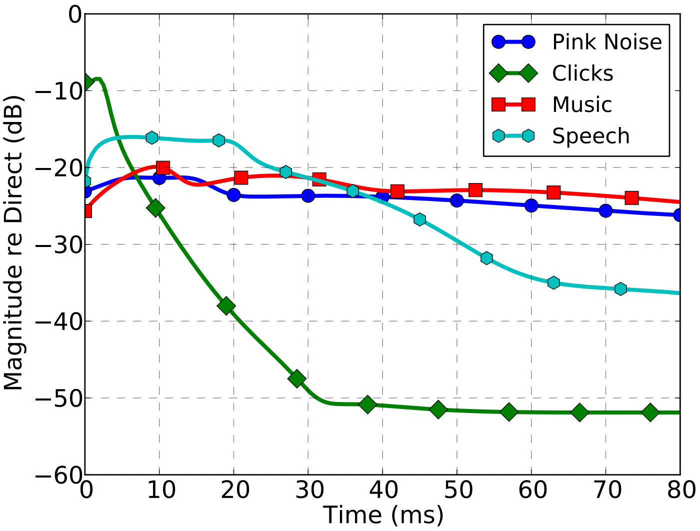


Figure 2.13: Detection thresholds of a single lateral reflection for classical music (squares, data fitted from (Barron, 1971)), pink noise (circles), clicks (diamonds) and speech (hexagons), data fitted from (Olive and Toole, 1989).

Rakerd and Hartmann (1985) studied the effects of a single reflective surface and showed that if directional cues are implausible relative to the acoustic environment, then the precedence effect breaks-down, thus the aforementioned thresholds may no longer be correct. Variables affecting plausibility may be the reflection's time of arrival, frequency spectrum or angle of incidence. When multiple reflections are involved, the localisation process becomes more complex. Very recently, Goupell and colleagues (2012) studied the effect of an additional reflection on the precedence effect. They showed that the precedence effect in a three source paradigm can be seen as combination of the precedence stages seen in the more typical two source situation. This includes summing localisation and localisation dominance, but the interaction with the third reflection may invlove different time scales. However, as it was shown that regardless of time scales, similar localisation mechanisms are involved, it was concluded that even though the two-source paradigm is somewhat of an simplified and idealised scenario, it still provides a strong basis for studying localisation in more complex listening situations.

Olive and Toole (1989) tested reflection audibility thresholds and image shift thresholds at three spatially distinct environments, as shown in Figure 2.14. They used the extreme case of anechoic conditions, an IEC (1985) standard room that may be considered to be a typical listening environment, and a modified version of the IEC room in which reflections were substantially absorbed, resulting in a *relatively reflection free* (RRF) room. It can be clearly seen that the more reverberant the environment is, the higher the audibility threshold, which is especially true for lateral reflections delayed by more than 30ms. It is also evident that in most cases, image-shift thresholds are higher than detection thresholds, meaning that reflections should be substantially higher in level in order to mislead the localisation judgement, than for mere detection. Bech (1998) studied the spatial and timbral perception of individual reflections in small rooms. In his research he used a simulated soundfield comprised of a direct sound, 17 individual reflections and a reverberant field. Bech pointed out that the spectral energy above 2kHz of an individual reflection is fundamental to its spatial effects. This also raises an interesting question regarding the desired frequency response of the desired acoustical treatment for rooms in which localisation is critical.

Summing up this section, it was shown that the perception of reflections in rooms is dependent on many variables. Not only time, level and direction of arrival play an important role, but also the transient nature of the stimulating signal, as well as its spectral characteristics. It can be asserted, therefore, that many different perceptual and cognitive mechanisms are involved in the processing of spatial information. Whilst many underlying neurophysiological processes are still not well understood, it is possible to draw from subjective data and design an objective, psychoacoustically plausible model, describing at least some of these preceptual phenomena.

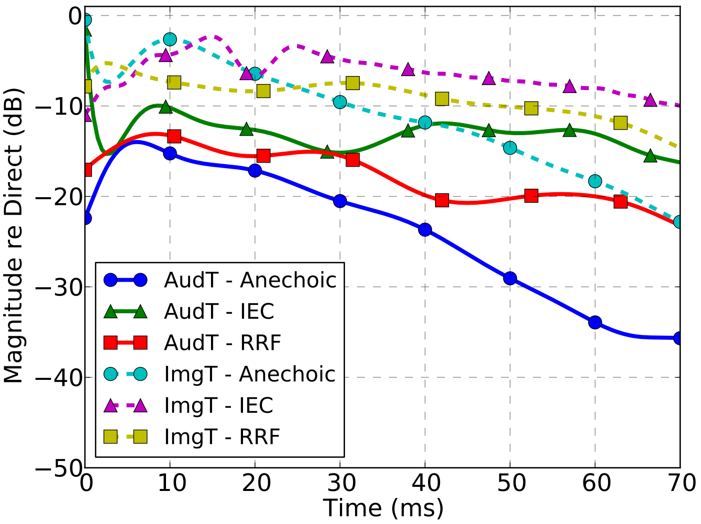


Figure 2.14: Detection thresholds (solid lines, AudT) and image-shift thresholds (dashed lines, ImgT) tested under anechoic conditions (circles), and in an IEC (triangles) and RRF (squares) rooms. data fitted from (Olive and Toole, 1989).

—Nobody believes simulations, except for those who conducted them; everybody believes experimental results, except for those who measured them.

Source Unknown

3

Elements of Computational Modelling

This chapter focuses on methods for modelling room acoustics, and the perception of directional sound in rooms. Section 3.1 provides a general overview of room simulation methods, whereas Section 3.2 elaborates on finite difference methods, which provides the underlying basis for the room model employed in this work. Lastly, common models of binaural sound localisation followed by a discussion on model applications are presented in Section 3.3.

3.1 Room Simulation Methods

Propagation of sound in a room can be seen as a boundary value problem, which may be solved using different mathematical techniques. At the absence of boundary conditions, a closed-form solution is trivial (see Section 2.1). However, in most practical cases of room acoustics, boundary conditions are arbitrary and an analytic solution is difficult if even possible at all. One approach is to make underlying assumptions in order to simplify the problem. Some assumptions may involve the geometry of the domain or its boundary conditions, resulting in a special case of the problem to which *analytic methods* can be applied. Other assumptions may be made about the underlying physics, resulting in a more simplified solution, for example in the case of *geometrical methods.* Another approach is to directly solve the governing PDEs numerically, which can be accomplished by use of *wave methods* in the time or frequency domains.

3.1.1 Analytical Methods

Analytic solutions to problems involving wave behaviour exist in different forms. In room acoustics, one of the most widely used approaches is the *modal decomposition* method for a rectilinear enclosure. When particle velocity on the boundaries is assumed to vanish (rigid boundaries), it is possible to solve the Helmholtz equation in closed-form by means of factorisation and separation of variables. Based on the modal theory of enclosures, the soundfield is then represented in the frequency domain as an infinite sum of contributing modes (Morse, 1948). This is an accurate and efficient way to solve the problem, however, in its traditional form the solution is limited to a rectilinear geometry with rigid boundaries[1], and as such, is not general enough to cover most realistic cases in room acoustics.

3.1.2 Geometrical Methods

Geometrical methods simplify the room acoustics problem by assuming that propagated wavelengths are significantly smaller than their containing space (includ-

---

[1] A solution also exists for damped boundaries.

ing any of its internal objects). By this, the differential problem is simplified and can be solved geometrically. The *image source method* (Allen and Berkley, 1979) considers the response at the receiving position as a sum of propagations directly from the direct source and its mirror images. Whilst the image source concept is an effective simplification of the problem, it suffers from two major issues. First, the number of image sources to be calculated grows exponentially with the order of reflection and with the number of reflecting surfaces in the room. Thus, if one wishes to compute a full impulse response for a geometrically complex structure, then the computational price becomes very high. Secondly, reflections modelled by the image source method are not always consistent with the underlying physics, as it reconstructs their phase function based on the assumption of minimum-phase, which is not a good approximation for absorptive surfaces. Furthermore, the image source method cannot handle diffuse reflections, meaning that scattering surfaces cannot be modelled. For a rectilinear room with perfectly reflecting surfaces the image source model approaches the exact solution to the wave equation, however, whilst such conditions present an idealised case of room acoustics, they are not always realistic.

Another popular geometrical method is the *ray tracing method* (Krokstad et al., 1968; Kulowski, 1985), in which wave propagation is geometrically simplified as an emission of rays, which upon hitting a surface are specularly or diffusely reflected. A receiver is defined as a small volume in the room, which records the time, level and direction of arrival of the different rays passing through it. This is a very computationally efficient approach, however it has drawbacks originating from two fundamental issues. First, the ray propagation assumption is only physically valid for high-frequencies, where wavelength is considerably short relative to the dimensions of the enclosed space. Secondly, as ray tracing is a *monte carlo* method, its accuracy is dependent on the number of rays being emitted. Thus, an accurate result is obtained only for high frequencies and when a large number of rays are used.

Hybrid methods have been introduced as an attempt to combine the advantages of both ray tracing and image source methods (Vorländer, 1989). Drawing from subjective studies indicating that individual early reflections have a stronger perceptual weight than late reflections, the image source method can be used to model the early part of an impulse response, whereas ray tracing can be use to model the remainder of the reverberation. Furthermore, ray tracing can be used in order to perform a *visibility test*, effectively indicating which image sources are visible from the receiving position. Accordingly, the image source method is applied only to these images, and considerable computational power is saved in the process. Nowadays, hybrid geometrical techniques are the most popular methods used in commercial room acoustics modelling software.

3.1.3 Wave Methods

Whilst geometrical methods are very mature in the field of room acoustics simulation, advances in computing technology contribute to the rising popularity of wave based methods, which require a considerable amount of computational power. Wave methods attempt to directly or indirectly solve the acoustic wave equation, and can be generally classified into three families, namely *Finite Element* (FE), *Boundary Element* (BE) and *Finite Difference* (FD) methods.

The finite element method originates from structural analysis (Courant, 1943), however it has also been generalised for problems in acoustics (Ihlenburg, 1998). In an FE-model, the volume is meshed into variously shaped elements each governed by a corresponding equation, essentially transforming a differential problem into an algebraic problem. The derived system of equations can then be solved using various numerical methods. The main advantage of the FE method lies within its remarkable ability to handle very complex geometries and propagation in inhomogeneous media. Its main drawback is that it requires a very fine mesh in order to solve for high frequencies, and as such, the computational efficiency is decreased with the increasing of modelled frequency. This makes the FE-method more suitable for solving problems involving mechanical vibration than general room acoustics, although some cases can be found in the literature, e.g. see (Geddes, 1984; Wright, 1995; Pietrzyk and Kleiner, 1997). More recently, the FE-method has been combined with geometrical methods (Aretz, 2012), which presents a promising approach in context of small room modelling.

The boundary element method does not attempt to solve the wave equation directly. Rather than that, it numerically solves the Kirchhoff-Helmholtz integral equation, which describes the calculated soundfield as a sum of contributions from sound pressures on the boundaries of the modelled space (Ciskowski and Brebbia, 1991). Since BE-method solves for surfaces rather than volumes, it reduces the dimensionality of the problem therefore offering a more efficient solution than FEM, especially for room acoustics where the modelled space is only sparsely filled with reflecting surfaces. Another advantage of the BEM is the ease in which open spaces can be modelled. Whilst BEM is traditionally solved in the frequency domain, some time domain formulations have also been proposed, e.g. see (Hargreaves and Cox, 2008).

Finite difference methods numerically solve differential equations governing sound propagation by means of finite difference approximations. This family of methods, and in particular the FDTD method, are discussed in length in Section 3.2. A closely related method is the *Digital Waveguide Mesh* (DWM), which originates from a sound synthesis model of the human vocal tract (Kelly and Lochbaum, 1962). A digital waveguide is referred to as a discretised version of the d'Alambert solution to the 1D wave equation (see Section 2.1), which can be effectively described as a bidirectional delay line, as shown in Figure 3.1. Thus, as opposed to approximating differential operators with discrete FD operators, an exact solution is directly discretised. Whilst this thesis does not make direct use of digital waveguides, some concepts and formulations discussed in Chapter 5 stem from the DWM, and as such, the method will be briefly overviewed here.

For multidimensional problems, it is possible to intersect two or more waveguides to form a Digital Waveguide Mesh (van Duyne and Smith, 1993). *Scattering Junctions* represent spatial sampling points to which incoming and outgoing waves arrive and are weighted according to the acoustic admittance of the junction. Calculation is performed in a two-pass process. First, the incoming and outgoing pressures are evaluated according to Equation 3.1, a calculation referred

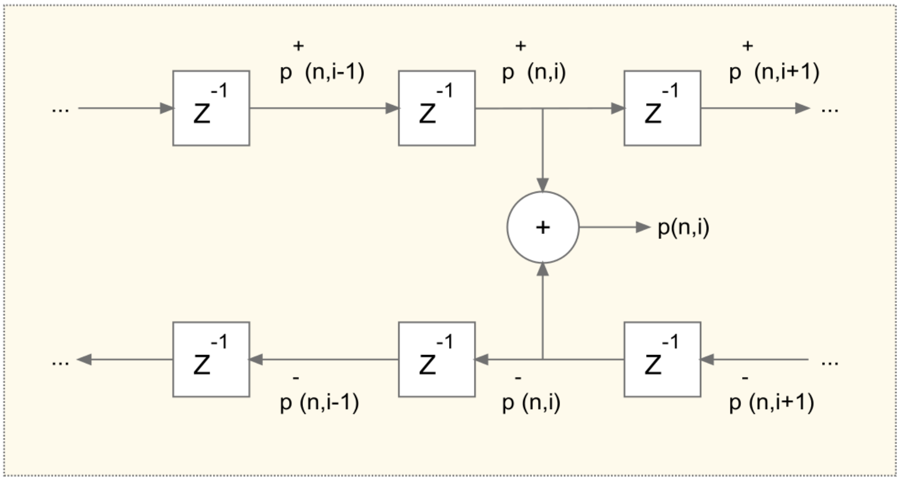


Figure 3.1: Graphical representation of a digital waveguide as a bidirectional delay line. p^+ and p^- represent incoming and outgoing wave components, respectively. Adapted from (Smith III, 1992).

to as the *scattering pass.*

$$p_J = \frac{2}{\sum_{k=1}^{N} Y_k} \sum_{k=1}^{N} Y_k p_k^+ \tag{3.1}$$

where p_J is the pressure on the calculated scattering junction, and Y_k and p_k^+ are the specific acoustic admittances and incoming pressure components at the k^{th} neighbouring junctions[2], respectively. Next, the positive and negative matrices representing pressure values for opposite directions are delayed and shifted. This is referred to as the *delay pass.* For example, a positive-right matrix becomes a negative-left matrix with its x-component delayed by one sample.

All in all, the DWM method provides a versatile and intuitive means for physical modelling of acoustic systems. However, in context of room modelling, its computational costs are disproportionally higher than its benefits, and similar performance can be achieved by the less-demanding FDTD method, which is reviewed in the following section.

---

[2] When waveguides are connected in a 3D rectilinear topology, then each junction has 6 neighbours. If one assumes propagation through air, where $Y_k = 1$ for all k, then Equation 3.1 is reduced to $p_J = \frac{1}{3}\sum_{k=1}^{6} p_k^+$

3.2 Finite Difference Modelling

Finite difference modelling is a family of simple yet powerful methods to solve partial differential equations. In particular, the *Finite Difference Time Domain* (FDTD) method is widely used in computational electrodynamics, and over the past few decades has also gained popularity in modelling acoustic systems, such as rooms (Kowalczyk and van Walstijn, 2011) and musical instruments (Bilbao, 2009). The rationale behind the FDTD methods, is to convert *differential* equations governing a physical system in *continuous* space-time, into *difference* equations which can be then used to obtain a numerical solution for the system in *discrete* space-time. To explain the underlying mathematical basis of the FDTD method, let us first consider an infinitely differentiable univariate function $u(x)$. Taylor series expansions of u about the two points neighboring x, namely $x + \Delta x$ and $x - \Delta x$, are given by

$$u(x \pm \Delta x) = u(x) \pm \Delta x \frac{\partial u(x)}{\partial x} + \frac{\Delta x^2}{2!} \frac{\partial^2 u(x)}{\partial x^2} \pm \frac{\Delta x^3}{3!} \frac{\partial^3 u(x)}{\partial x^3} + ... \tag{3.2}$$

When subtracting the two expansions in 3.2 and dividing by Δx, all terms with even-order derivatives cancel out and the second-order approximation for $\partial u / \partial x$ is given by

$$\frac{\partial u(x)}{\partial x} \approx \frac{u(x + \Delta x) - u(x - \Delta x)}{2\Delta x} + \mathcal{O}(\Delta x^2) \tag{3.3}$$

which is simply a *central difference* approximation for a first derivative. $\mathcal{O}(\cdot)$ denotes the order of truncation which is a result of omitting the higher order terms from the approximation. In this case, it means that approximation errors decay proportionally to Δx^2 as Δx is decreased, or it is simply said that the approximation is *second-order accurate.* Obtaining an approximation for the second-derivative of $u(x)$ is accomplished by summing the two series expansions in 3.2. Now, all terms with odd-order dervatives cancel out, resulting in

$$u(x + \Delta x) + u(x - \Delta x) = 2u(x) + \Delta x^2 \frac{\partial^2 u(x)}{\partial x^2} + \frac{2\Delta x^4}{4!} \frac{\partial^4 u(x)}{\partial x^4} + ... \tag{3.4}$$

which can be rearranged to show that

$$\frac{\partial^2 u(x)}{\partial x^2} \approx \frac{u(x+\Delta x) - 2u(x) + u(x-\Delta x)}{\Delta x^2} + \mathcal{O}(\Delta x^2) \tag{3.5}$$

Considering now a multivariate function u($\mathbf{x}$,t) describing a quantity changing in continuous time t, and continuous space $\mathbf{x} = (x, y, z) \in \mathbb{R}^3$, one can produce equivalent approximations in discrete time n and discrete space $\mathbf{i} = [l, m, i]$ by sampling the function such that

$$u\big|_{\mathbf{i}}^{n} = u(\mathbf{x}, t)\big|_{x=lX, y=mX, z=iX, t=nT} \tag{3.6}$$

where X is the spatial period (assumed to be uniform at all directions), and T is the temporal period related to the sample rate $f_s = 1/T$. In essence, these quantities represent the spatio-temporal increments, which are equivalent to Δx in the aforementioned univariate function $u(x)$. Accordingly, incrementing by T or X in the continuous domain is equivalent to a shift operation in the discrete domain, e.g. $u(\mathbf{x}, t+T) \equiv u\big|_{\mathbf{i}}^{n+1}$.

Following the conventions suggested in (Bilbao, 2009), we define *finite difference operators* which represent discrete approximations to partial derivatives. For example, the first and second order time derivatives of $u(\mathbf{x}, t)$ approximated by central differences, are denoted

$$\delta_{t\cdot} u\big|_{\mathbf{i}}^{n} \triangleq \frac{1}{2T}\left(u\big|_{\mathbf{i}}^{n+1} - u\big|_{\mathbf{i}}^{n-1}\right) \tag{3.7}$$

and

$$\delta_{t\cdot}^2 u\big|_{\mathbf{i}}^{n} \triangleq \frac{1}{T^2}\left(u\big|_{\mathbf{i}}^{n+1} - 2u\big|_{\mathbf{i}}^{n} + u\big|_{\mathbf{i}}^{n-1}\right) \tag{3.8}$$

3.2.1 Classical Yee Method

The roots of the FDTD method date back to Yee (Yee, 1966) which suggested a time stepping numerical scheme for solving Maxwell's equations. In acoustics modelling, a similar scheme is used to solve Euler's equations on two staggered grids for pressure and particle velocity (Botteldooren, 1995), as shown in Figure

3.2. Using finite difference operators, the conservation of momentum is approximated as

$$\delta_t p\big|_{\mathbf{i}}^{n} = -\rho_0 c\lambda \left(\delta_x u_x\big|_{\mathbf{i}}^{n} + \delta_y u_y\big|_{\mathbf{i}}^{n} + \delta_z u_z\big|_{\mathbf{i}}^{n}\right) \tag{3.9}$$

where u_x, u_y and u_z denote the orthogonal components of the particle velocity vector $\mathbf{u}$ in a Cartesian coordinate system. Similarly, the conservation of mass is approximated as

$$\delta_t u_x\big|_{\mathbf{i}}^{n} = -\frac{\lambda}{\rho_0 c}\delta_x p\big|_{\mathbf{i}}^{n} \tag{3.10a}$$

$$\delta_t u_y\big|_{\mathbf{i}}^{n} = -\frac{\lambda}{\rho_0 c}\delta_y p\big|_{\mathbf{i}}^{n} \tag{3.10b}$$

$$\delta_t u_z\big|_{\mathbf{i}}^{n} = -\frac{\lambda}{\rho_0 c}\delta_z p\big|_{\mathbf{i}}^{n} \tag{3.10c}$$

where p denotes acoustic pressure. The variable λ, also known as the *Courant number*, is used to relate the sampling periods in space and time according to the *Courant-Freidrichs-Lewy criterion* (Strikwerda, 2004)

$$\lambda = \frac{cT}{X} \tag{3.11}$$

As will be discussed in Section 3.2.5, the choice of λ is fundamental to the stability of the scheme. The finite difference operators are developed under the definition of two grids staggerred in time and space. Thus, spacing between particle velocity and pressure points on the grid correspond to half of a single step[3], and the corresponding operators are given by

$$\delta_t \mathbf{u}\big|_{\mathbf{i}}^{n} \equiv \mathbf{u}\big|_{\mathbf{i}}^{n+\frac{1}{2}} - \mathbf{u}\big|_{\mathbf{i}}^{n-\frac{1}{2}} \qquad \delta_t p\big|_{\mathbf{i}}^{n} \equiv p\big|_{\mathbf{i}}^{n+1} - p\big|_{\mathbf{i}}^{n} \tag{3.12a}$$

$$\delta_x \mathbf{u}\big|_{\mathbf{i}}^{n} \equiv \mathbf{u}\big|_{l+\frac{1}{2},m,i}^{n+\frac{1}{2}} - \mathbf{u}\big|_{l-\frac{1}{2},m,i}^{n+\frac{1}{2}} \qquad \delta_x p\big|_{\mathbf{i}}^{n} \equiv p\big|_{l+1,m,i}^{n} - p\big|_{l,m,i}^{n} \tag{3.12b}$$

$$\delta_y \mathbf{u}\big|_{\mathbf{i}}^{n} \equiv \mathbf{u}\big|_{l,m+\frac{1}{2},i}^{n+\frac{1}{2}} - \mathbf{u}\big|_{l,m-\frac{1}{2},i}^{n+\frac{1}{2}} \qquad \delta_y p\big|_{\mathbf{i}}^{n} \equiv p\big|_{l,m+1,i}^{n} - p\big|_{l,m,i}^{n} \tag{3.12c}$$

$$\delta_z \mathbf{u}\big|_{\mathbf{i}}^{n} \equiv \mathbf{u}\big|_{l,m,i+\frac{1}{2}}^{n+\frac{1}{2}} - \mathbf{u}\big|_{l,m,i-\frac{1}{2}}^{n+\frac{1}{2}} \qquad \delta_z p\big|_{\mathbf{i}}^{n} \equiv p\big|_{l,m,i+1}^{n} - p\big|_{l,m,i}^{n} \tag{3.12d}$$

Note that here the finite difference operators are *non-normalised*, as the required

---

[3] Although programatically one performs shift operations by a single increment/decrement, as non-integer indices are not possible.

division by the corresponding temporal or spatial periods is already expressed as coefficients in Equations 3.9 and 4.4. To implement the scheme in a computer program, one should iteratively solve Equations 3.9 and 4.4 for $p\big|_{\mathbf{i}}^{n+1}$ and $\mathbf{u}\big|_{\mathbf{i}}^{n+1}$ respectively.

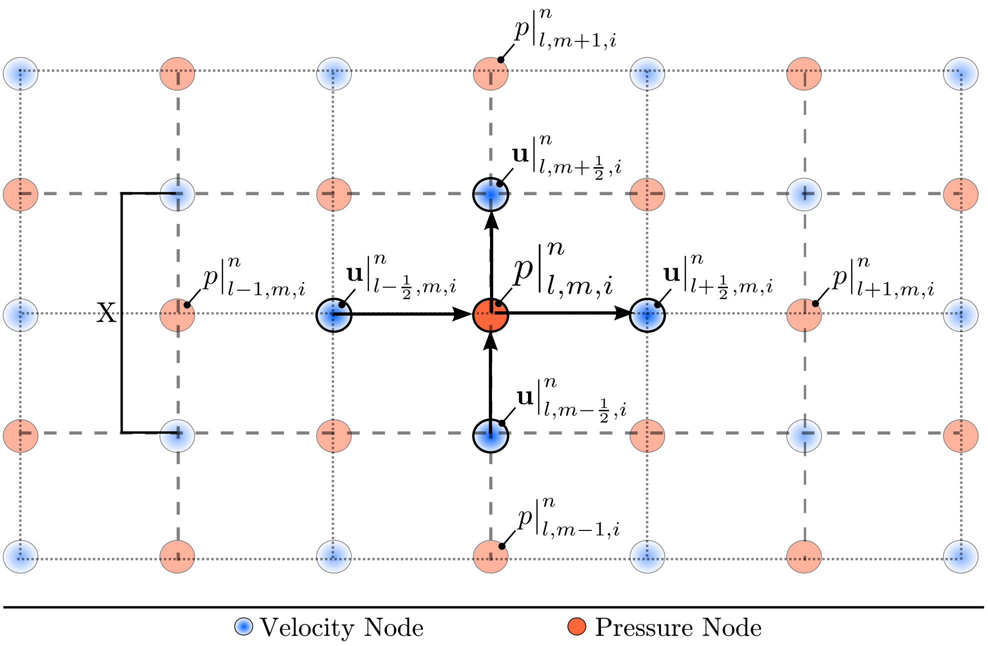


Figure 3.2: Staggerred grids for pressure (red-filled dots, p) and particle velocity (blue-gradient dots, $\mathbf{u}$), illustrated at discrete time n.

3.2.2 Compact Explicit Schemes for the Wave Equation

Whilst the Yee scheme is most widely known, it is by no means the most efficient solution. In fact, if knowledge of particle velocity is not required, one may employ a finite difference scheme approximating the scalar wave equation. In such case, the problem can be solved on a single grid only, which presents significant processing and memory savings (Kowalczyk and van Walstijn, 2011). In the numerical domain, the wave equation is modelled as

$$\left(\delta_t^2 - \lambda^2 \delta_{\mathbf{x}}^2\right) p\big|_{\mathbf{i}}^n = 0 \tag{3.13}$$

where the finite difference operators are given by

$$\delta_t^2 p_{\mathbf{i}}^n \equiv p\Big|_{\mathbf{i}}^{n+1} - 2p\Big|_{\mathbf{i}}^{n} + p\Big|_{\mathbf{i}}^{n-1} \tag{3.14a}$$

$$\delta_x^2 p_{\mathbf{i}}^n \equiv p\Big|_{l+1,m,i}^{n} - 2p\Big|_{l,m,i}^{n} + p\Big|_{l-1,m,i}^{n} \tag{3.14b}$$

$$\delta_y^2 p_{\mathbf{i}}^n \equiv p\Big|_{l,m+1,i}^{n} - 2p\Big|_{l,m,i}^{n} + p\Big|_{l,m-1,i}^{n} \tag{3.14c}$$

$$\delta_z^2 p_{\mathbf{i}}^n \equiv p\Big|_{l,m,i+1}^{n} - 2p\Big|_{l,m,i}^{n} + p\Big|_{l,m,i-1}^{n} \tag{3.14d}$$

and $\delta_{\mathbf{x}}^2$ represents the sum of the finite difference approximations in the three axial directions,

$$\delta_{\mathbf{x}}^2 = \delta_x^2 + \delta_y^2 + \delta_z^2 \tag{3.15}$$

As will be discussed in Section 3.2.6, the rectilinear nature of the scheme, results in direction and frequency dependent numerical dispersion which is detrimental to the solution. One option to control the amount of dispersion induced by the scheme, is to employ spatial interpolation (Savioja and Valimaki, 2000). Kowalczyk and Van Walstijn (2011) suggested a family of compact explicit numerical schemes for the wave equation, which are very suitable for room acoustics modelling. Following their approach, the wave equation can be modelled by Equation 3.13, but with the spatial finite difference operator modified to account not only for axially neighbouring nodes, but also for side-diagonal and diagonal nodes, as shown in Figure 3.3. Accordingly, $\delta_{\mathbf{x}}^2$ becomes

$$\delta_{\mathbf{x}}^2 = \delta_x^2 + \delta_y^2 + \delta_z^2 + a\left(\delta_x^2\delta_y^2 + \delta_x^2\delta_z^2 + \delta_y^2\delta_z^2\right) + b\delta_x^2\delta_y^2\delta_z^2 \tag{3.16}$$

where the free parameters a and b are chosen according to the desired numerical scheme, as shown in Table 3.1.

By applying the finite difference operators in 3.14 to the discrete wave equation 3.13, it is straightforward to show that the explicit update equation for an air node becomes

$$p\Big|_{l,m,i}^{n+1} = d_4 p\Big|_{l,m,i}^{n} - p\Big|_{l,m,i}^{n-1} + d_1 \underbrace{\sum p\Big|_a}_{axial} + d_2 \underbrace{\sum p\Big|_{sd}}_{side\text{-}diagonal} + d_3 \underbrace{\sum p\Big|_d}_{diagonal} \tag{3.17}$$

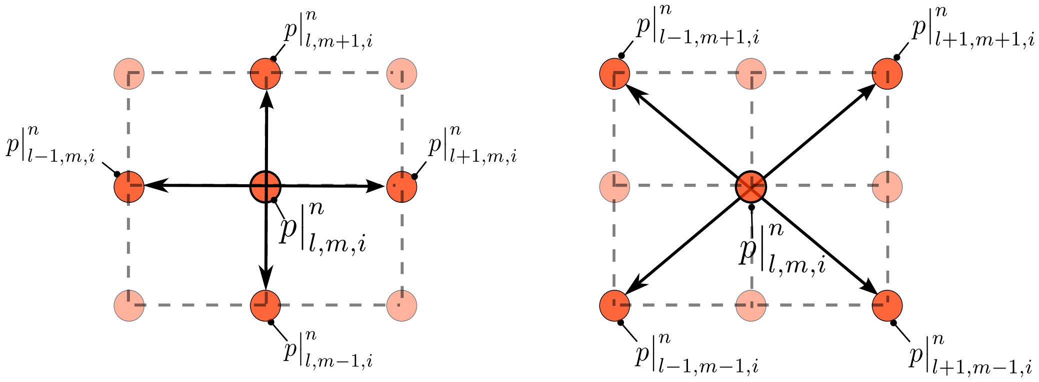


Figure 3.3: Illustration of spatial interpolation within numerical schemes for the wave equation. Left pane: axial node connections. Right pane: side-diagonal node connections along the i-plane. Similar side-diagonal connections exist along the l and m planes, as well as a full-diagonal connection. For a detailed stencil diagram, see e.g. (Kowalczyk and van Walstijn 2011, Figure 1).

Scheme	a	b	λ	accuracy	isotropy	points
Standard Rectilinear (SRL)	0	0	$\sqrt{\frac{1}{3}}$	$0.075 f_s$	$0.075 f_s$	7
Interpolated Wideband (IWB)	$\frac{1}{4}$	$\frac{1}{16}$	1	$0.186 f_s$	$0.186 f_s$	27

Table 3.1: Design parameters for two commonly-used numerical schemes. Accuracy and isotropy will be discussed in detail in Section 3.2.6. For a full list of numerical schemes and their parameters, readers are referred to (Kowalczyk and van Walstijn, 2011).

where $\sum p\big|_a$, $\sum p\big|_{sd}$ and $\sum p\big|_d$ are the sums of pressures of the neighbouring nodes at the axial, side-diagonal and diagonal directions, respectively, and the coefficients d_1, d_2, d_3 and d_4 are given by (Kowalczyk and van Walstijn, 2011)

$$\begin{aligned} d_1 &= \lambda^2(1-4a+4b) \\ d_2 &= \lambda^2(a-2b) \\ d_3 &= \lambda^2 b \\ d_4 &= 2+\lambda^2\left(12a-8b-6\right) \end{aligned} \tag{3.18}$$

An interesting point in case is that if one chooses the grid parameters for a standard rectilinear (SRL) scheme, then Equation 3.17 can be shown to be equivalent to its DWM counterpart, that is, Equation 3.1 and its following delay pass (Smith III, 2004; Karjalainen and Erkut, 2004).

3.2.3 Grid Excitation

In FDTD simulation, the grid can be excited either by employing a time-varying source on a single grid node, e.g. in (Schneider et al., 1998a), or by pre-determining the pressure distribution across the entire domain, e.g. in (Sakamoto et al., 2002). In this work we are only concerned with the former, as we wish to investigate the build up of the soundfield, rather than to analyse its steady state. When exciting the grid with a time-varying source, two issues are considered. First, the desired source function should be identified. The choice of this function is normally constrained by the mechanical behaviour of the modelled source as well as by the properties of the numerical scheme. Secondly, the chosen source function should be injected into the grid, which can be accomplished using different methods. This topic is covered in detailed on Chapter 4, and as such, shall not be discussed here any further.

3.2.4 Boundary Models

At the boundaries of the domain the grid should be truncated. If the domain is assumed to be unbounded, then *absorbing boundary conditions* (ABCs) should be employed in order to effectively emulate an anechoic environment. ABCs have long been a prominent topic of research, and perhaps the most notable example is the *perfectly matched layer* (PML) originally suggested in (Berenger, 1994). Formulations based on other physical or mathematical assumptions can also be found in the literature, e.g. in (Enquist and Majda, 1977; Yuan et al., 1997; Murphy and Mullen, 2002). An alternative approach to employing ABCs, is to create a pseudo-anechoic environment, by terminating the simulation before any reflections arrive at the observation position. Whilst this approach is straightforward, it requires a rather large domain and is not suitable when long excitation signals are used. Nevertheless, if the excitation signal is compact in time, then this is a simple and computationally efficient approach for achieving an anechoic simulation. In this study we are mostly concerned with impulse response measurements, and as such, the pseudo-anechoic approach is taken whenever an unbounded domain is simulated.

In the case of a room model, one normally assumes that at least *some* energy is reflected off the boundaries. As such, boundary models which adhere to LRS theory (see Chapter 2) are a good choice. Kowalczyk and van Walstijn (2011) suggest such a boundary model for the compact explicit schemes. In the case of non-conformal[4], 3D boundary methods the geometry is sampled and quantised in space, resulting in a *stair-case* representation of the surface. Each boundary node could then be part of one of 6 possible *faces*, 12 *inner-edges*, 12 *outer-edges*, 8 *inner-corners* or 8 *re-entrant corners.* For illustrative purposes, three types of boundary nodes are exemplified Figure 3.4. Each of these boundary nodes needs to satisfy a different boundary condition, which in a discrete-time formulation results in different types of boundary update equations. Here, the boundary model is exemplified by considering the case of a right-face. For an elaborate representation of different types of boundaries, readers are referred to (Kowalczyk and van Walstijn, 2011) .

If the boundary impedance ξ_w is constant across frequency, then arriving at an

---

[4] Whilst conformal methods have been developed for FDTD models (Tolan and Schneider, 2003), they are considerably more complex and computationally demanding.

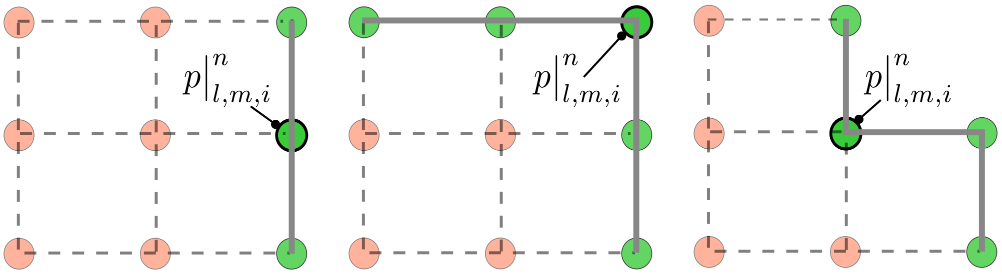


Figure 3.4: Example of different types of boundary nodes. Left - a right 2D edge (or 3D face). Middle - a right-top 2D inner-corner (or 3D inner-edge). Right - a bottom-left 2D re-entrant corner (or 3D re-entrant edge). For illustrations of other types of boundary nodes, readers are referred to (Kowalczyk and van Walstijn, 2011)

update equation for a right-face can be simply accomplished by approximating Equation 2.17 using finite differences. However, recall from Section 2.2 that boundary conditions, in most realistic cases of surface absorption, are frequency-dependent. To solve this, the boundary impedance $\xi_w(\omega)$ can be represented using a *digital impedance filter* (DIF) of the form

$$\xi_w(z) = \frac{b_0 + \sum_{k=1}^{N} b_k z^{-k}}{a_0 + \sum_{k=1}^{N} a_k z^{-k}} \tag{3.19}$$

The update equation for a right-face is then given by (Kowalczyk and van Walstijn, 2011)

$$\begin{aligned} p|^{n+1}_{l,m,i} = A_0 \Big[& d_1 \sum p_a + d_2 \sum p_{sd} + d_3 \sum p_d + d_4 p|^{n}_{l,m,i} \\ & + \frac{\lambda^2}{b_0} g|^{n} + \left(\frac{\lambda a_0}{b_0} p|^{n-1}_{l,m,i} \right) \Big] \end{aligned} \tag{3.20}$$

where

$$A_0 = \frac{1}{1 + \frac{\lambda a_0}{b_0}} \tag{3.21}$$

In this case, the sums of the axial, side-diagonal and diagonal neighbouring nodes are solved such that nodes lying within the boundaries are algebraically eliminated. Accordingly, these quantities are given by

$$\sum p_a = 2p\Big|_{l-1,m,i}^{n} + p\Big|_{l,m+1,i}^{n} + p\Big|_{l,m-1,i}^{n} + p\Big|_{l,m,i+1}^{n} + p\Big|_{l,m,i-1}^{n} \tag{3.22a}$$

$$\begin{aligned}\sum p_{sd} = {} & 2\left(p\Big|_{l-1,m+1,i}^{n} + p\Big|_{l-1,m-1,i}^{n} + p\Big|_{l-1,m,i+1}^{n} + p\Big|_{l-1,m,i-1}^{n}\right) \\ & + \left(p\Big|_{l,m+1,i+1}^{n} + p\Big|_{l,m+1,i-1}^{n} + p\Big|_{l,m-1,i+1}^{n} + p\Big|_{l,m-1,i-1}^{n}\right)\end{aligned} \tag{3.22b}$$

$$\sum p_d = 2\left(p\Big|_{l-1,m+1,i+1}^{n} + p\Big|_{l-1,m-1,i+1}^{n} + p\Big|_{l-1,m+1,i-1}^{n} + p\Big|_{l-1,m-1,i-1}^{n}\right) \tag{3.22c}$$

In the time domain, the DIF approximating the frequency dependent boundary is solved using the intermediate functions

$$x\Big|^{n} = \frac{a_0}{\lambda b_0}\left(p\Big|_{l,m,i}^{n+1} - p\Big|_{l,m,i}^{n-1}\right) - \frac{g\Big|^{n}}{b_0} \tag{3.23a}$$

$$y\Big|^{n} = \frac{1}{a_0} b_0 x\Big|^{n} + g\Big|^{n} \tag{3.23b}$$

$$g\Big|^{n} = \sum_{k=1}^{N} b_k x\Big|^{n-k} - a_k y\Big|^{n-k} \tag{3.23c}$$

where $x\Big|^{n}$ and $y\Big|^{n}$ represent the digital impedance filter's inputs and outputs. The approximation using a digital impedance filter is not directly used in this thesis, however, a frequency independent case is derived from it in Chapter 5.

3.2.5 Errors and Stability

In finite difference modelling, numerical artefacts can be traced back to two sources of error, namely *precision* (or round-off) and *truncation* errors. Precision accuracy is strictly dependent on the capabilities of the computing machine, which typically corresponds to either *single* or *double* precision arithmetic. As will be further discussed in Chapter 5, simulations in this thesis are executed on a general purpose GPU which has a limited amount of memory. Accordingly, it has been opted to employ single precision arithmetic, which allow larger systems to be modelled at the expense of compromising precision accuracy. Truncation errors occur due to discretisation of the problem in time and space. Recall from Equations 3.3 and 3.5 that infinite series expansions are truncated in order to arrive at finite difference approximations of the derivatives. As the spatial and temporal periods X and T are decreased, so do their associated truncation errors. If at the limit, where $(T, X \to 0)$, the finite difference scheme aproaches the values of the modelled differential equation, then the scheme is said to be *consistent*. However, according to the *Lax-Richtmyer Equivalence Theorem*, in order for a consistent scheme[5] to be regarded to as *convergent*, it must also be *numerically stable* (Strikwerda, 2004).

For a numerical scheme to be considered stable, it should not allow any growing solutions to occur. Observing that the FDTD update equations are solved *recursively*, it can be concluded that if the digital filter approximating the PDE is at least *marginally stable*, then no growing solutions will occur. The FDTD update equations cannot be checked for stability in their native form, as recursive relationships depend on both time *and* space. Nevertheless, investigating the stability criterion of a numerical scheme can be accomplished by employing a *von Neumann analysis* (Strikwerda, 2004), effectively revealing which values of λ will cause a growing solution. In a von-Neumann analysis, one performs a Fourier transform on all spatial dimensions, which then reduces the multidimensional scheme into a digital filter whose recursive properties depend only on the time dimension. In other words, spatial finite differences are transformed into *k-space*, whereas temporal finite differences are left in the time domain. Accordingly, if the resulting digital filter is at least marginally-stable for all spatial frequencies, then the numerical scheme is said to be stable. As this dissertation is not concerned with the development of finite difference schemes, the entire mathematical process is not shown here. Examples of stability analyses for common numerical schemes can be found in the literature, e.g. see (Strikwerda, 2004; Bilbao, 2004). In context of the compact explicit schemes used in this thesis, a stability analysis shows that the constraint on choosing λ is (Kowalczyk and van Walstijn, 2011)

$$\lambda^2 \leq \min\left(1, \frac{1}{2-4a}, \frac{1}{3-12a+16b}\right) \tag{3.24}$$

where a and b are defined in Table 3.1.

---

[5] Assuming of course, that it is used to solve a *well-posed* initial-value problem.

3.2.6 Numerical Dispersion

When sound waves propagate in a non-dispersive medium such as air, their propagation velocity is isotropic and independent of frequency. However, modelling such a system using FDTD induces numerical artefacts, namely that phase velocities of waves in the numerical domain, deviate from their theoretical counterparts. This deviation is dependent on frequency as well as on the direction of propagation, and is referred to in the literature as *numerical dispersion* (Bilbao, 2004). In general, numerical dispersion becomes more dominant with increasing frequency, and as such, imposes a limitation on the bandwidth of the modelled system. As both bandwidth and isotropy are of importance to modelling sound localisation, a mathematical analysis of dispersion is offered in this section. For simplicity, we focus on the special case of the SRL scheme. A generalisation for all compact explicit schemes is found in (Kowalczyk and van Walstijn, 2011).

Let us consider a multidimensional z-transform such that the variables in the time and z-domains are

$$(n, l, m, i) \xrightarrow{\text{Z}} (z_t, z_x, z_y, z_z) \tag{3.25}$$

By the z-transform shift theorem, the finite difference operators defined in Equation 3.14 can now be rewritten as

$$D_t^2 p\big|_{\mathbf{i}}^n \equiv (z_t - 2 + z_t^{-1})p\big|_{\mathbf{i}}^n \tag{3.26a}$$

$$D_x^2 p\big|_{\mathbf{i}}^n \equiv (z_x - 2 + z_x^{-1})p\big|_{\mathbf{i}}^n \tag{3.26b}$$

$$D_y^2 p\big|_{\mathbf{i}}^n \equiv (z_y - 2 + z_y^{-1})p\big|_{\mathbf{i}}^n \tag{3.26c}$$

$$D_z^2 p\big|_{\mathbf{i}}^n \equiv (z_z - 2 + z_z^{-1})p\big|_{\mathbf{i}}^n \tag{3.26d}$$

where the complex frequency variables are $z_t = e^{j\omega T}$, $z_x = e^{j\hat{\beta}_x X}$, $z_y = e^{j\hat{\beta}_y X}$, $z_z = e^{j\hat{\beta}_z X}$ and $\hat{\beta}$ is the numerical wavenumber given by

$$\hat{\beta}^2 = \hat{\beta}_x^2 + \hat{\beta}_y^2 + \hat{\beta}_z^2 \tag{3.27}$$

Accordingly, the discrete wave equation 3.13 can be written as

$$D_t^2 p\big|_{\mathbf{i}}^n = \lambda^2 \left(D_x^2 + D_y^2 + D_z^2\right) p\big|_{\mathbf{i}}^n \tag{3.28}$$

Noting that

$$D_t^2 = \left(e^{j\omega T} - 2 + e^{-j\omega T}\right) = -4\sin^2\left(\frac{\omega T}{2}\right) \tag{3.29}$$

and similarly

$$D_x^2 = -4\sin^2\left(\frac{\hat{\beta}_x X}{2}\right) \tag{3.30a}$$

$$D_y^2 = -4\sin^2\left(\frac{\hat{\beta}_y X}{2}\right) \tag{3.30b}$$

$$D_z^2 = -4\sin^2\left(\frac{\hat{\beta}_z X}{2}\right) \tag{3.30c}$$

then by direct substitution into Equation 3.28 we obtain

$$\sin^2\left(\frac{\omega T}{2}\right) = \lambda^2\left[\sin^2\left(\frac{\hat{\beta}_x X}{2}\right) + \sin^2\left(\frac{\hat{\beta}_y X}{2}\right) + \sin^2\left(\frac{\hat{\beta}_z X}{2}\right)\right] \tag{3.31}$$

which is the well known dispersion relation for a rectilinear grid. This expression is fundamental to the analysis of numerical dispersion, as it interrelates the grid parameters to temporal and spatial frequencies. For example, if one wishes to study the directional effects of dispersion, it is common to solve Equation 3.31 for ω and then calculate the ratio of the numerical phase velocity $\hat{c}$ to the theoretical phase velocity c, by

$$v_{err}(\hat{\beta}_x, \hat{\beta}_y, \hat{\beta}_z) = \frac{\hat{c}}{c} = \frac{\omega}{\hat{\beta} c} \tag{3.32}$$

A dispersion plot for a section plane of the 3D rectilinear grid is shown in Figure 3.5. As discussed in Section 2.4, ITD cues play an important role in localisation of sound on the horizontal plane. The question arises, whether the anisotropic characteristics of numerical dispersion could corrupt ITDs in such a way that they will mislead a localisation judgement. To answer this question, let us consider the isotropy of different numerical schemes. The dispersion relation shown

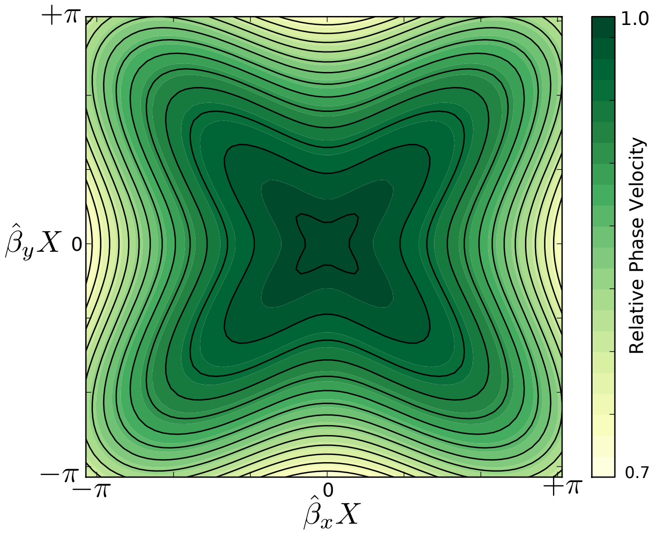


Figure 3.5: Dispersion error, v_{err} (ratio of numerical to theoretical phase velocities) calculated for the $x-y$ plane (i.e. $\hat{\beta}_z=0$) of a standard rectilinear scheme. Note that errors are lowest for the diagonal directions (small decrease in relative phase velocity with increasing of frequency), and highest for the axial directions.

in Equation 3.31 can be rearranged to calculate the relative phase velocity as function of frequency for a predefined propagation angle (i.e a predefined combination of $\hat{\beta}_x, \hat{\beta}_y, \hat{\beta}_z$). Kowalczyk and Van Walstijn (2011) calculated this function for different angles of propagation[6] and defined a measure of *isotropy error* as the difference between the phase velocities at the "best" and "worst" directions[7]. Isotropy errors for the SRL and IWB schemes are shown in Figure 3.6.

As it is not exactly clear what is the minimum isotropy error required for the numerical deviation in ITD to cause a noticeable shift in source locus, here we adopt a safe approach. Figure 3.6 indicates that the SRL and IWB schemes are almost entirely isotropic up to $0.015f_s$ and $0.3f_s$ respectively (observe the frequencies at which the curves begin to deviate from zero). Thus, if one assumes that the auditory system relies on interaural phase differences below $\sim$ 700Hz (see Section 2.4), then one requires a minimum sample rate of $\sim$ 46kHz (SRL) or $\sim$ 23kHz (IWB) in order to ensure that anisotropy does not affect localisation judgement. Note that this is a 'worst-case' analysis, as isotropy error is normally measured between two extreme directions, whereas considering the size of a hu-

---

[6] For the complete mathematical procedure see (van Walstijn and Kowalczyk, 2008)

[7] By "best" and "worst" we refer to the directions in which the highest and lowest amount of change in relative phase velocity occur over frequency. These are dependent on the numerical scheme being used.

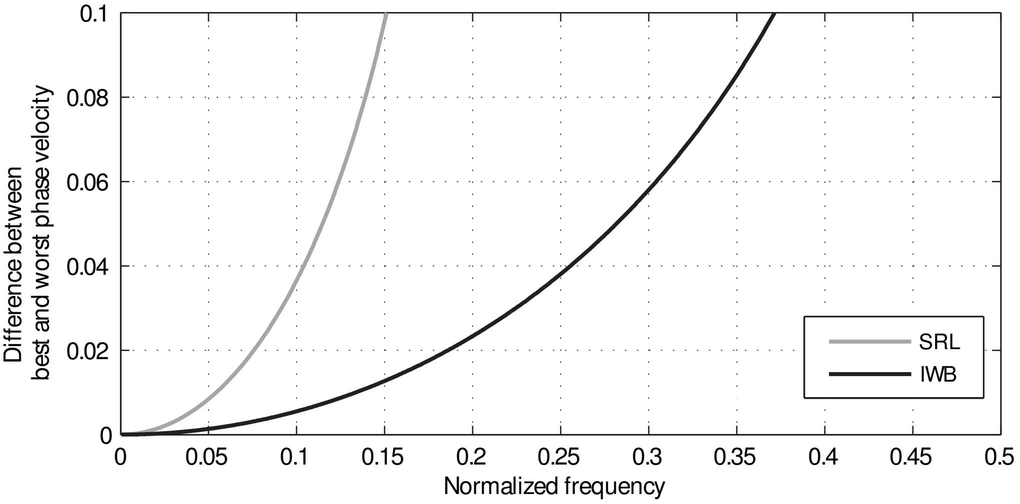


Figure 3.6: Isotropy error for the SRL and IWB schemes. Modified from (Kowalczyk and van Walstijn, 2011). Used with kind permission from the authors.

man head, the difference in propagation angles from the source to the left and right ears is rather minimal. It should be borne in mind that numerical dispersion accumulates as the wave travels a distance, meaning that the greater the distance between the source and the receiver, the more pronounced will be the numerical dispersion errors. Yet, the limits of isotropy discussed here ensure that even for large distances, the accumulated dispersion will be independent of the direction of propagation, and as such, should not affect interaural phase differences.

3.3 Models of Sound Localisation

Modelling sound localisation is one of the foremost challenges in the field of computational auditory scene analysis, with direct impact on research in robotics, architectural acoustics and auditory neuroscience, amongst others. Accordingly, a vast amount of knowledge can be found in the literature, and in order to keep this review as focused as possible, we define three main criteria for models which are relevant to this work. First, the model must account for sound localisation by human listeners, meaning that it should not necessarily be as accurate as possible, but to be able to faithfully reproduce choices similar to those that would have been made by real human subjects. Secondly, focus is given to models which are based on binaural processing, as the goal of this work is to provide a framework for modelling localisation in the horizontal plane. Lastly, we focus on *functional* models which are applicable to the field of architectural acoustics. In other words, whilst there are models designed to study very specific binaural processes, here we are interested in models which have a clear potential to be generalised to a wider range of problems involving localisation in enclosed spaces.

3.3.1 Fundamental Modelling Concepts

Models of binaural localisation can be classified according to their underlying principles. *Physiologically motivated* methods generally draw from neurophysiological findings, and may either simulate neural networks directly or rely on mathematical descriptions of biologically-inspired processes. Another approach is to use *analytically motivated* expressions whose only goal is to predict the results of subjective tests. This does not mean that such expressions are merely a mathematical fit to the data, but that they do not make any underlying assumptions based-on or even related-to physiological knowledge. Thus, analytically-motived methods are *psychoacoustically plausible*, whereas physiologically-motived methods are also *physiologically plausible*. This section provides a general literature review of different strategies for modelling binaural sound localisation.

Jeffress (1948) published a landmark paper in which he described a method for sound localisation based on ITD cues. The paper is considered remarkable, especially when recognising the little amount of knowledge that was available at the time that it was published. An extensive review of its physiological basis is available in (Yin, 2002). Jeffress made three underlying assumptions. First, he assumed that input signals from the peripheral auditory system carry reliable binaural timing information. Secondly, he assumed that the auditory system employs an array of binaural detection cells (equivalent to MSO neurons, see Section 2.3) which have a maximum activity when input signals coincide. Lastly, he assumed that the bilateral neural pathways conveying information to the central nervous system are formed in opposing directions. From a signal processing perspective this can be seen as a bidirectional delay line, whose intersections correspond to coincidence detectors. Since processing is performed in separate frequency channels, then multiple delay lines exist, each corresponding to data in a different critical band. According to Jeffress, the place within this network in which maximal activity occurs is inherently dependent on the ITD and corresponds to the perceived location of the sound source. From a mathematical point of view, this is equivalent to calculating a running cross-correlation function between the ipsilateral and contralateral signals at a frequency band of interest (Joris et al., 1998).

Lindemann (1986a; 1986b) extended Jeffress' model by adding contralateral inhibitory elements to it. He assumed that the two signals in a bidirectional delay line inhibit one another, which resembles physiological processes of ILD encoding (see Section 2.3). In turn, this accounts for time-intensity trading mechanisms exhibited by the human auditory system. Gaik (1993) adjusted this mechanism by introducing additional weighting functions which simulate naturally occurring time-intensity combinations, thus improving the psychoacoustical plausibility of the model. Another important benefit of the contralateral inhibition process concerns the ability to perceive lead-lag type stimuli. In (Lindemann, 1986b), it is shown that the process leads to emphasis of the lead sound, which implies that the model can account for the precedence effect. A modelling strategy founded on coincidence detection, with all of its extensions, is often referred to as a *Jeffress-Lindemann-Gaik* model.

Breebaart and colleagues (2001) suggested a different approach to model binaural hearing. Rather than relying on cross-correlation algorithms, they based their model on *Equalisation-Cancellation* (EC) theory (Durlach, 1963) and the physiological findings of Reed and Blum (1990). They developed a conceptual type of nerve cells, combining ITD coincidence detectors (similar to those of a Jeffress model) with ILD processors inspired by the findings of Reed and Blum. According to EC-theory, the auditory system eliminates binaural masking components in a two pass process. First, spectral and temporal transformations are applied to input stimuli in an attempt to equalise the masking components (equalisation, or E-process). Then, one stimulus is subtracted from the other and the identical components are removed (cancellation, or C-process). Although binaural masking is not the main concern of this study, the EC-process implies a type of auditory signal processing approach directly relevant to sound localisation. As it is assumed that EC is accomplished by a contralateral inhibition mechanism, Breebaart and colleagues partially based their model on EI-type neurons found in the LSO and IC. The overall output of this model is somewhat complex, as analysis is performed in different critical bands at different instants in time. Therefore, for a single frequency band this results in a two-dimensional time varying activity pattern, describing the change of EI activity over time.

Faller and Merimaa (2004b) have suggested a single modelling approach to analyse localisation in complex listening situations, often involving multiple sources and reflections. Their model is analytically motivated and relies on evaluation of directional cues using *interaural coherence*. When multiple sources and reflections exist, localisation cues presented to the auditory system are considerably different than in an anechoic environment. The model's underlying assumption is that the locus of a sound source can be determined by considering localisation cues only at time instances where they resemble free field cues. This is one of the core components used in the integrated model described in this thesis, and as such, will be further detailed in Chapter 6.

Goodman and Brette (2010) modelled a biological neural network for processing binaural cues. In their study, they showed that due to the HRTF filtering of the outer ear, the location-specific structure of binaural signals is transformed into location-specific synchrony patterns in the brain, which may correspond to patterns of free-field localisation cues previously learned by the auditory system. Although the Goodman-Brette model seems promising and provides an in depth view of the biological processes fundamental to sound localisation, directly employing their model is somewhat beyond the scope of this work, as it requires a substantial understanding of auditory neuroscience which is not within the author's background.

3.3.2 Applied Models of Sound Localisation

The modelling approaches presented in 3.3.1 may all be applicable to investigating problems in room acoustics and spatial audio, however such problems normally require a more complete modelling framework. Such a framework would not only account for processing and encoding of binaural cues, but also for the function of auditory peripheral system (lower stages) and spatial mapping of encoded localisation cues (higher stages).

Macpherson (1991) suggested one of the early models for localisation of auditory events generated by means of stereophonic reproduction. His model consisted of a peripheral processor and two mechanisms extracting ITDs and ILDs from the input stimuli. He also included a simplified precedence model, by which reflections later than 1ms from stimulus onset are thought to be irrelevant and are trimmed respectively. Macpherson also introduced an approach to arrive at a single localisation judgement from predicted interaural cues, by making use of lookup tables. This approach, and its relevancy to this work, is further discussed in Chapter 6.

Pulkki and colleagues (1999) investigated localisation of artificially generated images in different sound reproduction paradigms using a binaural model. They used a peripheral processor combined with a binaural stage in which ITDs are extracted from Jeffress cross-correlation function and ILDs are calculated from the time averaged signal power. Whilst their model provides important insight into spatial audio reproduction, it is not designed to operate in reverberant conditions, and as such, is less suitable for research in room acoustics.

Braasch (2005a) evaluated spatial images created with different stereo recording techniques using binaural models. This is slightly different from what we try to achieve in this study in the sense that the transducers are microphones (i.e. receivers, not sources) and reflections come from the recording room (rather than the listening room). Nevertheless, as the core components of the filtered stimuli are present in both cases (i.e. reflections, multiple sources), it can be assumed that the model would be applicable to this research to some extent. Braasch's setup consisted of a real source in a hall which was recorded by different stereo microphone configurations. Measured signals were convolved with HRTFs at ±30°, essentially simulating listening through a stereophonic loudspeaker system in a (reflection-free) room. In the binaural stage, Braasch used a combination of the running IACC (to account for ITDs), a modified Breebaart algorithm (adapted to account only for ILDs) and a modified Lindemann algorithm to account for the precedence effect. Although in theory, Braasch's model could be used in context of this research, it should be noted that the listening situations he studied are different than those investigated here. This is because, when reproducing a reverberant recording over a traditional stereo system, the directional information of the recorded reflections is lost.

Park and colleagues (2008) suggested two types of pattern recognition algorithms to serve as central processors or decision-makers in a Breebaart-type model. In their study, they have rigorously tested the model's ability to predict stereophonically reproduced images. Whilst Breebaart's model is capable of handling time-varying signals, Park and colleagues focused on time averaged responses, as they assumed that the localisation task is performed under anechoic conditions.

—You know the sound of two
hands clapping; tell me, what
is the sound of one hand?

Hakuin Ekaku

4

Physical and Numerical Constraints in Sound Source Modelling

In finite difference time domain simulation of room acoustics, source functions are subjected to various constraints. These depend on the way sources are injected into the grid and on the chosen parameters of the numerical scheme being used. Constraints may be physical, meaning that they arise from processes governing the generation of sound, whereas others may be numerical, meaning that they arise from limitations imposed by choosing parameters for a particular finite difference scheme. In this chapter, the process of designing and injecting sources is generalised by introducing a system of three cascaded filters characterising the source mechanics, sound transduction and restriction of numerical dispersion. It is shown that hard, soft and transparent sources can be seen as special cases of the cascaded filters approach. We then suggest an optimal combination of filters, resulting in a source which has a maximally flat bandwidth, does not scatter incoming waves, and is correctly scaled with the grid parameters. The model, which we call a *Physically Constrained Source*, is developed by considering the mechanics of a small pulsating sphere. By deriving the model from first principles, the mechanics of the source is related to the injection process, and a physically plausible method to design sources which can be parametrically controlled in a fashion resembling real transducer operation is suggested. The physically-constrained source is compared to existing source models, and is shown to be the only method which, by definition, generates a soundfield that converges with a closed-form solution to the wave equation. Some of the concepts, formulations and results presented in this chapter have been featured in (Sheaffer et al., 2012).

4.1 Background

The issue of grid excitation in FDTD methods has received only little attention and corresponding literature is relatively sparse. It appears that there is no single correct way to excite a grid and the method of source excitation is, in most cases, application specific. Here, focus is given to the excitation of 3D grids for simulation of longitudinal acoustic wave propagation, which is a different challenge than excitation of other grids, for example for simulating transversely vibrating mechanical systems (Fontana and Rocchesso, 1998). In the specific case of room acoustics, the grid is normally excited either by employing a time-varying function at a single grid node (Schneider et al., 1998a) approximating a unit impulse, or by pre-determining a field-wide pressure distribution (Sakamoto et al., 2002). Whilst the former approach is useful for studying the transient build-up of the sound field, the latter is more suitable for investigating its steady-state. In this chapter, only time-varying sources are considered, as they more readily resemble the way one would perform an acoustic measurement of a real space. It is also assumed here, that one measures or models a space in order to *objectively study* its acoustics, and/or to *visualise* wave behaviour in it, and/or to *auralise* it.

In electromagnetic (EM) FDTD, where the numerical scheme approximates a solution to Maxwell's equations (Yee, 1966), one normally excites the grid using voltage or current sources by replacing or adding them to the calculated magnetic or electric fields (Taflove and Hagness, 2000; Gedney, 2011). When voltage sources are used to *replace* the electric field, waves propagating back to the source position get reflected by the source (Buechler et al., 1995). In acoustics modelling, this is equivalent to *imposing* pressure on the field, and is often referred to as a *Hard Source* (HS) implementation. When voltage sources are *added* to the calculated electric field, the aforementioned scatterring does not occur, and in acoustics, this is often referred to as a *Soft Source* (SS) implementation, where pressure is being *superimposed* on the field. This superposition of voltage (or pressure), however, results in a field which bears little resemblence to that of a hard-source, and such sources have been shown to be equivalent to current (or velocity) sources (Schneider et al., 1998a). The underlying numerical causes for this will be further discussed in this chapter. In order to generate a sound field which is similar to that of a hard source but whithout the scattering effects, Schneider and colleagues (1998b; 1998a) have formulated *Transparent Sources* (TS) which rely on measuring and compensating for the grid's impulse response. These sources, however, are computationally expensive and require significant pre-processing.

In the acoustics literature, FDTD *excitation signals* and source *injection methods* are often treated separately. In other words, it is commonly believed that one has the freedom to choose the excitation signal regardless of the chosen injection method, and *vice versa*. As will be further discussed in this chapter, there are different ways in which the two are related, and these interrelations are dictated by the numerical parameters of the model as well as by the physics of the governing system. In order to decide on an appropriate excitation signal and injection method, one should consider the different constraints imposed by the mechanics of sound generation, the physics of fluid emergence and their representations in a finite difference model.

It appears that acoustics FDTD has blindly inherited source methods from electromagnetics, and the relationship to the physics of sound generation is often overlooked in the literature. The purpose of this chapter is twofold. Its first part provides a mathematical analysis of the physical and numerical constraints imposed by sound generation in an FDTD system (Section 4.3), which is then followed by a systematic review of existing source methods (Section 4.4). In its second part, the problem of designing appropriate sound sources is addressed by introducing a network of cascaded filters, each representing a different physical system (Section 4.5). This signal processing scheme is used to convert any arbitrarily chosen *excitation signal* into a final *source function* which, depending on chosen parameters, may adhere to some or all constraints. Once an understanding of the differenent constraints has been established, an improved method for modelling sound sources is suggested (Section 4.5.4). Finally, the method is compared to existing source models (Section 4.6) and results are discussed in light of room acoustics modelling (Section 4.7).

4.2 The Anatomy of Sources in Acoustics FDTD

4.2.1 Sources in a Yee-type Method

In Chapter 3, the conservation laws of mass and momentum and their relationship to the Yee-type FDTD method was discussed. Let us now consider the same equation but with sources. In their Eulerian form, these are given by (Morse and Ingard, 1986)

$$\frac{1}{c^2}\frac{\partial p(\mathbf{x},t)}{\partial t} + \rho_0 \nabla \cdot \mathbf{u}(\mathbf{x},t) = q(\mathbf{x},t) \tag{4.1}$$

$$\rho_0 \frac{\partial \mathbf{u}(\mathbf{x},t)}{\partial t} + \nabla p(\mathbf{x},t) = \widetilde{\mathbf{F}}(\mathbf{x},t) \tag{4.2}$$

Here, the function $q(\mathbf{x},t)$ denotes the *rate of fluid emergence* in the system in the dimensions of density per unit time (kg m^{-3} s^{-1}), and the function $\widetilde{\mathbf{F}}(\mathbf{x},t)$ is the *acoustic force* exerted upon the source volume. For simplicity, we assume that all considered excitation functions represent volume velocity sources, and as such, the force term in Equation (4.2) is neglected. Accordingly, Equations (4.1) and (4.2) can be approximated using finite difference operators as

$$\delta_t p\big|_{\mathbf{i}}^{n} = \underbrace{c^2 T q\big|_{\mathbf{i}}^{n}}_{\textit{Source Term}} - z_0 \lambda \left(\delta_x u_x\big|_{\mathbf{i}}^{n} + \delta_y u_y\big|_{\mathbf{i}}^{n} + \delta_z u_z\big|_{\mathbf{i}}^{n} \right) \tag{4.3}$$

and

$$\delta_t u_x\big|_{\mathbf{i}}^{n} = -\frac{\lambda}{z_0}\delta_x p\big|_{\mathbf{i}}^{n} \tag{4.4a}$$

$$\delta_t u_y\big|_{\mathbf{i}}^{n} = -\frac{\lambda}{z_0}\delta_y p\big|_{\mathbf{i}}^{n} \tag{4.4b}$$

$$\delta_t u_z\big|_{\mathbf{i}}^{n} = -\frac{\lambda}{z_0}\delta_z p\big|_{\mathbf{i}}^{n} \tag{4.4c}$$

By eliminating any source terms, one obtains the update equations for the medium, as shown in Chapter 3.

4.2.2 Sources in a Wave Equation Method

In a similar fashion, when sources are present in the domain, then one should consider the inhomogeneous wave equation,

$$\frac{1}{c^2}\frac{\partial^2 p(\mathbf{x},t)}{\partial t^2} - \nabla^2 p(\mathbf{x},t) = \psi(\mathbf{x},t) \tag{4.5}$$

To enable a direct comparison with other studies, here $\psi(\mathbf{x},t)$ is defined as a general source driving function, whose physical relation to fluid emergence in the system shall be further discussed in Section 4.3. Using the same nomenclature, the wave equation can be modelled as

$$\left(\delta_t^2 - \lambda^2\delta_{\mathbf{x}}^2\right) p\big|_{\mathbf{i}}^{n} = \underbrace{c^2T^2\psi\big|_{\mathbf{i}}^{n}}_{\textit{Source Term}} \tag{4.6}$$

Once again, by eliminating the right hand side of Equation (4.6), one obtains the update equations for the medium as shown in Chapter 3.

4.3 Constraints in Sound Source Modelling

When designing sound sources for an FDTD simulation, one normally faces two principal questions. First, an excitation signal should be chosen based on the desired temporal and spectral characteristics of the source. Next, one chooses an injection method, which most generally involves imposing or superimposing the excitation signal on a grid node. Ideally, one would have the freedom to choose any arbitrary excitation signal and to transparently inject it into the grid without any ramifications. However, several physical and numerical systems govern the process of exciting an FDTD grid, and these impose constraints that should be taken into consideration when designing sources. In this section these constraints are identified by analysing the physical and numerical systems that govern the source.

4.3.1 Transduction Constraints

In Equation (4.1), the time derivative of pressure and space derivatives of particle velocity are related to the rate of fluid emergence, $q(\mathbf{x}, t)$. Considering a source positioned at a single node of an FDTD grid, in which each cell occupies a volume equal to X^3, then fluid emergence in the system can be expressed by means of the source's volume velocity $\hat{q}(t)$, as follows

$$q(\mathbf{x}, t) = \frac{\rho_0}{X^3}\hat{q}(t)\delta(\mathbf{x} - \mathbf{x}') \tag{4.7}$$

where $\mathbf{x}' = (x', y', z') \in \mathbb{R}^3$ denotes the source position. Equation (4.7) presents a *scaling constraint* which relates the amplitude of the source to the volume it occupies. By combining equations (4.2) and (4.1), the particle velocity vector is eliminated and the wave equation is derived. Note that here equation (4.2) is used without the acoustic force term. It follows from this derivation that the source term in the wave equation becomes

$$\psi(\mathbf{x}, t) = \frac{\partial q(\mathbf{x}, t)}{\partial t} = \frac{\rho_0}{X^3}\frac{d}{dt}\hat{q}(t)\delta(\mathbf{x} - \mathbf{x}') \tag{4.8}$$

Physically, the quantity $\psi(\mathbf{x}, t)$ has the dimensions of density per unit time squared (kg m^{-3} s^{-2}), and can be thought of as fluid emergence due to volume acceleration of the source. Thus, when designing sources for use in wave equation schemes, a *differentiation constraint* also applies, meaning that volume velocity should be injected as its first time derivative. Observe that the source terms in Equations (4.1) and (4.5) are supplemental to the fundamental time-space relationships, that is, if one sets $q(\mathbf{x}, t) = 0$ then the homogeneous wave equation is obtained. This implies that fluid emergence is an *additive* process, which numerically means that source nodes should also be evaluated with the FDTD update equations for air. Therefore, there are either two or three transduction constraints, depending on the chosen method. For Yee-type methods, the governing physics show that volume velocity sources should be appropriately scaled (*scaling constraint*), as well as superimposed on the grid (*superposition constraint*). In wave-equation methods the same constraints apply, but in addition, source functions should be differentiated in time (*differentiation constraint*).

4.3.2 Mechanical Constraints

In order to generate a volume velocity at the source, some mechanical process is required. This normally corresponds to some vibrating object, for example, a pulsating sphere or a moving piston. Such mechanical systems are governed by the laws of motion, and accordingly introduce further modelling constraints. Depending on the complexity of the modelled transducer, these can range from a simple second order linear system, to a full physical representation of a loudspeaker (Small, 1972). Whilst many mechanical constraints are specific to the chosen transducer, some can be generalised to the problem of sound production. Continuous DC flow, for example, is something that transducers generally cannot produce, and as such, one would expect that

$$\int_{-\infty}^{\infty} \psi(\mathbf{x}, t) dt = 0 \tag{4.9}$$

which naturally occurs if the *differentiation constraint* described in Section 4.3.1 is adhered to. However, if one decides to arbitrarily choose $\psi(\mathbf{x}, t)$, then failure to adhere to this constraint may have detrimental ramifications, as will be further discussed.

4.3.3 Numerical Constraints

Finite difference methods are subject to numerical dispersion, which increases as the ratio of the sample rate to the modelled frequency is decreased. This results in waves whose phase velocities are dependent on frequency and in non-isotropic schemes, also on the direction of propagation. When the grid is excited at frequencies which are prone to substantial dispersion, the resulting response is contaminated with numerical errors, which not only impair the ability to perform visual analysis, but also introduce undesired audible artefacts. Thus, if one employs a simulation in order to visually investigate wave phenomena, then it is important that high frequencies are removed from the excitation signal, to prevent these from being generated in the model. This is referred to as *Bandwidth Constraint*.

When transient phenomena are studied, an impulsive source is required and therefore bandlimiting the excitation function presents a tradeoff between the source's bandwidth and its time compactness. If the goal of the simulation is only to produce auralisations, then one may opt to excite the grid directly with the programme material to be auralised. A more efficient way is to model the room's response using a unit impulse, and then employ convolution in the post-processing stage. Either way, if one wishes to eliminate the audible effects of dispersion, the result should be appropriately bandlimited.

Another numerical constraint involves the temporal length of the source function. If the excitation signal is not finite in time, then one should truncate it at points such that any discontinuity errors arising from the truncation will be lower than any other numerical errors inherent in the scheme. Thus, it is useful that the excitation signal is finite by definition, which is here referred to as a *Finiteness Constraint*. Ideally, the value of the source and all of its derivatives up to the truncation order of the scheme would be zero at simulation onset, however this has been shown to be more prominent in higher order numerical schemes (Botts et al., 2010).

4.4 Existing Source Models

At this point, it is useful to review some of the specific requirements from a sound source in an idealised room measurement. First, it is desired that the bandwidth of the source is wide enough to cover the entire frequency range of interest, and that it is sufficiently flat within that range. Secondly, the sound source should generate a prescribed pressure field, meaning that one should be able to predict its behaviour in free field. Thirdly, it is useful to have a source that can excite the room equally in all directions at all frequencies of interest. Lastly, although never physically feasible, it is occasionally useful to be able to excite the room without introducing scattering effects from the source itself.

In a numerical model, some additional requirements may apply. It is important that the process of grid excitation is *numerically consistent*, meaning that a change in grid parameters would not affect the magnitude of the sound field generated by the source. Also, when transient phenomena are investigated, it is desired that the source excitation signal is sufficiently compact in time, so that temporal overlap between discrete reflections does not occur. In this section, existing source models are reviewed in light of the above requirements and the physical and numerical constraints described in Section 4.3.

4.4.1 Hard Sources

A hard source is the simplest form of grid excitation, according to which an acoustic quantity is directly imposed on the source node. This quantity may be sound pressure or particle velocity, and is here represented in the discrete domain by the excitation signal $s_p\big|^n$, whose notation will be further explained. Thus, the update equation for a hard source injected into a pressure grid is

$$p\big|_{\mathbf{i}'}^{n+1} = s_p\big|^{n+1} \tag{4.10}$$

where $\mathbf{i}' = [l', m', i']$ denotes the index position of the source. The first thing evident from Equation (4.10) is that the conservation laws of mass and momentum are not satisfied at the source node. This implies that acoustic pressure appears with no underlying physical cause, which means that the hard source paradigm does not adhere to the aforementioned transduction constraints. Because update equations for air cannot operate over a hard source node, any incoming waves get scattered by the source. Accordingly, the node is often loosely thought of as a sound radiating boundary node. This description, however, is not precise, as such an element should adhere to boundary conditions which are not evident in the hard source formulation. One could argue that in a real measurement scenario, a loudspeaker would inevitably be present in the room, and therefore scattering from a hard source is not an unrealistic outcome. However, it should be borne in mind that in an FDTD simulation, the *physical size* of the sound radiating node is directly dependent on the spatial sample period, meaning that the scattering effects of the hard source are numerically inconsistent.

Since the hard source is directly imposed on the grid, the transient change in sound pressure at the source node takes the form of the source function, which is a desired feature, as one can directly employ excitation signals that generate a prescribed pressure. However, as the equations of mass and momentum are not satisfied at the source node, the inability of air particles to effectively perform rarefaction may cause unwanted low-frequency artefacts when certain excitation signals are used (Jeong and Lam, 2012). Thus, it is postulated that the hard source paradigm is fundamentally flawed in the sense that it does not adhere to physical laws governing the medium nor the boundaries.

4.4.2 Soft Sources

The scattering and low-frequency problems of hard sources can be overcome by employing *Soft Sources*. In such cases, the excitation signal is superimposed on source node, which has already been evaluated by the medium update equations. The update equation for a soft source node on a pressure grid is therefore

$$p\big|_{\mathbf{i}'}^{n+1} = \left\{p\big|_{\mathbf{i}'}^{n+1}\right\} + s_p\big|^{n+1} \tag{4.11}$$

where $\left\{p\big|_{\mathbf{i}'}^{n+1}\right\}$ represents the result of updating the node with the general update equation for air. Soft sources may have different effects depending on the type of scheme being used. In Yee-type grids, e.g. see (Botteldooren, 1995), a soft source function is differentiated due to the staggered nature of the scheme. The update equation for pressure, progresses time in only one half of a step and the following half-step consists of updating particle velocity by evaluating the derivatives of pressure. Thus, in Yee-type schemes, the signal $s_p\big|^n$ is injected as its first time derivative. This inherent differentiation process is important, as it ensures that no DC component is being injected into the grid, however it also severely modifies the spectrum of the source function, by generating an undesired roll-off in low frequencies.

In wave-equation methods, the soft source does not get automatically differentiated, and as such, behaves in a different fashion. The resulting pressure at the grid node directly takes on the shape of $s_p\big|^n$, which is a desired feature. Because of this, however, one is not free to arbitrarily choose the excitation signal. Consider, for example, the case of an excitation signal which has a DC component, and is placed in front of a reflecting surface on the x-plane. For a plane wave source of arbitrary amplitude A, interacting with a surface having a reflection coefficient $\hat{r}$, the total sound pressure along the plane is given by

$$p(x,t) = Ae^{j(\omega t - kx)} + \hat{r}Ae^{j(\omega t + kx)} \tag{4.12}$$

As such, the sound pressure at DC is uniformly $p = A(1+\hat{r})$ along the plane. Since the soft source is being added to existing pressure, then for any $\hat{r} > 0$ a pre-existing DC component would constructively superimpose on itself at the source node, resulting in an incremental offset in the response. Similar effects have been observed in the field of computational electrodynamics (Su et al., 2005). In context of the source modelling constraints, such *arbitrary soft sources* always fulfil the superposition requirements dictated by the equation of continuity, however, the differentiation constraints are inherently adhered to only in Yee-type grids.

Based on digital waveguide analysis, Karjalainen and Erkut (2004) identified the requirement to superimpose, differentiate and scale soft sources in wave-equation FDTD schemes. Their formulation, which shall be further referred to as a *differentiated soft source*, yields the following update equation

$$p\Big|_{l',m',i'}^{n+1} = \left\{ p\Big|_{l',m',i'}^{n+1} \right\} + \frac{\rho_0 c X}{2A_w} \left(\hat{q}\Big|^{n+1} - \hat{q}\Big|^{n-1} \right) \tag{4.13}$$

where A_w denotes the cross-sectional area of the waveguide, occupied by the source. Note that here the excitation function is explicitly defined as a volume velocity. Furthermore, the formulation adheres to both superposition and differentiation constraints. However, being drawn from 1D waveguide theory, the scaling factor is only correct for one dimensional schemes, as will be further shown.

A side effect of all soft sources is that the injected excitation function is modified by the grid's impulse response, which occurs because the grid's update equations operate over the source node (Schneider et al., 1998a). It is important to distinguish between the effects of the grid's impulse response which are minimal, and the differentiation process which severely modifies the source spectrum. In Yee-grids, the two effects can be easily confused, however, one should bear in mind that the former is a *numerical* artefact which can be traced back to finite difference approximation errors, whereas the latter is a *physical* effect and is unrelated to the discrete nature of the model.

4.4.3 Transparent Sources

Schneider and colleagues (1998a) have addressed some issues of source modelling by making use of *Transparent Sources*, which do not scatter incoming waves and do not get modified by the grid's impulse response. This approach requires that the grid's impulse response is measured prior to the simulation stage and is compensated for during simulation, and can be mathematically described by

$$p\Big|_{\mathbf{i}'}^{n+1} = \left\{ p\Big|_{\mathbf{i}'}^{n+1} \right\} + s_p\Big|^{n+1} - \sum_{\mu=0}^{n} I\Big|^{n-\mu+1} s_p\Big|^{\mu} \tag{4.14}$$

where $I\Big|^n$ denotes the pre-measured impulse response of the grid, which is obtained by exciting the grid with a unit impulse and capturing the result of updating the source node with the update equation for air (Schneider et al., 1998a). Therefore, transparent sources not only compensate for the grid's impulse response, but also reverse the effects of source differentiation, effectively resulting in a sound field similar to that of a hard source, but without scattering incoming waves. As was discussed by Jeong and Lam (2012), transparent sources suffer from the same low frequency artefacts as hard sources.

It should be noted that the grid's impulse response must be obtained at the absence of any scattering objects, which for long simulation times entails modelling a large domain. Thus, the requirement to pre-measure the grid's impulse response constitutes a less attractive method, as it is more computationally expensive and less intuitive to implement. Furthermore, transparent sources do not adhere to any scaling constraints, and due to the grid compensation process, neither to the differentiation constraint.

4.4.4 Issues with Existing Source Models

Table 4.1 summarises the transduction constraints of the different source models discussed in this section. It can be seen that for the case of a three-dimensional problem, none of the existing source models fully adhere to all transduction constraints. It should also be noted that existing source models are only concerned with the way excitation functions are injected into the grid. It was shown in this section that some constraints are inter-related. For example, some excitation functions may cause problems when a hard or transparent source injection method is chosen, and an excitation function injected as an arbitrary soft source cannot contain a DC component. It is therefore advocated that a complete source model should not only be concerned with grid injection, but also with the entire process of fluid emergence in a *numerical* model. In other words, a complete source model should adhere not only to transduction constraints, but also to mechanical and numerical constraints.

4.5 A Cascaded Filters Approach

The process of modelling sound sources can be represented in a structural, parametric manner, by considering a system of three cascaded filters whose input is

Table 4.1: Summary of existing source models in light of the constraints discussed in Section 4.3. Note that whilst existing source models do not explicitly discuss the shape of the excitation signal, in principle, they could be designed such that the method would adhere to numerical and mechanical constraints. HS - Hard source, SS - Soft source, DSS - Differentiated soft source, TS - Transparent source. Methods marked by ± adhere to the constraint in special cases only.

	HS	SS	DSS	TS
Scaling	-	-	±	-
Differentiation	-	±	+	-
Superposition	-	+	+	+
Mechanical	-	-	-	-
Bandwidth	-	-	-	-
Finiteness	-	-	-	-

a Kronecker delta, as shown in Figure 4.1. The delta function is first passed through a *pulse shaping filter*, $H_p(z)$, which ensures that the system is driven using an appropriately bandlimited signal, resulting in the excitation signal $s_p\big|^n$. The output of the pulse-shaping filter is used to drive a *mechanical filter*, $H_m(z)$, which shapes the excitation signal according to the behaviour of the modelled sound generating mechanism. Lastly, an *injection filter*, $H_i(z)$, is used in order to shape the excitation signal according to the transduction constraints, resulting in a source function, $s_g\big|^n_{\mathbf{i}'}$, which can be used to directly drive the grid.

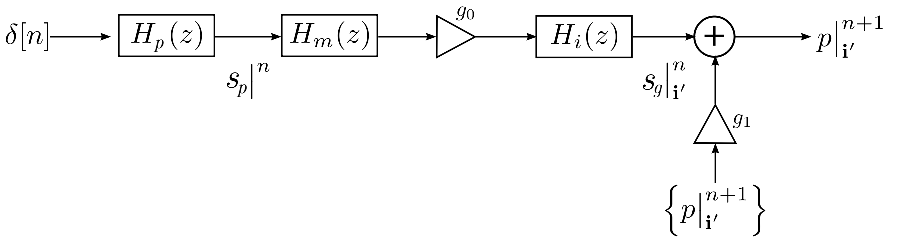


Figure 4.1: Cascaded filter representation of source model. $H_p(z)$ pulse-shaping filter, $H_m(z)$ mechanical filter, $H_i(z)$ injection filter, $s_p\big|^n$ excitation signal, $s_g\big|^n_{\mathbf{i}'}$ final grid signal to be injected.

4.5.1 The Pulse-Shaping Filter, $H_p(z)$

If one does not require to visually inspect the simulation response (e.g. in the case of auralisation), then removal of dispersive frequencies can be accomplished by lowpass filtering in the post-processing stage. In such case, one can quite freely choose a filter suitable for the task. Yet, in other cases one would excite the grid with a pre-processed signal. Ideally such a signal would be appropriately bandlimited, and preferably finite in length. It should have clearly defined cutoff frequencies, a flat passband, a steep transition band and a ripple-free stopband, thus effectively preventing grid excitation at frequencies prone to strong numerical dispersion. In this section, some commonly used excitation signals are reviewed, which in context of the cascaded filters approach represent the impulse response of the pulse-shaping filter, $H_p(z)$. Since the output of this filter is effectively the excitation signal $s\big|^n$ itself, then for simplicity, formulations are given directly in the discrete time domain. In order to provide some visual support, Figure 4.2 depicts the different filters discussed in this section, in the time and frequency domains.

Gaussian filter

The time-shifted Gaussian pulse is a popular excitation signal in FDTD methods. It is defined, in the discrete time-domain, by the following equation

$$s\big|^n = A_p \exp\left(\frac{-T^2\left[n-n_0\right]^2}{2\sigma^2}\right) \tag{4.15}$$

where A_p is the peak amplitude of the function, n_0 is the number of samples by which the pulse is shifted, and σ is the Gaussian variance related to its cutoff (half-power bandwidth) frequency by

$$f_c = \frac{\sqrt{2\ln 2}}{2\pi\sigma} \tag{4.16}$$

The Gaussian pulse is a unipolar function which decays to zero only at $n = \pm\infty$, and is symmetrical around $n = 0$. Thus, to ensure smoothness, the signal should be appropriately time shifted (i.e., $n' >> 0$) and carefully truncated.

Modulated Gaussian filter

The sine-modulated Gaussian pulse is widely used in electrodynamics (Gedney, 2011) and has been recently introduced for acoustic FDTD modelling (Jeong and Lam, 2012). It's impulse response is given by

$$s\big|^{n} = A_p \sin\left(\omega_m T\left[n - n_0\right]\right) \exp\left(\frac{-T^2\left[n - n_0\right]^2}{2\sigma^2}\right) \tag{4.17}$$

where $\omega_m = 2\pi f_m$ is the angular modulating frequency. The sine modulation results in a bipolar function, which is suitable for implementation as a hard source (Jeong and Lam, 2012) as well as an arbitrary (undifferentiated) soft source. However, this results in a gradual roll-off of low-frequencies which render the modulated-Gaussian inappropriate for applications requiring a flat bandwidth.

Ricker wavelet

Ricker wavelets have been originally introduced for geophysical modelling (Ricker, 1953), and have several documented uses in acoustics FDTD (Yuan et al., 1999; Tolan and Schneider, 2003; Redondo et al., 2007). The wavelet is a normalised second-derivative of a Gaussian function, and can be described by the following equation

$$s\big|^{n} = \left(1 - 2\gamma\left[n - n_0\right]^2\right) \exp\left(-\gamma\left[n - n_0\right]^2\right) \tag{4.18}$$

where $\gamma = \left(\pi f_c T\right)^2$, and f_c is the peak frequency of the wavelet.

Blackman-Harris window

The impulse response of the Blackman-Harris window is a common excitation signal in computational electrodynamics (Gedney, 2011), and is described by

$$s\big|^{n} = \begin{cases} \sum_{\mu=0}^{3} a_\mu cos\left(\frac{\mu\omega_c nT}{1.55}\right) & , 0 < n < \frac{1.55}{f_c T} \\ 0 & \text{otherwise} \end{cases} \tag{4.19}$$

where $\omega_c = 2\pi f_c$ is the angular half-bandwidth of the function, and a_μ is given by

$$a_\mu = \begin{pmatrix} 0.353222222 \\ -0.488 \\ 0.145 \\ -0.010222222 \end{pmatrix} \tag{4.20}$$

The Blackman-Harris window has a flat passband and a steep transition band, which makes it a good candidate for simulations which require bandwidth maximisation. However, its inherent stopband ripples may induce numerical dispersion, even though they are of relatively small magnitude. This is especially evident when lower cutoff frequencies are required.

Maximally-flat filter

Maximally-flat FIR filters (MAXFLATs) (Herrmann, 1971) are family of simple yet accurate filters, which have a smooth and ripple-free magnitude response. Traditionally, MAXFLAT magnitudes are calculated using polynomial expressions and then tap coefficients are obtained by employing an inverse Discrete Fourier Transform. More recently, Khan and Ohba (2000) derived explicit formulae for calculating the impulse response of MAXFLAT filters directly in the discrete time domain. With some notational changes, their formulation for a lowpass filter is given by

$$\begin{aligned} s\big|^0 &= \omega_c T \\ s\big|^n &= \frac{(2N-1)!!^2 \sin(n\omega_c T)}{\hat{b}n(2N+n-1)!!(2N-n-1)!!} \end{aligned} \tag{4.21}$$

where !! is the double factorial, ω_c is the angular cutoff frequency, N is the filter order and $\hat{b}$ is given by

$$\hat{b} = \begin{cases} 2 & , n = \pm 1, 3, 5, \ldots, 2N-1 \\ \pi & , n = \pm 2, 4, 6, \ldots, 2N-2 \end{cases} \tag{4.22}$$

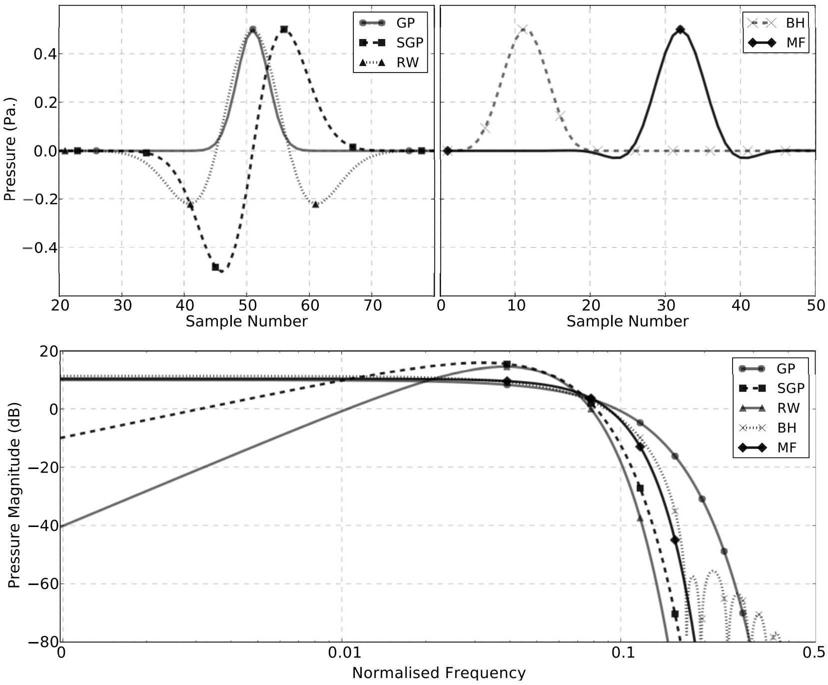


Figure 4.2: Impulse response (upper panes) and magnitude response (lower pane) of different anti-dispersion filters. GP Gaussian pulse, SGP Sine-moduled Gaussian pulse, RW Ricker wavelet, BH Blackman-Harris window, MF Maximally flat FIR. All filters were designed such that their −3dB cutoff points correspond to $0.075f_s$.

Note that the response is calculated for positive *and* negative n, therefore $s\Big|^n$ should be shifted by $(2N-1)$ samples to ensure that the filter is causal. The use of MAXFLATs as excitation functions has several advantages. Being non-recursive filters, MAXFLATs have a finite impulse response, and as such, there is no need to truncate the signal. Additionally, a flat magnitude response and steep transition band can be achieved even with relatively low order filters, meaning that the excitation signal is compact in time.

4.5.2 The Mechanical Filter, $H_m(z)$

The purpose of the mechanical filter, denoted by $H_m(z)$, is to ensure that excitation signals are physically plausible, and adhere to the mechanical constraints discussed in Section 4.3. In other words, if one would inject an excitation signal directly into the grid, then it may end up representing a function which cannot be produced by real world mechanical systems. For example, if one applies a transient pulse of force to a loudspeaker driver, then it will inherently modify the signal according to its mechanical characteristics. The mechanical constraints governing the transducer can be normally represented as a linear system of masses, dampers and springs, whose complexity depends on the modelled transducer. Such systems may range from a simple piston to a fully parametric model of a loudspeaker driver in a cabinet. As shown in Section 4.3, the most general constraint imposed by mechanical systems is that they cannot produce continuous DC. Thus, the simplest and most general form of a mechanical filter is a DC-blocker given by

$$H_m(z) = \frac{1 - z^{-1}}{1 - \alpha z^{-1}} \tag{4.23}$$

where α is typically between 0.9 and 1.0 (Smith III, 2006). Alternatively, one may convolve the excitation signal with a measured impulse response of a transducer. However, in this case the system's response may be compromised by engineering constraints (resulting in a non-flat frequency response, for example), which can be avoided when taking a modelling approach.

4.5.3 The Injection Network

The injection filter, denoted by $H_i(z)$ and its corresponding gain functions g_0 and g_1, represent the final stage in transforming the excitation signal $s_p\big|^n$ into the source function $s_g\big|^n_{\mathbf{i}'}$. This DSP structure is a digital representation of the transduction constraints discussed in Section 4.3. The gain function g_0 is used to convert surface velocity to source strength by accounting for the different *scaling constraints*. The signal is then routed through a general purpose digital filter which can perform different operations such as differentiation (to adhere to *differentiation constraints*) or grid IR compensation (in the case of a transparent source). Lastly, the gain function g_1 controls the *superposition constraint*, and may take on the values 0 or 1 depending on whether the source funciton is imposed or superimposed on the grid.

4.5.4 A Physically Constrained Source Model

We now formulate a combination of cascaded filters which constitute a parametric source model that adheres to the physical constraints discussed in Section 4.3.

Governing equations

Let us consider a pulsating sphere of (small) radius a_0 whose surface velocity $\nu(t)$, in vacuum, is governed by

$$M\frac{\partial \nu(t)}{\partial t} = -R\nu(t) - K\int \nu(t)dt + F(t) \tag{4.24}$$

where M, R, and K are respectively, the mass, damping and elasticity constants characterising the mechanical system, and $F(t)$ is the force driving the sphere pulsation. With air surrounding the sphere, the mechanical impedance of the system is

$$Z(\omega) = Z_{\mathrm{v}}(\omega) + Z_{\mathrm{a}}(\omega) \tag{4.25}$$

where

$$Z_{\mathrm{v}}(\omega) = Mj\omega + R + K/(j\omega) \tag{4.26}$$

is the impedance of the system in vacuum and

$$Z_{\mathrm{a}}(\omega) = \rho_0 A a_0 \left(j\omega + (a_0/c)\omega\right) \tag{4.27}$$

is the mechanical impedance of the surrounding air (Morse and Ingard, 1986). However, the latter term may be omitted since a_0 is very small, meaning that $|Z_{\mathrm{v}}(\omega)| >> |Z_{\mathrm{a}}(\omega)|$ in all practical cases. Hence the system may be characterised by the transfer function

$$H_m(s) = \frac{s}{Ms^2 + Rs + K} \tag{4.28}$$

which has the dimension of mechanical admittance. In the time domain, the impulse response of the system is given by

$$h_m(t) = \left[\cos(\omega_r t) - \frac{\alpha}{\omega_r}\sin(\omega_r t)\right] M e^{-\alpha t} \tag{4.29}$$

where $\alpha = \frac{R}{2M}$ is the damping factor, $\omega_0 = \sqrt{\frac{K}{M}}$ is the system's undamped resonant frequency and $\omega_r = \sqrt{\omega_0^2 - \alpha^2}$. At the source, the sphere's surface velocity equals the particle velocity of air, which can be mathematically expressed as convolution between the driving force and the system's impulse response, $\nu(t) = F(t) * h_m(t)$. The pulsation of the sphere causes fluid to be pushed into and extracted from the region bordering the source sphere surface, which is characterised by a *volume velocity*,

$$\hat{q}(t) = \nu(t) A_s \tag{4.30}$$

having the dimensions of volume per unit time ($m^3 s^{-1}$), where $A_s = 4\pi a_0^2$ is the surface area of the sphere. This volume velocity can then by used in conjunction with Equation (4.7) to generate a physically constrained source function.

Numerical formulation

In the z-domain, the transfer function of the PCS model's mechanical filter, $H_m(z)$, can be obtained by applying a bilinear transform to $H_m(s)$. This choice is mainly because, unlike other discretisation methods, the bilinear transform does not place any stability limits on the values of M, R and K, thus allowing them to be freely chosen. Taking the bilinear transform of Equation (4.28) we obtain the following digital filter

$$H_m(z) = \frac{b_0 + b_2 z^{-2}}{1 + a_1 z^{-1} + a_2 z^{-2}} \tag{4.31}$$

with the coefficients given by

$$\begin{aligned} b_0 &= \frac{\beta}{M\beta^2 + R\beta + K} \\ b_2 &= -\frac{\beta}{M\beta^2 + R\beta + K} \\ a_1 &= \frac{2\left(K - M\beta^2\right)}{M\beta^2 + R\beta + K} \\ a_2 &= 1 - \frac{2R\beta}{M\beta^2 + R\beta + K} \end{aligned} \tag{4.32}$$

where β is the bilinear operator, which for a pre-warped ω_0 is given by

$$\beta = \frac{\omega_0}{\tan(\omega_0 T/2)} \tag{4.33}$$

Thus, in the PCS method, one considers the quantity represented by the excitation signal $s_p\big|^n$ to be a force driving the sphere, that is, the discrete time equivalent of $F(t)$. Accordingly, passing this signal through $H_m(z)$, yields the sphere's surface velocity $\nu\big|^n$, which is then used in the final injection network. To derive the PCS injection filter and its corresponding coefficients g_0 and g_1, one needs to consider the type of scheme being used. Recall from Equation (4.1), that the source function represents the rate of fluid emergence in the system. The source term in a Yee-type scheme is respectively represented by the function $q\big|_{\mathbf{i}'}^n$ which can be related to the sphere's surface velocity by

$$q\big|_{\mathbf{i}'}^n = \frac{\rho_0 A_s}{X^3}\nu\big|^n \delta[\mathbf{i} - \mathbf{i}'] \tag{4.34}$$

Taking into account the additional scaling factors for the source term in Equation (4.3), the coefficient g_0 is given by

$$g_0 = \frac{z_0 \lambda A_s}{X^2} \tag{4.35}$$

For the Yee-type scheme there is no differentiation constraint, therefore the injection filter's transfer function is $H_i(z) = 1$. Considering the superposition constraint, we have $g_1 = 1$ to allow the update equation for air to operate over the source node. Accordingly, the final update equation for a Yee-type source node becomes

$$\begin{aligned} p\big|_{\mathbf{i}'}^{n+1} &= \left\{ p\big|_{\mathbf{i}'}^{n+1} \right\} + g_0 \nu\big|_{\mathbf{i}'}^{n+1} \\ &= \left\{ p\big|_{\mathbf{i}'}^{n+1} \right\} + (c^2T) q\big|_{\mathbf{i}'}^{n+1} \end{aligned} \tag{4.36}$$

To develop the injection filter for the wave equation method, the physical definition of $\psi(\mathbf{x}, t)$ is followed. In the numerical domain, we adhere to the differentiation constraint described in Equation (4.8), by employing central finite differences approximating the time derivative of $q(\mathbf{x}, t)$. Accordingly, the transfer function of the injection filter for the wave equation becomes

$$H_i(z) = \frac{1}{2T}\left(1 - z^{-2}\right) \tag{4.37}$$

Considering the scaling constraints drawn from the formulation of $q\big|_{\mathbf{i}'}^{n}$, the coefficient g_0 for a wave-equation source is given by

$$g_0 = \frac{\lambda^2 \rho_0 A_s}{X} \tag{4.38}$$

Adhering to the superposition constraint, we set $g_1 = 1$, and the final update equation for a wave-equation source node becomes

$$\begin{aligned} p\big|_{\mathbf{i}'}^{n+1} &= \left\{ p\big|_{\mathbf{i}'}^{n+1} \right\} + \frac{g_0}{2T}\left(\nu\big|_{\mathbf{i}'}^{n+1} - \nu\big|_{\mathbf{i}'}^{n-1} \right) \\ &= \left\{ p\big|_{\mathbf{i}'}^{n+1} \right\} + \frac{c^2T}{2}\left(q\big|_{\mathbf{i}'}^{n+1} - q\big|_{\mathbf{i}'}^{n-1} \right) \end{aligned} \tag{4.39}$$

Note that the processing shown in Figure 4.1 outputs a signal which is delayed by one-sample in comparison to the result of equation (4.39).

4.5.5 Generalising Source Models

The cascaded-filters network generalises the process of modelling a sound source for FDTD simulation. All existing source models, as well as the PCS model, can be seen as special cases of the cascaded filter method. To summarise this, Table 4.2 shows the different transfer functions and coefficients which may be

Table 4.2: Generalisation of source models using the cascaded filters approach. HS - Hard source, SS - Soft source, DSS - Differentiated soft source, TS - Transparent source, PCS - Physically constrained source. Any filter block or multiplier that does not apply to a specific source model is set to unity here so as to not affect the corresponding grid signal.

	$H_m(z)$	g_0	$H_i(z)$	g_1
HS	1	1	1	0
SS	1	1	1	1
DSS	1	$\frac{1}{2A_s}\rho_0 cX$	$1 - z^{-2}$	1
TS	1	1	$1 - I(z)$	1
PCS (Yee)	Eq. (4.31)	$\frac{1}{X^2}2z_0\lambda A_s$	1	1
PCS (Wave)	Eq. (4.31)	$\frac{1}{X}\lambda^2\rho_0 A_s$	$\frac{1}{2T}\left(1 - z^{-2}\right)$	1

used in the filter network in order to model different sources. For hard and soft sources the grid source function simply equals the excitation signal at the source position, that is $s_g\big|_{\mathbf{i}'}^{n} \equiv s_p\big|^{n}$ with the only difference being the value of g_1 which controls the superposition constraint. In Yee-type schemes, the dimension of a hard source is pressure and the dimension of a soft source is velocity (due to the inherent differentiation), whereas in wave equation schemes both sources have the dimension of pressure. Differentiated soft sources used in wave equation schemes correct this by calculating the signal's time derivative in the injection filter, which effectively changes the dimension of the excitation signal. Transparent sources feature a processing chain similar to that of soft sources, with the injection filter designed to compensate for the grid IR and reverse the effects of inherent differentiation (if it is present, such as in Yee methods), effectively resulting in a soundfield similar to that of a hard source.

For the PCS method, the dimension of $s_p\big|^{n}$ is mechanical force, and after the complete signal processing chain, the source function represents source density (in Yee methods), i.e. $s_g\big|_{\mathbf{i}'}^{n} = q\big|_{\mathbf{i}'}^{n}$, or its first time derivative (in wave equation methods), i.e. $s_g\big|_{\mathbf{i}'}^{n} = \psi\big|_{\mathbf{i}'}^{n}$.

4.6 Results

4.6.1 Prescribed Pressure

In room acoustics simulation, it is desirable to design source functions that generate a sound field of predictable magnitude. In this section, the PCS model is employed to accomplish this task. The goal is to design an excitation signal which would propagate omni-directionally (within the dispersion limitations of the FDTD method), and would feature a flat response in a frequency range between two defined cutoff frequencies. In a spherically-symmetrical system, a closed-form solution for Equation (4.5) in free field is possible. In such case, the spherical Laplace operator in (4.5) becomes angle-independent, and with a point-source approximation, the sound pressure at the distance $r = \|\mathbf{x} - \mathbf{x}'\|$ is given by (Morse and Ingard, 1986)

$$p(r,t) = \frac{\rho_0}{4\pi r}\frac{d}{dt}\hat{q}\left(t - \frac{r}{c}\right) \tag{4.40}$$

Thus, Equation (4.40) provides a good reference to verify that the PCS method converges with the wave equation, in terms of both scaling and wave shape. To test this, a receiver was placed at the centre of a 6x6x6m domain, which was solved using the standard rectilinear scheme ($a = 0$ and $b = 0$) at a sample rate of 16kHz. A PCS source was placed at a radial distance of $r = 1.5$m and an azimuth of 45° on the same plane as the receiver. The simulation was executed for long enough for the entire signal to propagate from the source to receiver without introducing any reflections from the boundaries.

For the pulse-shaping filter, the impulse response of a MAXFLAT FIR was used. The filter was designed according to Equation (4.21), with $N = 16$ and a high cutoff frequency of $0.075 f_s$, which corresponds to the 2% dispersion criterion for the standard rectilinear scheme, as described in (Kowalczyk and van Walstijn, 2011). The magnitude of excitation was chosen such that the peak amplitude of the filter's output is normalised to a driving force of 250μN. The MAXFLAT FIR features a steep transition band and a highly attenuated stop-band whilst still remaining compact and finite in time, which makes it an ideal candidate for FDTD excitation, but it is noted that the PCS method may be used in conjunction with any other pulse shaping filter.

The mechanical filter of the PCS is characterised by the system resonance ω_0 and quality factor Q. The former controls the low cut-off frequency whilst the latter defines the peakedness of the transition between the rolled-off and usable frequency bands. In an optimal transducer design process, the designer would specify the desired values for these parameters and the remaining electro-mechanical quantities would be engineered accordingly. Here it is assumed that the source has some mass, M, and the remaining damping and stiffness coefficients are calculated from $R = \frac{\omega_0 M}{Q}$ and $K = M\omega_0^2$, respectively. In this experiment, the surface area of the sphere is considered to be equal to the surface area of a single FDTD grid cell, and its mechanical constants correspond to values $M = 25$g, $f_0 = 100$Hz and $Q = 0.7$. Simulation was repeated using both the wave equation method and the Yee-type method, and a reference response was calculated by passing the PCS volume velocity through Equation 4.40. As shown in Figure 4.3, when using the PCS model, both methods converge well with the closed form solution.

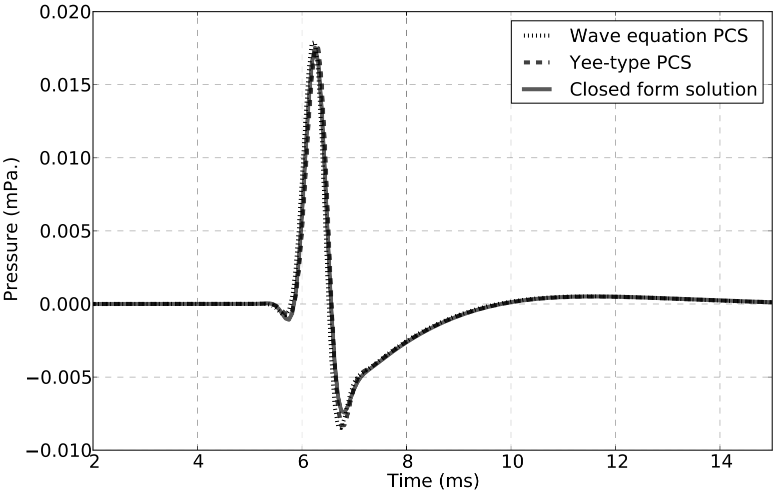


Figure 4.3: Sound pressure at the receiving position of a domain excited using the PCS method. Dotted line - wave equation method, dashed line - Yee type method, solid line - closed form solution.

To study the frequency response of the PCS model, the same experiment was conducted only using the wave equation method based on the interpolated wideband scheme ($a = 1/4$ and $b = 1/16$), allowing for the high cutoff frequency to be increased to $0.25f_s$. The PCS resonance was kept at $f_0 = 100\text{Hz}$, which corresponds to a normalised frequency of $0.0063f_s$. To exemplify the effects of the system's Q on the frequency response, the simulation was repeated for different values of Q ranging from 0.5 to 2.0. As seen in Figure 4.4, the PCS model allows generation of sources having a flat bandwidth between the system's resonance and the cutoff frequency of the pulse-shaping filter. As expected from a second order linear system, adjusting Q controls the tradeoff between the steepness of the low-frequency transition band and the magnitude of resonance.

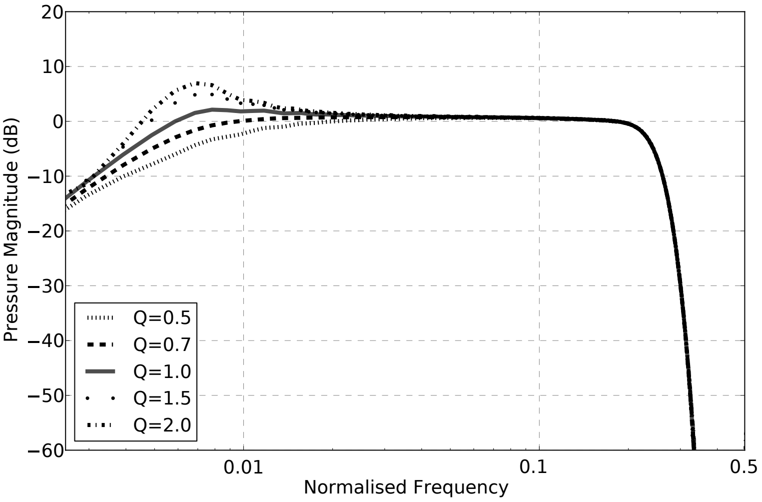


Figure 4.4: Frequency response at the receiving position of a domain excited using the PCS method, with five different values of Q. The system's resonance is at $0.0063f_s$ and the cutoff frequency of the pulse shaping filter is at $0.25f_s$.

4.6.2 Numerical Consistency

When simulating a physical system, modification of numerical parameters should only affect the accuracy of the model and not the physical behaviour of the system. In context of finite difference modelling, changing the model's sample rate should not affect the magnitude of the generated sound field. This is directly related to the scaling constraint discussed in Section 4.3, which relates fluid emergence, which is a physical quantity, to the spatial sample period of the grid, which is a numerical property. Accordingly, it is expected that source models which do not adhere to this constraint, will not yield consistent results across different sample rates.

To test this, a standard rectilinear grid was solved using the wave equation FDTD method, using three representative source models, namely a hard source, a differentiated soft source and a physically constrained source. Transparent sources and undifferentiated soft sources have the same scaling coefficients as hard sources (see Table 4.2), thus, as far as the magnitude of the soundfield is concerned, results can be appropriately deduced from the hard source example. The simulation was repeated for three sample rates, namely $f_s = 8$kHz ($X = 74.37$mm), $f_s = 12$kHz ($X = 49.58$mm) and $f_s = 18$kHz ($X = 33.05$mm). A MAXFLAT pulse-shaping filter of order $N = 16$ and $f_c = 600$Hz was used in all simulations, ensuring that anomalies do not occur due to differences in the shape of the excitation signal. For the PCS method, the surface area of the sphere modelled by the mechanical filter was also held constant at a value of $a_0 = 0.1208\text{m}^2$ for all three sample rates. This is feasible as a_0 is a physical quantity describing the source, rather than a numerical property of the FDTD method.

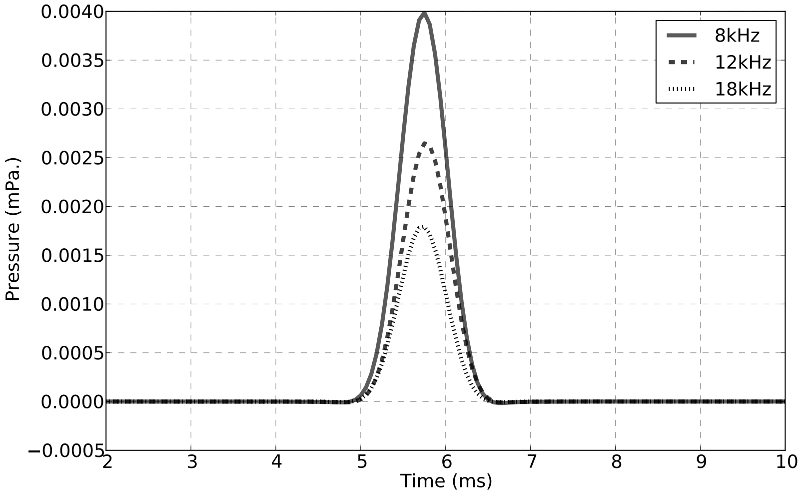


Figure 4.5: Pressure at the receiving position for a grid excited by a hard source, shown at three different sample rates.

It can be seen in Figures 4.5, 4.6 and 4.7 that the PCS is the only source model that maintains its numerical consistency. Nevertheless, in a one-dimensional prob-

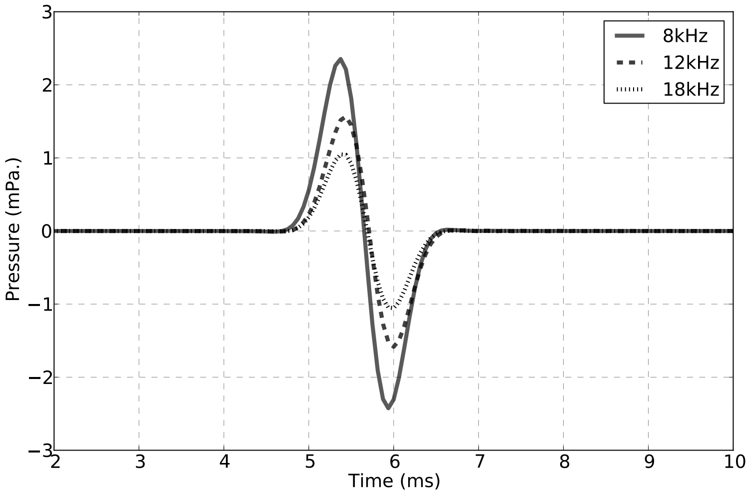


Figure 4.6: Pressure at the receiving position for a grid excited by a differentiated soft source after (Karjalainen and Erkut, 2004), shown at three different sample rates.

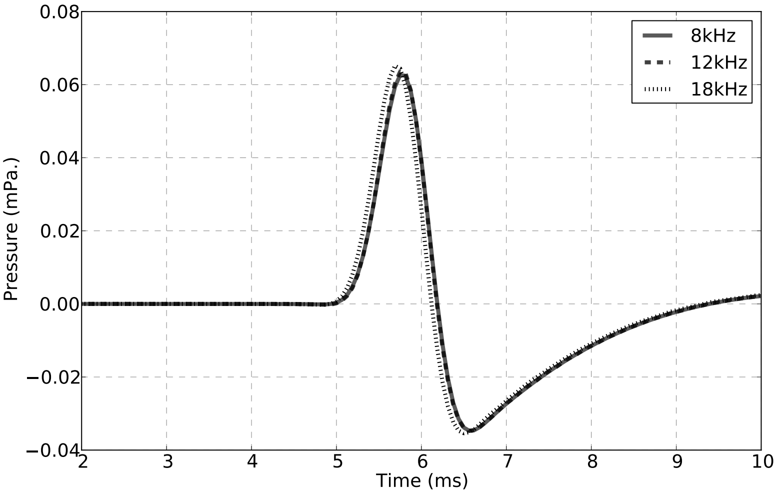


Figure 4.7: Pressure at the receiving position for a grid excited by a physically constrained source, shown at three different sample rates.

lem, one would expect similar consistency for the case of a differentiated soft source, when it is appropriately scaled as described in (Karjalainen and Erkut, 2004). Additionally, some small deviations in time are evident in all source models. These deviations are attributed to spatial quantisation errors due to the different cell sizes being used.

4.6.3 Frequency Response

To study how the different models modify the frequency response of the system, an interpolated wideband scheme was excited using a hard source[1], a differentiated soft source[2] and a physically constrained source, in three different simulations. All simulations used a MAXFLAT pulse-shaping filter with $f_c = 0.25 f_s$, and the PCS was designed with a low resonance at $f_0 = 0.167 f_s$ and $Q = 0.7$. For visual clarity, simulation outputs were normalised such that the peak value of each resulting impulse response is unity. As seen in Figure 4.8, the soft source suffers from a severe roll off in low frequencies, which is to be expected due to differentiation (be it inherited the source formulation or in the grid update equations in the case of a Yee method). It is tempting to think that in wave equation methods, one could directly inject the undifferentiated excitation signal into the grid, thus avoiding the problems of scattering and differentiation artefacts altogether. However, as will be shown in the following section, failure to adhere to the differentiation constraint may be detrimental to the result.

4.6.4 DC Artefacts

In Section 4.4 it was theoretically established that any additive source whose grid signal $s_g\big|^n$ contains a DC component, would result in a growing solution. It is important to point out that the injection filter $H_i(z)$ of source methods adhering to the differentiation constraint, automatically filter any DC from the excitation signal $s_p\big|^n$ which is passed through it. Thus, a concern arises when one uses an arbitrary soft source, such as described by Equation (4.13), where the grid signal

---

[1] Also representative of the frequency response of a transparent and an undifferentiated soft source

[2] Also representative of a Yee-type soft source.

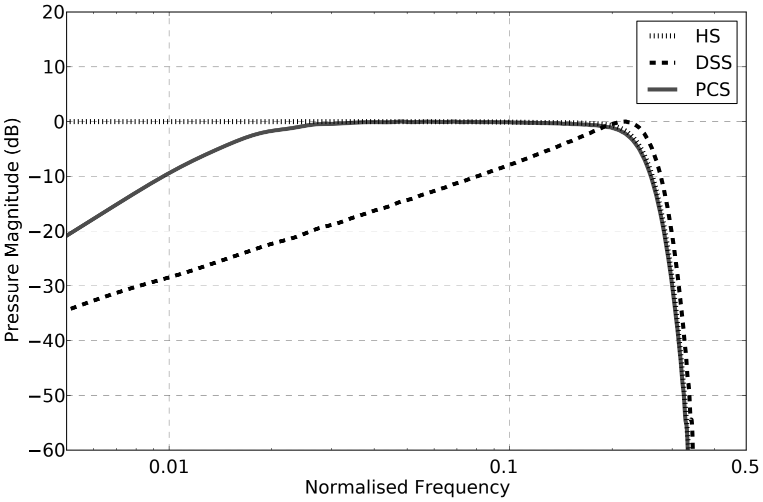


Figure 4.8: Calculated frequency response for three different source models, HS- hard source, DSS- differentiated soft source, PCS - physically constrained source. Excitation signals are MAXFLAT impulse responses of $N = 16$ and $f_c = 0.25 f_s$. PCS resonance is at $f_0 = 0.167 f_s$.

directly equals the excitation signal, and as such, may contain energy around DC. To test this, let us consider two source models, an arbitrary soft source and a physically constrained source, both of which are designed using a Gaussian pulse shaping filter, as described by Equation (4.15). Here, the Gaussian filter is chosen as it is a unipolar function, hence featuring a strong DC component. Simulations were executed for 400ms and results were recorded 1.5m from the source, as seen in Figure 4.9. For visual clarity, results were normalised such that the direct component of the soundfield equals 1Pa. in both cases. It is evident that the PCS maintains its average symmetry around the x axis, whereas the soft source solution is linearly growing. This growth is attributed to the accumulation of DC in the soundfield, and is unrelated to stability issues which normally cause exponential growth.

Such behaviour is also sensible from a physical perspective, as a DC component in $s_g\big|^n$ indicates that $q(t)$ is not of finite length, meaning that the source endlessly generates a volume velocity. To explain this, let us discuss the physical meaning of using a Gaussian as an excitation signal in an undifferentiated soft source model. Since such a soft source does not adhere to the differentiation constraint nor any other mechanical constraints, then the grid signal $s_g\big|^n$ becomes a

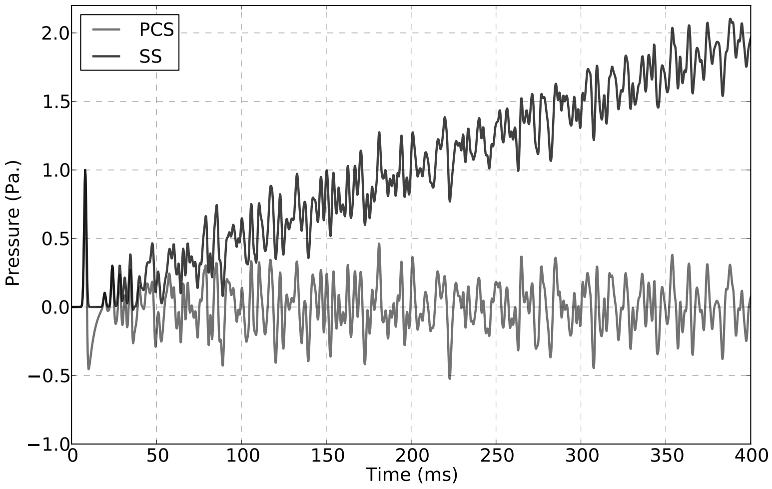


Figure 4.9: Sound pressure at the receiving position for a grid excited by an undifferentiated soft source (SS) and a physically constrained source (PCS). Both source models employ a Gaussian pulse shaping filter. Results are normalised for visual clarity.

direct numerical representation of $\psi(\mathbf{x},t)$, thus

$$s_g\Big|^n = s_p\Big|^n \equiv \psi(\mathbf{x},t)\Big|_{t=nT} \tag{4.41}$$

Since $\psi(\mathbf{x},t)$ is defined as the first time derivative of $q(\mathbf{x},t)$, then following Equation (4.8), the rate of fluid emergence due to the soft source is obtained by taking the integral of a Gaussian function, which yields

$$q(\mathbf{x},t) = \int \psi(\mathbf{x},t)dt = \int A_p \exp\left(\frac{-\left[t-t_0\right]^2}{2\sigma^2}\right)dt = \sqrt{\frac{\pi}{2}}A_p\sigma\textsc{erf}\left(\frac{t-t_0}{\sqrt{2}\sigma}\right) \tag{4.42}$$

where $\textsc{erf}(\cdot)$ is the Gauss error function. Figures 4.10 and 4.11 depict $\psi(t)$ and $q(t)$, for such an undifferentiated soft source and for a physically constrained source.

For the PCS, both $q(t)$ and $\psi(t)$ start at zero and decay to zero, indicating a finite source. However, this is not the case for the arbitrary soft source. The fact that the grid signal represented by $\psi(t)$ seems time-limited can be misleading, as it only physically means that the source generating mechanism does not accelerate before or after the excitation period (see the physical definintion of ψ in Section

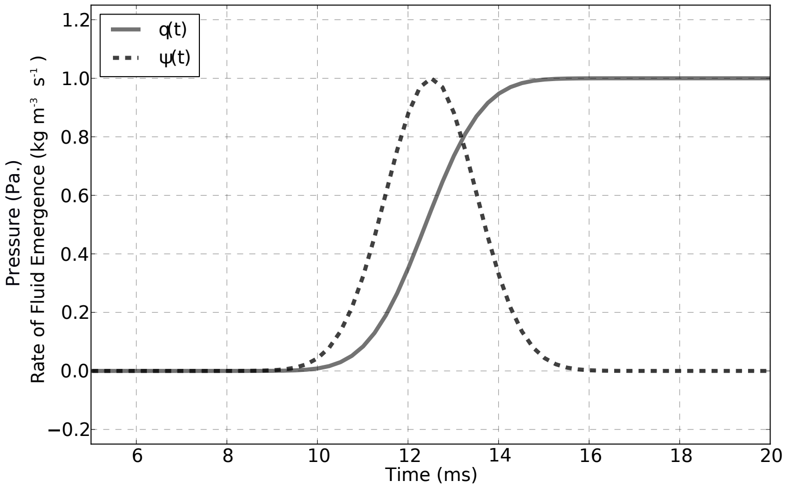


Figure 4.10: Source function, $\psi(t)$ and rate of fluid emergence, $q(t)$ at the source node, for an undifferentiated soft source excited using a Gaussian pulse-shaping filter. Results have been normalised to 1Pa. and $1\text{kgm}^{-3}\text{s}^{-1}$ for visual clarity.

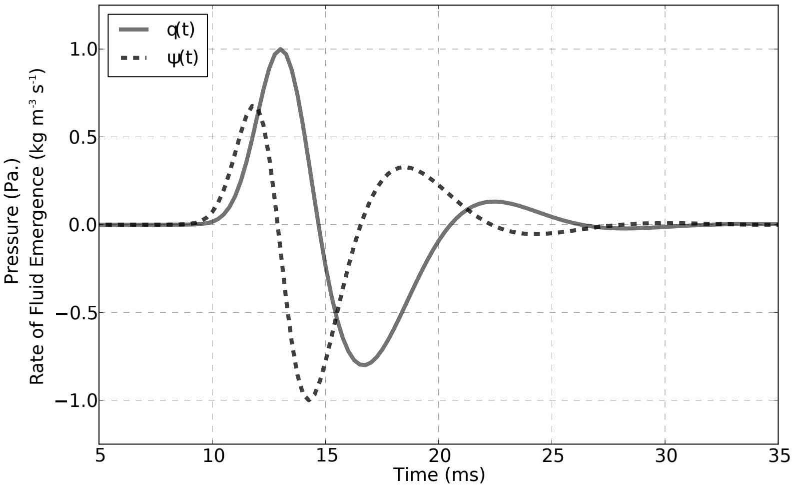


Figure 4.11: Source function, $\psi(t)$ and rate of fluid emergence, $q(t)$ at the source node, for a physically constrained source excited using a Gaussian pulse-shaping filter. Results have been normalised to ±1Pa. and $\pm 1\text{kgm}^{-3}\text{s}^{-1}$ for visual clarity.

4.3). This, of course, does not mean that the source is not active. In fact, it can be seen that when $\psi(t)$ decays, $q(t)$ rises and stays at a constant value through the remaining simulation period. This indicates that even when $\psi(t)$ is time limited, the source mechanism may still generate volume velocity. As one would expect, $q(t)$ remains at a constant positive value which is equivalent the generation of DC flow, meaning that the soundfield continuously gets pressurised by the source.

4.7 Discussion

The PCS model introduced in this chapter provides a means to design sources that adhere to all required physical and numerical constraints. In a free field, results are in good agreement with the closed-form solution to the wave equation, which underlines the physical basis of the approach. Discrepancies are mostly attributed to the numerical artefacts of the system, which include spatial quantisation and dispersion. Both of these can be greatly reduced by means of oversampling, choosing an appropriate numerical scheme, and/or spatial interpolation. To demonstrate the practical applications of the PCS model, the system was designed in conjunction with a MAXFLAT pulse-shaping filter, which has a flat transition band, is time compact and finite by definition. Since the PCS model adheres to the superposition constraint, it does not suffer from source scattering issues nor from low frequency artefacts which are evident in hard and transparent sources. Thus, as seen in the results, the PCS model fulfils the requirements from an FDTD source, as were established in Section 4.4.

The importance of scaling and differentiation has been clearly identified in this work. Correct magnitude scaling of the excitation signal is essential for obtaining consistent results when numerical parameters are altered. More importantly, the magnitude scaling of the PCS model yields results that converge well with the analytic solution, which is not the case for other sources. The model inherently handles differentiation of the excitation signal, which eliminates any existing DC component from the grid signal, thus avoiding any undesirable solution growth.

Following Equation (4.8), the source function in the frequency domain is

$$\Psi(\omega) = j\omega Q(\omega) \tag{4.43}$$

which is null for $\omega = 0$, and requires, from the Fourier differentiation theorem, that $q(\mathbf{x}, \pm\infty) = 0$. This condition is satisfied if the pulsating sphere represented by the PCS mechanical filter, begins at its resting position and returns to that position after the excitation period. Observation of Equation (4.29) shows that

$$\lim_{t \to \infty} Me^{-\alpha t} = 0 \quad , \quad \alpha > 0 \tag{4.44}$$

therefore if $F(t)$ begins and ends at zero and α is positive, then both $q(t)$ and $\psi(t)$ are appropriately time-limited, as has been shown to be the case for the PCS. A suitable $F(t)$ can be obtained by employing a pulse shaping filter having a finite impulse response, and a positive α simply means that the mechanical system must be damped. This affirms that the PCS mechanical filter acts as a natural DC blocking filter, however, similar results can be achieved when employing the most general form of a mechanical filter, as was described in Section 4.5.

Similar issues concerning scaling and differentiation have been briefly discussed in (Karjalainen and Erkut, 2004). However, being drawn from 1D digital waveguide theory, the proposed scaling factor is inadequate for 3D schemes and hence does not generate a prescribed pressure. Furthermore, the soft source does not explicitly define a mechanical nor a pulse shaping filter. Therefore, it does not fulfil the flatness requirement (if it is differentiated) or the mechanical constraint preventing solution growth (if it is undifferentiated). In the PCS model, the mass reactance of the sphere acts as an integrator which, in a physical manner, counters the effects of differentiation. Below its resonant frequency, the system is stiffness dominated, and as such, naturally acts as a DC-blocking filter. The result is a source having a near-flat pressure spectrum whose physical properties can be freely chosen by adjusting Q and ω_0. Thus, the PCS model adheres to physical laws but is not limited by real-world engineering constraints.

In comparison to hard sources, the spectrum of the PCS is not flat down to DC, however the low-frequency roll off is essential as it has been shown that a DC component is undesirable. Whilst hard sources fulfil the requirement of a prescribed pressure, they do not adhere to any physical constraints, and as such do not meet the transparency (i.e., no scattering) and numerical consistency requirements. Transparent sources emulate hard sources while avoiding source scattering and thus meeting the transparency requirement. However, they do not adhere to any transduction or mechanical constraints, and are considerably more difficult to implement and computationally expensive in comparison to the PCS.

In (Jeong and Lam, 2012), some low frequency artefacts have been identified in hard and transparent sources when excited by Gaussian functions. The authors showed that these artefacts occur due to the inability of air to rarefact at the source node as it gets replaced by the source function. In this work, it was shown that the imposition of pressure on the source node is in contrast with the superposition constraint dictated by the physical laws governing fluid emergence in an acoustic system. Jeong and Lam address this problem by employing a sine-modulated Gaussian as an excitation signal. This solution comes at the cost of meeting the flatness requirement, as the sine-modulated Gaussian filter features a roll off at low frequencies (see Figure 4.2), although somewhat less pronounced than that of differentiated sources (see Figure 4.8), which naturally occur with Yee type soft sources. However, their solution relies only on changing the pulse-shaping filter used with hard or transparent source injection methods, and as such, does not meet the numerical consistency requirement whilst still compromising the flatness requirement to some extent. In fact, the low frequency problem can be avoided altogether by adhering to all specified physical constraints, which is achieved by the PCS model. Furthermore, Jeong and Lam suggest that the transparency requirements can be met by time-limiting the source function, however, in wave equation methods the grid signal represents fluid emergence due to volume acceleration (see Figure 4.10), and as such, even when grid signal has returned to zero, any abrupt time limiting may cause significant discontinuity errors.

4.8 Conclusion

In this chapter, physical and numerical constraints in FDTD sound source modelling were studied. Five constraints were identified, namely, scaling, differentiation, superposition, mechnical, bandwidth and finiteness. Typical requirements of sound sources for room acoustics simulation were discussed, and existing source models were reviewed in light of these requirements and constraints. A novel cascaded-filter approach was proposed to generalise the process of FDTD source modelling, and existing source models have been shown to be special cases of this approach. Then, an optimal combination of filters referred to as a PCS model, was suggested. This model can be thought of as a *discrete point-monopole* which is constrained by physical laws. Two systems govern this source model, the first being the mechanical pulsation of a small sphere, and the second being the transduction of motion into acoustic pressure. To the authors' knowledge, this is the first time where these two systems have been integrated into a single physically-plausible model, at least in context of FDTD simulation.

This approach offers various benefits over existing source models. It provides a means to design and inject sources which generate a prescribed pressure field, do not scatter incoming waves, and have a flat frequency response without causing any low-frequency artefacts. Furthermore, the proposed PCS model is correctly scaled with FDTD parameters, and thus is numerically consistent across different sample rates, which is not the case for any of the previously suggested source models.

—We shape our buildings;
thereafter they shape us.

Winston Churchill

5

Practical Aspects of Room and Listener Modelling

Whilst the roots of the FDTD method are dated a few decades back (Yee, 1966), in acoustics modelling it is not considered as mature as geometrical methods, or frequency domain wave methods such as FEM and BEM. In room acoustics research, there are some documented cases where the FDTD method was used to study problems involving sound propagation in enclosed or semi-enclosed spaces, e.g. see (LoVetri et al., 1996; Drumm et al., 2008; Sakamoto et al., 2008; Collecchia et al., 2012; Lokki et al., 2013), however the method is not widely used in the industry, most likely due to the unavailability of a commercial or community-supported software package. Ideally, such a software package would allow its user to:

1. Model wave propagation in a domain to which frequency dependent bound-

ary conditions can be applied, with minimal numerical errors, and at a reasonable computation time.

2. Model arbitrary room geometries.
3. Model acoustic sources as directive, frequency dependent entities.
4. Capture acoustic quantities, such as pressure and particle velocity, at single or multiple receiving positions.
5. Capture binaural room impulse responses (BRIRs) that may be directly used for auralisation.
6. Visually inspect the time varying soundfield, in a clear and efficient way.

As part of this research, the author has developed computer software, termed *WaveCloud*, which apart from serving as the computational engine used in this thesis, is intended to serve as basis for an open source FDTD project, as is discussed in Section 5.1. This chapter deals with the practical aspects of realising the FDTD method for room acoustics simulation.

Before proceeding, it is useful to review the above requirements in light of current knowledge of acoustics FDTD. Aspects 1 and 2 relate to numerical dispersion which is the most evident shortcoming of finite difference methods, and as such have received much attention from the research community (Savioja and Valimaki, 2000; Christopher and Schneider, 2005; Kowalczyk and van Walstijn, 2011; Potter et al., 2011; Bilbao, 2012). Whilst such efforts have resulted in highly optimised numerical schemes, their integration in a complete room acoustics simulator may still present a technical challenge. For example, most of these methods have only been thoroughly tested using simple rectangular boundaries. As will be further discussed in Section 5.3, it is the author's experience that for geometrically complicated objects, the simplest non-interpolated standard rectilinear scheme is the most trouble-free method, even if it requires significant oversampling in order to reduce its inherent dispersion. Accordingly, these aspects are addressed in WaveCloud by employing high-performance computing technology based on general purpose graphics hardware (GPGPU). This is discussed in Section 5.2, which in part relies on material published by the author in (Sheaffer and Fazenda, 2010).

The issue of modelling directional sources has been studied in (Hacihabiboglu et al., 2008) for the DWM method, however, this approach has not been adapted for use in the FDTD method yet. Escolano and colleagues (2007) have proposed a method to model directive sources in FDTD methods, by employing an array of sinusoidal monopoles in the near field of a virtual source. The method has been extended to using broadband sources in (Escolano et al., 2009), and requires preprocessing in the frequency domain in order to calculate the different weights for the near field array. In (Southern and Murphy, 2009) the weights for first order directive sources have been discussed, effectively providing an efficient means of modelling sources of simple directivity patterns. The method, however, has not been generalised for sources whose directivity is frequency dependent, which is needed in many practical cases. In Section 5.5, a generalised approach of solving frequency dependent problems in FDTD methods is proposed, which is predominantly based on material published by the author in (Sheaffer et al., 2011). The approach, which has been implemented in WaveCloud, provides a simple way to turn frequency independent sources and boundaries into frequency dependent ones, at the expense of having to run multiple simulations.

Another challenge in FDTD simulation of room acoustics, is modelling binaural receivers (aspect 5). Unfortunately, this is not as straightforward as in geometrical methods, as the wave nature of the method and the need to bandlimit the grid results in temporal overlap between the various reflected components of the soundfield. Thus, in order to convolve each reflection with a corresponding pre-measured HRIR, one would first have to decompose the response into its discrete reflections which is non-trivial and requires post-processing. A versatile yet indirect method is to capture and spatially encode the output of an FDTD model (Southern et al., 2012). Using this approach, n^{th}-order ambisonic impulse responses are obtained, which can be further post-processed to generate auralisations.

A simpler way to approximate a binaural receiver, is to record the sound pressure at two spaced receivers. However, in the absence of an efficient shadowing object between the two probes, interaural level differences (ILDs) are not correctly reproduced, and interaural time differences (ITDs) are only roughly estimated. Murphy and Beeson (2003; 2007) have shown that in a 2D Digital Waveguide Mesh, approximation of correct ITD values is possible by embedding a circular object between two spaced receivers, however the applicability of this to ILD cues has not been looked at. A more complete way of modelling HRIRs, including all binaural and monaural cues, was successfully achieved in (Xiao and Liu, 2003; Mokhtari et al., 2008) using specialised FDTD schemes for propagation of sound in inhomogeneous media. Whilst their results are in good agreement with measured responses, employing such an approach in room acoustics FDTD is challenging. First, the spatial period required to achieve such accuracy is in the order of a single millimetre which is computationally expensive given a uniformly sampled room. Furthermore, it is not clear whether sufficient accuracy can be achieved using efficient numerical schemes more commonly used for room acoustics simulation, as they are principally different and normally employed in a coarser grid setup. Accordingly, section 5.4, which is partly based on material published by the author in (Sheaffer et al., 2013), discusses the merits and limitations of obtaining BRIRs by means of embedding listener geometries in an FDTD grid. Lastly, Sections 5.3 and 5.6 complement the aspects of generalising the FDTD method for use in room acoustics, by briefly discussing methods for handling arbitrary geometries and for visualisation of soundfields generated in WaveCloud.

The purpose of this chapter is twofold. First and foremost, it is establishes the underlying methodology for obtaining the BRIRs used in Chapters 6 and 7. An additional goal is to complement the current know-how of implementing a general purpose FDTD-based room acoustics simulator. Accordingly, some other aspects such as source directivity and frequency dependent boundaries, which are not directly used in this thesis but are important for the development of the WaveCloud project, are also covered.

5.1 The WaveCloud Project

One of the strengths of the FDTD method is that it's core algorithm is simple enough for many scientists and engineers to implement, hence making it a good candidate for a collaborative, open-source computing project. Implementing more advanced features may require expertise and a deeper understanding of the method, thus having a readily available core FDTD engine would encourage researchers to focus on more specific features of an FDTD simulator. Recently, a freely available FDTD toolbox was released (Hill and Hawksford, 2011), however it is aimed at solving low frequency problems and therefore lacks many features required for more general applications in room acoustics. Furthermore, this toolkit was not available at the time this research commenced. Other software packages reported in the literature (Beeson and Murphy, 2004; Lopez et al., 2007; Lokki et al., 2011) which potentially include more advanced features, unfortunately, have not been made available to the general research community. An exception to this is a computer program reported in (Drumm and Lam, 2007; Drumm, 2007) which is available for download, however its source code was not made open.

In this section the general structure of *WaveCloud*, the software package developed by the author, is presented. At this time, the software was only used for research conducted at Salford University, however, the code is intended to serve the basis of an open-source, community-supported project, which would potentially be of interest not only to those working to improve the FDTD method, but also for individuals conducting research in the wider domain of room acoustics.

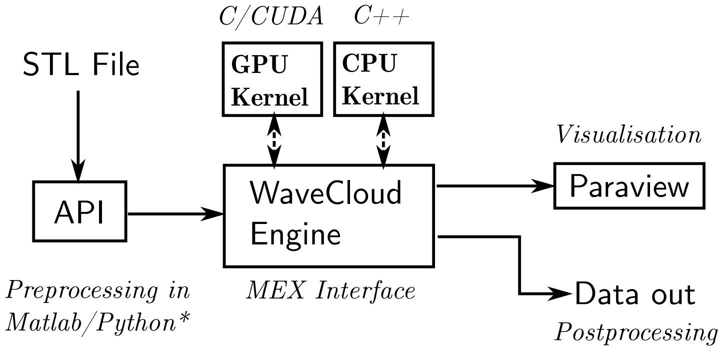


Figure 5.1: General structure of the *WaveCloud* open software package. *Python API is work in progress.

The WaveCloud project, whose general structure is shown in Figure 5.1, is a multi-environment software package, which is intended to integrate with the typical workflow used in scientific research. The WaveCloud engine provides an *Application Programming Interface* (API) which can be directly used inside Matlab (a Python API is currently work in progress), thus allowing the user to easily write scripts for FDTD simulation. A sample WaveCloud script is shown in Appendix A. The API also provides functions to load, voxellise and embed arbitrary geometries in the FDTD grid, based on predesigned CAD drawings saved in the STL file format (cross-conversion from other popular vector formats is straightforward). Inclusion of numerous sources[1] is optional, making it possible to model some more composite sound reproduction paradigms such as line arrays, and wavefield synthesis.

Accordingly, the user defines all required simulation parameters inside a Matlab *struct* object and then passes it on to the computational engine. Depending on the available computer architecture, the engine would then execute the simulation either on the hosting computer's CPU or if possible, on a dedicated GPGPU according to the principles discussed in Section 5.2. Once simulation has completed, the engine outputs variables containing the recorded pressure and particle velocity at the defined receiving positions. Because computation does not happen in real time, then when visualisation is opted for, the model saves the data representing the time varying soundfield, which is then opened in an external visualisation application.

5.2 Oversampled Implementations on a General Purpose GPU

As was discussed in Chapter 3, the most prominent shortcoming of the FDTD method is numerical dispersion. Even with highly optimised numerical methods, the bandwidth of the method is limited. For example, with the interpolated-wideband scheme (Kowalczyk and van Walstijn, 2011), the 2% dispersion error

---

[1] The total amount of possible sources is only limited by the physical size of the modelled domain.

limit is at a frequency of $0.186f_s$ both in terms of accuracy and isotropy. Thus, to be able to simulate the entire audible spectrum, one would need a sample rate of over 100kHz. With the more widely known standard rectilinear scheme, achieving this would require a sample rate of over 250kHz, as the dispersion limit is at $0.075f_s$. Thus it appears that even when considering the state of the art in room acoustics FDTD, oversampling is required in order to yield accurate results in a wide band of frequencies.

Naturally, oversampling results in a significant increase in problem size which requires more computational resources. Being grid-based and time-stepping, the explicit FDTD method is highly suitable for parallelisation, which is possible using multicore computer architectures. Harnessing the power of graphics processors for general purposes (GPGPU) is gaining high popularity in many computational sciences, with recent evidence of successful finite difference applications in electro-dynamics (Sypek et al., 2009) and seismic modeling (Michéa and Komatitsch, 2010). The introduction of the CUDA programming language (Nvidia, 2008a) provides an attractive solution for implementing algorithms on GPGPUs at a comfortable learning curve, hence allowing scientists and engineers to easily and quickly adapt their codes to run on graphics hardware. In acoustics, an implementation for 2D structures has been proposed by Southern and colleagues (2010), and more recently focus on 3D problems has been discussed in (Sheaffer and Fazenda, 2010; Savioja, 2010; Webb and Bilbao, 2011). Some authors have also suggested novel numerical methods which are related to FDTD but more efficiently exploit the strengths of GPU computing (Mehra et al., 2012). In this section, a parallel implementation of the wave-equation compact explicit FDTD method is presented, effectively allowing for solving significantly larger problems than with the traditional CPU paradigm. The implementation presented here is used as the core part of WaveCloud's GPU engine.

5.2.1 Algorithm Overview

CUDA, or *Compute Unified Device Architecture*, is a parallel computing framework designed to be employed on Nvidia-based graphics hardware. One of CU-

DAs main advantages is its accessibility to a wide range of programmers, as using it does not require prior knowledge or experience in developing applications for graphics hardware. Any CUDA-enabled GPU can be programmed using a high-level language termed CUDA-C, which is essentially C with GPU extensions and some restrictions. Programmers may opt to write their code directly in CUDA-C or make use of wrappers, binders and plugins currently available for Python, Matlab and Fortran amongst others.

In Figure 5.2, A CUDA program flow is portrayed for a typical time-iterative algorithm such as the explicit FDTD method. First, memory is allocated on the host computer alongside any other preparations needed. Once the data have been prepared, it is copied onto the device's global memory for further processing. Next, a loop on the host executes a GPU kernel in each time step. The kernel is designed to compute values for the entire soundfield and, for the wave equation method, requires knowledge of the soundfield at z^{-1} and z^{-2}. Once the entire computation cycle is completed, the required data is copied back to the host memory for further post-processing and output.

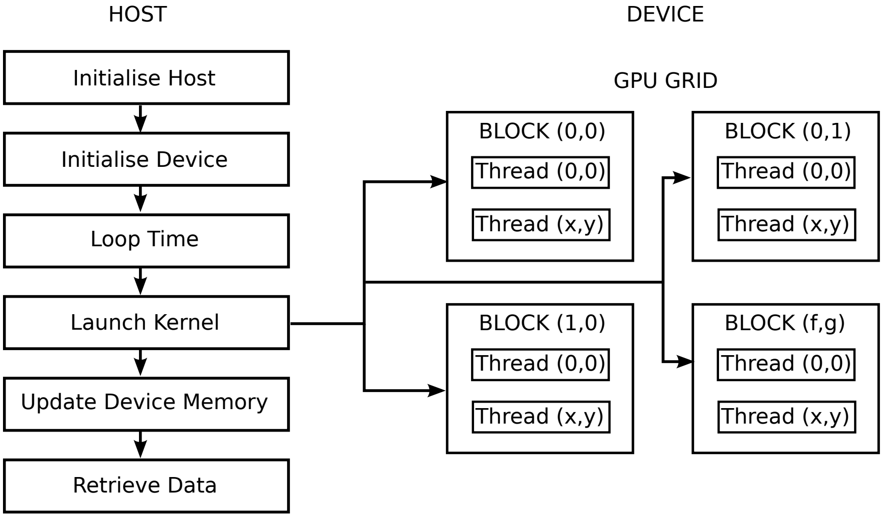


Figure 5.2: Schematic representation of a parralelised FDTD algorithm.

5.2.2 Thread and Memory Hierarchy

The GPU programming architecture is based on a *Single-Instruction, Multiple-Thread* (SIMT) paradigm, meaning that a predefined set of computational operations is used on a large data set. In practice, this is achieved by executing the code on a scalable array of streaming multiprocessors (Nvidia, 2008a), each containing $8 - 48$ CUDA cores depending on the *compute capability* of the GPU. From a logical point of view, the programmer creates a *kernel* defining a set of explicit operations that are performed unanimously on the data. Whilst these operations are performed serially, a launched kernel activates a group of threads which are executed in parallel. A kernel can therefore be seen as a set of instructions and an active thread would correspond to that set of instructions performed on a designated part of the stored data. In an ideal parallel environment, the number of possible active threads would be comparable to the size of the data, that is, one thread per data point. In reality, the amount of active threads is explicitly limited by hardware architecture, hence requiring a mechanism to schedule and synchronise their operation.

From the programmer's viewpoint, threads are grouped into *blocks* which are further clustered in *grids*. When a kernel is launched, data in a grid is processed block by block, invoking its own group of threads each having a unique identifier code. From a physical point of view, once a kernel has been launched, active threads in a block execute concurrently on a multiprocessor. The multiprocessor is therefore in charge of administering thread execution, a process which is typically done in groups of 32 threads, commonly known as *warps*. Whilst warps may execute independently, their operation is controlled by a warp scheduler on the multiprocessor. An important point in case is that warps are designed to execute a single common instruction, so maximum computation efficiency is achieved when all threads in a warp follow a unanimous operation scheme. Divergence and branching due to thread-distinctive instructions results in serial execution, which in turn reduces efficiency. An example of this would be the inclusion of conditional statements within a kernel, which would require different threads within a warp to perform different operations at arbitrary memory locations.

In a traditional CPU programming paradigm, computational operations are thought of as ”expensive” whereas memory access is considered ”cheap”. In other words, programmers prefer to pre-calculate any elements that are intended for re-usability and store them in memory until needed. Conversely, in a GPU programming paradigm, memory transactions take longer to perform than simple arithmetic. It is important to identify four types of memory available in a GPGPU programming environment. *Host memory* is the term used to describe the RAM provided by the hosting computer, *global memory* is the RAM available on the GPU, *shared memory* is a specifically dedicated 16KB of cached memory that can be shared between threads inside the same block, and *local memory* is a small amount of cached and thread-specific memory.

It is often necessary to control memory usage on a thread or warp level to ensure maximum performance. Memory coalescing is a common technique to minimize memory latency for a given set of instructions, allowing access to 16 words in a single memory transaction. In general, three conditions must be met:

1. Data should be stored in 4, 8 or 16-byte words.
2. All accessed words must reside in the same memory segment.
3. Words must be sequentially accessed by threads in a half-warp.

With recent progress in GPU technology memory coalescing is becoming less and less of a concern, however, it is still beneficial for ensuring backward compatibility for users who still posses older GPUs.

5.2.3 Processing 3D Data Structures

At the time that this part of the study was conducted, CUDA technology allowed implementation of only 1D and 2D grids, meaning that 3D data structures had to be remapped in order to be accessed. In (Sheaffer and Fazenda, 2010), the author has suggested, implemented and compared two methods to map 3D data structures onto 2D grids, allowing for an efficient implementation of the 3D FDTD method on a GPGPU. Since the introduction of CUDA 4.0, it is natively possible to calculate 3D grids and the code for the computational engine used in this thesis was adapted accordingly. Thus, the proposed method is no longer directly relevant to this work, and as such, have been omitted from this chapter. Interested readers may refer to (Sheaffer and Fazenda, 2010) for further details[2].

---

[2] The document may be accessed online at `http://usir.salford.ac.uk/11568/`

5.2.4 Algorithm Optimisation

In order to exploit the benefits offered by GPGPU technology, the FDTD implementation needs to be adapted according to the GPU programming paradigm. The first and foremost concern is memory access, and as such, memory transfers between the host and the device should be kept to a minimum. Depending on the size of the domain and the sample rate required, it may be possible to preallocate all the required memory on the device. Otherwise, the domain needs to be solved in a number of data batches, which greatly reduce efficiency due to a lower computation-communication ratio. In such a case, it is advisable to perform a small number of large memory transfers, rather than many small ones. It should be pointed out that whilst the FDTD formulation for the wave equation indicate that three matrices are used to describe the soundfield at three consecutive time steps, it is possible to reduce memory requirements by 1/3, by overwriting the memory allocated to $p\big|_{\mathbf{i}}^{n-1}$ with the value solved for $p\big|_{\mathbf{i}}^{n+1}$. This is possible as $p\big|_{\mathbf{i}}^{n-1}$ is only required at the position $\mathbf{i}$, and read-after-write errors can be safely avoided.

Additionally, any extraneous memory accesses should be eliminated, which in context of an FDTD algorithm normally correspond to the various coefficients used in the explicit update equations. Since the GPU favours computation over communication, these should be recomputed at each kernel launch, rather than precomputed and stored in memory. This is particularly true as calculating the FDTD coefficients involve only basic arithmetic. If more complex computations are needed, for example in the case of solving boundary impedance filters, then shared memory should be used as it has a faster throughput than global memory (Nvidia, 2008a). The ideal computation-communication ratio may not be straightforward to predict, and the use of dedicated profiling tools are often required to fully optimise the algorithm. Such tools, which are provided with CUDA, allow an effective assessment of thread occupancy, that is, the ratio of active warps per multiprocessor to the total possible active warps (Nvidia, 2008b). The variables affecting occupancy, namely thread block sizes, should be adjusted such that the maximum amount of warps would be active. An important point in

Table 5.1: Numerical domain size (total number of nodes) and computation times for a CPU and a GPU as function of sample rate for a standard rectilinear grid solved at the Courant limit (i.e. $\lambda = \sqrt{1/3}$). Limiting frequency calculated based on 2% dispersion.

Sample Rate (Hz)	Limiting Freq. (Hz)	Nodes	CPU (s)	GPU (s)
11,025	826	54,872	7.09	1.73
22,050	$1,687$	421,875	129.47	4.86
32,000	$2,400$	1,259,712	992.65	12.39
44,100	$3,307$	3,307,949	-	36.59
65,000	$4,875$	10,503,459	-	153.57

case is that optimal performance can often be achieved even with non-maximal occupancy, however low occupancy would normally degrade performance. It is also advisable that the total number of threads per block is set as a multiple of 32.

In order to avoid thread divergence, it is useful to define task-specific kernels. In context of room acoustics, it can be normally assumed that the domain is largely empty, thus one kernel can be defined to solve the explicit update equations for air, whereas a different kernel would be used to solve the boundaries. Even after the entire domain has been solved for, any internal objects may be calculated with a separate kernel and resulting pressure can be overwritten at the respective points in the soundfield. Such approach has been taken in (Sheaffer et al., 2013), and was shown to be highly efficient.

5.2.5 Results and Discussion

To exemplify the benefits of GPU computing to the FDTD method, an empty domain of 2x2x2m was solved on a GPU (Nvidia GTX-580 1.5GB RAM) in single precision arithmetic and at five different sample rates, according to the guidelines presented in this section. For reference, the same grid was solved on a CPU (Intel Core i5 2.8GHz). The numerical domain sizes, limiting frequencies (for to 2% dispersion) and resulting computation times for the two cases are detailed in Table 5.1, and graphically shown in Figure 5.3. Results from the CPU are not given above 32kHz as the computation time was too long.

It can be seen that the higher the sample rate is, the more one benefits from

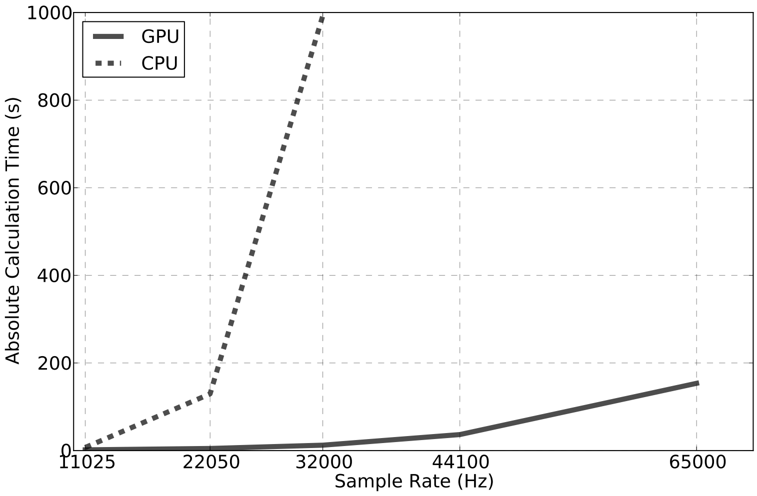


Figure 5.3: Computation times of FDTD model using a GPU and a CPU.

executing the code on the GPU. More specifically, at 11, 025Hz the computation time ratio of the GPU to the CPU is only 4.09, whereas at 32kHz the ratio increases to 80.09. This is to be expected, as the GPU architecture becomes more efficient as the amount of unanimous operations increase. Increasing the sample rate means that the room's volume is represented by more nodes, and since the same kernel operates over all air nodes, then efficiency increases. However, in rooms of non-rectilinear geometry, the efficiency would be constrained by the ratio of volumes occupied by air (efficient) and boundaries (less efficient), as was discussed in (Savioja, 2010; Sheaffer et al., 2011). The GPU used in this investigation (Nvidia GTX-580 1.5GB RAM) is a high-end consumer GPU, whose manufacturer-suggested retail price was less than £300 at the time of writing this chapter. Thus, it is clear from this investigation that even with a modest financial investment, significant improvement in computation speed can be obtained when employing GPGPU technology. In context of this thesis, the use of a GPGPU is essential for generating impulse responses at 65kHz allowing for the perceptual model described in Chapter 6, to be applied at high frequencies as well as at low frequencies.

5.3 Handling Arbitrary Room Geometries

Computer-aided architectural designs are usually stored in vector graphics format, which normally contains vertex locations defining the geometry of different objects. Thus, solid objects are not represented directly and are rendered from the vector definitions by the graphics software being used. In order to include such a data format in an FDTD simulation, it is required that the model geometry is discretised according to the spatial parameters of the grid (Drumm and Lam, 2007). The process of converting a vector definition into such a discrete volumetric representation is known in the field of computer graphics as *voxellisation*, which is typically performed by means of ray-tracing, e.g. see (Patil and Ravi, 2005) for a description of the voxellisation method used in this thesis.

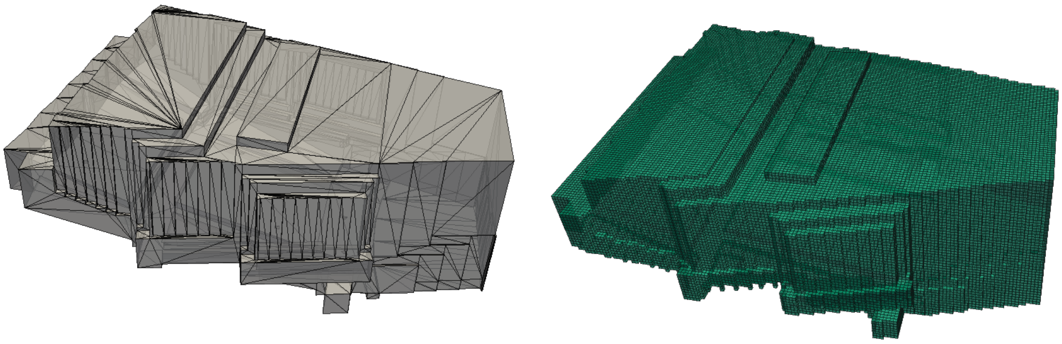


Figure 5.4: Computer model of the Elmia concert hall, which is typically used in acoustics simulation software tests (Bork, 2000). Left - original vector model, Right - voxellised model with a spatial resolution of $X = 29.75$mm, correspoding to a sample rate of $f_s = 2$kHz in an SRL scheme.

An inherent problem of voxelised models is that oblique surfaces become staircase-represented, as can be seen in Figure 5.4. If the spatial sample period is small enough in comparison to the considered wavelength, then a good approximation is possible. A more accurate approach is to employ conformal boundary models (Dey and Mittra, 1997; Tolan and Schneider, 2003), however, these are more computationally complex and significantly more complicated to implement, especially when interpolated schemes are used. In this thesis, modelled rooms involve only right-angles, and as such, a non-conformal boundary model was opted for. Additionally, even if one were to use WaveCloud to model rooms of more complex geometries, it is sensible that oversampling would compensate for the compromised accuracy in the stair-case representation of the boundaries. For example, a standard rectilinear scheme solved at $f_s = 65,000$Hz would have an upper (2% dispersion) limiting frequency of 4875Hz whose wavelength is 70.5mm. At this sample rate, the spatial period is only 9.15mm which is significantly lower than the shortest considered wavelength.

In non-conformal methods, once the model has been voxellised, each boundary node is classified according to the specific boundary condition which is applied to it and boundary impedance value (or values, in the case of frequency dependent boundaries) is assigned to it. This classification process is essential, as each type of boundary node (e.g. front face, right-back edge, front-back-left corner) involves solving a difference explicit boundary update equation (see Section 3.2.4). This approach, which is a physically correct way to solve the problem, is easily applied to most cases of model geometries. However, if parts of the model involve geometrical shapes of high complexity (such as the human head model described in Section 5.4), then one may encounter cases where it is difficult to uniquely classify a boundary node to an update equation. If such an ambiguity is not resolved, then the stability of the boundary model is compromised. This is especially evident in interpolated schemes, as for the standard rectilinear method the update equation for any re-entrant (outer) corners or edges reduces to the update equation for air (Kowalczyk and van Walstijn, 2008). Accordingly, a different boundary model is used for complex interior objects, as will be discussed in the following section.

5.4 Modelling Binaural Receivers

In this section, a method of calculating BRIRs by directly embedding a listener geometry in the grid is studied. Unlike previous studies, in which authors used specialised FDTD methods to accurately model HRIRs, here the focus is on room modelling. Thus, the idea is to assess to what extent the numerical scheme used for modelling the room is applicable to obtaining appropriate binaural cues, when an embedded listener method is used.

5.4.1 Listener Geometry

Using a rigid sphere to approximate the human head is a simple means of obtaining interaural localisation cues (Algazi et al., 2002). In context of this investigation, it is useful for two other reasons. First, lacking any spectral cues, responses generated using a rigid sphere model are easy to visually inspect, and any errors arising from the applied boundary model are easier to identify. Secondly, the closed-form solution for diffraction around a rigid sphere (see Section 2.4) serves as an exact reference for comparisons.

A more accurate representation of a listener, is an actual model of a human head. Here a laser scan of a KEMAR mannikin[3]. is used, originally used in (Kahana and Nelson, 2007), which have been polygon-reduced and voxellised as shown in Figure 5.5. The amount of spatial detail which is preserved after discretising the model is dependent on the temporal sample rate, as such, the shape of the head approaches that of a sphere as the sampling frequency is reduced. As reference to the numerical model, the measured KEMAR set from MIT (Gardner et al., 1994) is used.

5.4.2 Boundary Model

Whilst it is possible to achieve accurate geometrical representations of objects and numerically consistent results using conformal boundary methods (Tolan and Schneider, 2003), many authors use the FDTD method for its simplicity and ease

---

[3] Laser scan courtesy of Yuvi Kahana and the ISVR, University of Southampton

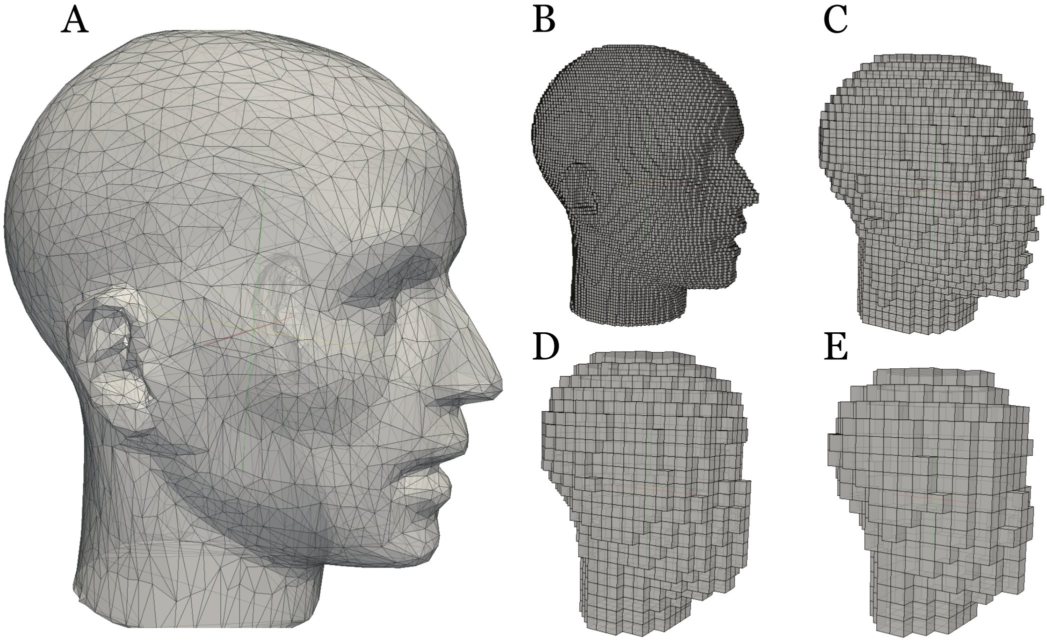


Figure 5.5: Geometry of a KEMAR mannikin used in this study. A. Polygon reduced laser scans, B. Voxellised for 176.4kHz ($X = 3.37$mm), C. Voxellised for 65kHz ($X = 9.15$mm), D. Voxellised for 44.1kHz ($X = 13.5$mm) and E. Voxellised for 32kHz ($X = 18.59$mm).

of implementation, and as such, refrain from employing complex realisations. Furthermore, many room models do not feature curved or oblique boundaries, and non-conformal methods are the appropriate choice. As was discussed in Section 5.3, a key aspect in the accuracy of numerical models is the realisation of appropriate boundary conditions, and particularly the classification of the different boundary nodes, which may be ambiguous for models of high complexity. A different approach is to model any internal objects as impedance discontinuities in the grid using an admittance boundary condition, originally derived for the digital waveguide mesh (Kelloniemi, 2005). Consider a boundary node at the discrete position $\mathbf{i} = [l, m, i]$, whose surface admittance is expressed as function of its reflection factor $\hat{r}$, by

$$Y\Big|_{\mathbf{i}} = \frac{1 - \hat{r}}{1 + \hat{r}} \tag{5.1}$$

If the node is considered to be the FDTD equivalent of a lossless scattering junction in a rectilinear digital waveguide mesh (see Section 3.1.3), then the explicit update equation for the node becomes

$$\begin{aligned} p\big|_{l,m,i}^{n+1} = A(&p\big|_{l-1,m,i}^{n}Y\big|_{l-1,m,i} + p\big|_{l+1,m,i}^{n}Y\big|_{l+1,m,i} + p\big|_{l,m-1,i}^{n}Y\big|_{l,m-1,i} \\ &+ p\big|_{l,m+1,i}^{n}Y\big|_{l,m+1,i} + p\big|_{l,m,i-1}^{n}Y\big|_{l,m,i-1} + p\big|_{l,m,i+1}^{n}Y\big|_{l,m,i+1}) - p\big|_{l,m,i}^{n-1} \end{aligned} \tag{5.2}$$

where the coefficient A is given by

$$A = \frac{2}{Y\big|_{l-1,m,i} + Y\big|_{l+1,m,i} + Y\big|_{l,m-1,i} + Y\big|_{l,m+1,i} + Y\big|_{l,m,i-1} + Y\big|_{l,m,i+1}} \tag{5.3}$$

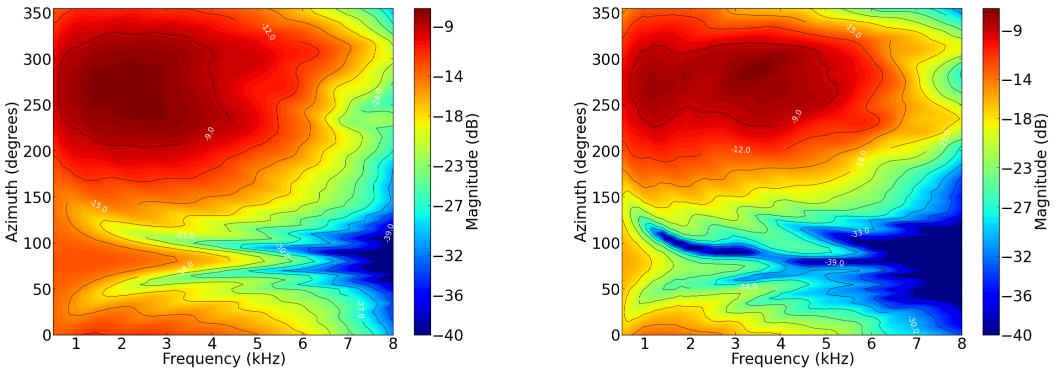
Unlike boundary models which are used to truncate the grid, this formulation necessitates that only the primary role of the node is identified. For example, if the node represents air, then its admittance is set to unity, i.e. $Y\big|_{\mathbf{i}} = 1$, and when all of its surrounding nodes are also air nodes, then Equation (5.2) reduces to the update equation for for the acoustic medium in a standard rectilinear grid. If the node represents a reflecting surface, then a low admittance value is used. In this study, a value of $Y = 1.25 \cdot 10^{-4}$ is used to represent the surface of the object approximating the head. It should be noted that as the formulation is predicated on a lossless approach, then any fraction of energy not reflected off the surface gets transmitted through the object. Given the high impedance used for the head object, the fraction of energy transmitted through the boundary is expected to have a negligible effect on the computed binaural cues.

5.4.3 Results and Discussion

As was established in Section 5.1, the technology used in this dissertation constrains simulations to a sample rate of 65kHz. For brevity, only results obtained at this sample rate are discussed here. Readers interested in a similar investigation for other sample rates are referred to (Sheaffer et al., 2013).

To assess the listener models, a MAXFLAT-based physically constrained source (f_0 = 100Hz, Q = 0.7, f_c = 6kHz) was placed at a radial distance of 1m from a binaural receiver, in a virtual domain of 27m^3 calculated using the SRL scheme. The source was rotated 360° around the listener in intervals of 1°, and the resulting pressure was recorded on the surface of the object at the positions of the left and right ears. The simulation was executed long enough for the signals to fully reach the receiving positions, but was terminated before any reflections arrived from the boundaries of the domain, hence effectively emulating an anechoic measurement. The process was repeated to obtain responses for models of a sphere as well as a laser-scanned Kemar.

The resulting left-ear HRTFs are depicted in Figure 5.6. Figure 2.9 (see Chapter 2) shows similar HRTFs for the closed form solution of a sphere and for the measured Kemar responses. Note that since HRTFs show absolute pressure, then for visual clarity reference HRTFs were convolved with the same source function used to generate the numerical results. Comparing Figures 5.6 with 2.9, it can be seen that the sphere model matches its reference quite closely, at least up to the high cutoff frequency. Divergence in the general shape of the HRTF, is especially evident in the circular high pressure area on the ipsilateral angles, which loses its shape at high frequencies in the numerical model. This is attributed to numerical dispersion, which is anisotropic and grows stronger with frequency. For the Kemar model, the numerical results resemble those of a sphere much more than the reference measurements. This is to be expected as at the sample rate of 65kHz, only the general shape of the head is represented, and the fine geometrical structures of the pinna are lost in the discretisation process.

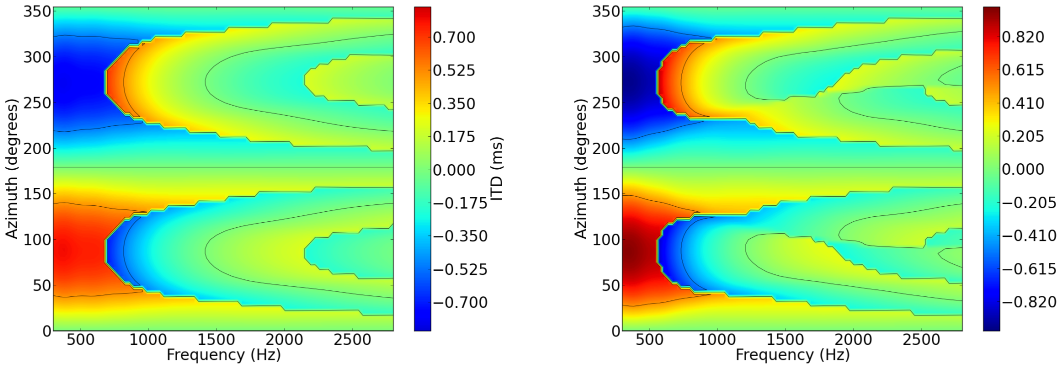


(a) HRTF - Sphere Model (b) HRTF - Kemar Model

Figure 5.6: Calculated left-ear HRTFs for a sphere model (left) and a Kemar model (right), solved at a sample rate of 65kHz.

Whilst HRTFs provide an effective means for visually inspecting the transfer function of the model approximating the head, it is also useful to evaluate how well these models perform in emulating interaural cues which are required for sound localisation. This is especially important in context of this thesis, as the perceptual model discussed in Chapter 6 is restricted to localisation on the horizontal plane, and as such, strongly relies on interaural cues. The interaural time differences for the sphere and KEMAR models, obtained using Equation 2.31, are shown in Figure 5.7. Comparing to their reference values (Figure 2.11), it can be seen that very good matching is evident for both models.

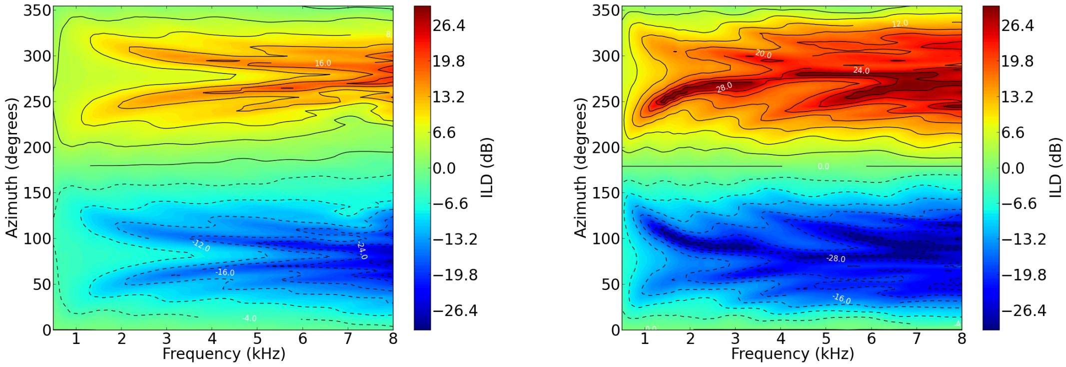
An interesting point is the highest frequency at which interaural phase differences can be unambiguously resolved by the auditory system. This frequency can be identified as the tip of the abrupt colour change in the ITD pattern, which for both reference models is situated at around 750Hz (for the two extreme lateral angles). In the Kemar numerical model, this frequency is considerably lower, at around $500 - 600$Hz. Since ITDs are calculated from phase-delays, this implies that there is not a perfect alignment between the physical dimensions of the measured Kemar mannikin and the discretised laser scans, which is possibly due to spatial quantisation errors. Nevertheless, since the perceptual model discussed in Chapter 6 involves supervised learning, then such a discrepancy is akin to differences between individual listeners.



(a) ITD - Sphere Model (b) ITD - Kemar Model

Figure 5.7: Calculated interaural time differences for a sphere model (left) and a Kemar model (right), solved at a sample rate of 65kHz.

The interaural level differences for the sphere and Kemal models is shown in Figure 5.8. In comparison to Figure 2.10, there is a generally good match for the sphere model, with some divergence at high frequencies. This can be seen in the curves around 0dB ILD, which for the reference model remain striaght with frequency at the angles of 0° and more clearly at 180°. For the numerical sphere model, these curves diverge as the frequency is increased, which implies an effect of dispersion, however, one would expect similar behaviour in the numerical Kemar model, and this is not quite evident in the results. When comparing the numerical Kemar model to its reference measurements, there is resemblance in the general ILD pattern, however many of the subtle features attributed to the fine geometry of the head, are not correctly reproduced.



(a) ILD - Sphere Model

(b) ILD - Kemar Model

Figure 5.8: Calculated interaural level differences for a sphere model (left) and a Kemar model (right), solved at a sample rate of 65kHz.

The above results imply that, at least at a sample rate of 65kHz, there is not much advantage to attempting to model a head with a high geometrical accuracy, as most of the fine features of the head are vanished in the discretisation process. There is, however, one principal difference between the numerical models of the sphere and Kemar, in terms of their inherent interaural cues. For the sphere model, both ITDs and ILDs are symmetrical around the two lateral angles, which is to be expected due to the symmetrical shape of the sphere. For the Kemar model, even at such a low discretisation resolution, there is noticable difference between interaural cues of the front and back hemispheres, which is evident in assymetries around 90° and 270°. This exemplifies the uniqueness of interaural cues obtained with a more accurate representation of a head, which in context of this thesis, play an important role in resolving front to back confusions in sound localisation, as will be discussed in Chapter 6.

In terms of more general applications of the FDTD method, if one requires only interaural cues, then implementing a *snowman model* (Algazi et al., 2002) could provide a good alternative that does not require to go through the process of laser-scanning and voxellising a complex geometrical object. In Wave-Cloud, the functions provided by the preprocessing API automatically voxellise the laser-scanned head according to the defined sample rate. Thus, in terms of future-proofing the software package, it is sensible to use the laser-scanning approach, even though due to technological constraints, at the moment a full representation of the head is not possible in most practical cases. Nevertheless, the fact that directly embedding a listener geometry in the grid is, in principle, numerically inconsistent, emphasises the need to develop improved methods of obtaining BRIRs in FDTD simulation.

5.5 Modelling Frequency-Dependent Elements

Most realistic cases of room acoustics involve elements which need to be modelled as frequency dependent quantities. These are typically radiation patterns of sound sources, boundary impedances or receiver directivity patterns. As described in Chapter 3, an effective way to tackle frequency dependent boundaries is to employ digital impedance filters which can be designed according to analytical principles or empirical data. In (Savioja, 2010) this approach was taken, yet it was suggested that when modelling rooms of high geometrical complexity on a GPGPU, it may be simpler, and possibly more efficient, to run multiple frequency independent simulations with different boundary impedance values. In this chapter, this approach is generalised to modelling frequency dependent sources as well as boundaries, and its merits and drawbacks are discussed in context of an integrated FDTD modelling environment, such as WaveCloud.

A diagram of the underlying structure of this method, which will herein be referred to as the *multiband method*, is shown in Figure 5.9.

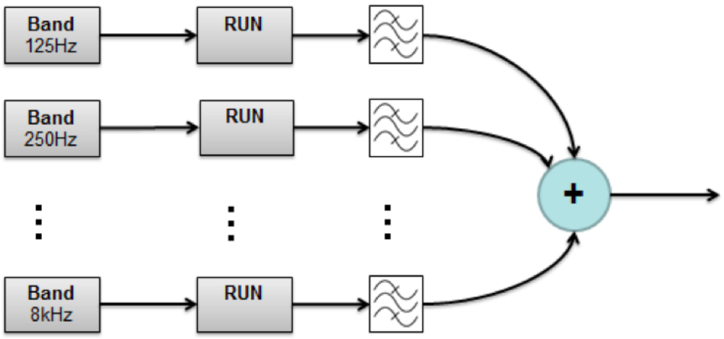


Figure 5.9: General structure of the multiband method.

The user executes a number of separate simulations, each corresponding to boundary impedance and source directivity data in a different frequency band. In room acoustics, empirical data is typically available between 63Hz and 8kHz in octave or third octave bands, thus in order to cover the entire bandwidth one would need to run $7 - 22$ simulations. Once all data is calculated, the resulting response of each simulation is passed through a full or third octave bandpass filter, and all filtered responses are summed into a single frequency dependent response. Concerns about the efficiency of such an approach are further discussed in Section 5.5.3, however it should be pointed out that the motivation here is to generalise a method for solving frequency dependency, rather than to find the most computationally-efficient way to accomplish this task.

5.5.1 Boundary Model

In contrast to Section 5.4 where the boundary model was predicated on impedance discontinuity, in this case it is assumed that the boundaries represent the edges of the computational domain, and as such require that the grid is spatially terminated. This is accomplished by considering a special case of the frequency dependent LRS boundary model presented in Section 3.2.4, in which a zero-order filter is used to represent a frequency independent quantity. Accordingly, Equation 3.22 reduces to

$$p\Big|_{\mathbf{i}}^{n+1} = D_0 \left[d_1 \sum p_a + d_2 \sum p_{sd} + d_3 \sum p_d + d_4 p\Big|_{\mathbf{i}}^{n} + \left(\frac{\lambda}{\xi_w} - 1 \right) p\Big|_{\mathbf{i}}^{n-1} \right] \quad (5.4)$$

where ξ_w is the normalised impedance of the boundary node, $D_0 = \left(1 + \frac{\lambda}{\xi_w}\right)^{-1}$, and p_a, p_{sd} and p_d are given in Equation 3.18. To asses the accuracy of the multiband method with the above boundary model, the experimental procedure suggested in (Kowalczyk and van Walstijn, 2008) is followed. A domain of 373^3 nodes corresponding to a physical volume of 4096m^3 is modelled on a GPGPU using the IWB scheme. A wall is placed in the centre of the domain, effectively splitting it into half, as shown in Figure 5.10. In the domain, a source and a receiver are placed at ±45° and a radial distance of 94 nodes from the wall, which is sufficiently far to assume plane wave propagation for most frequencies of interest. All simulations are executed for a duration long enough for the reflection from the wall to arrive at the receiving position, but short enough such that no reflections from the outer boundaries are recorded. This can be seen as an emulation of a surface measurement in a virtual anechoic chamber, as shown in Figure 5.10.

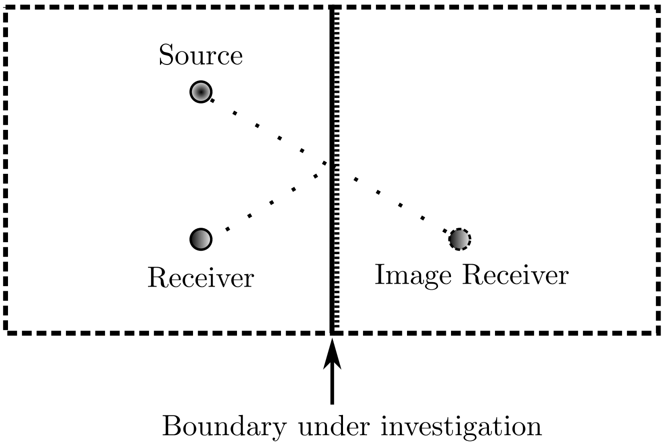


Figure 5.10: Experimental setup for the multiband boundary model investigations.

The measurement procedure is completed in two steps. First, simulation is executed with the wall under investigation in place, resulting in the recorded signal $x_{(tot)}\big|^n$ whose soundfield contains a direct component and a single reflection from the wall. Next, the wall is lifted and two signals are recorded. At the receiving position, the recorded signal $x_d\big|^n$ contains only the direct component of $x_{(tot)}\big|^n$, and at the mirror receiving position the signal $x_m\big|^n$ emulates only the reflected component of $x_{(tot)}\big|^n$, but without any level or phase change due to the wall. Accordingly, a signal carrying only the numerical reflection can be obtained by

$$x_{(rn)}\big|^n = x_{(tot)}\big|^n - x_d\big|^n \tag{5.5}$$

Assuming a real reflection coefficient $R(f)$ representing the boundary impedance at the simulation of the f^{th} frequency band, the expected theoretical reflection is given by

$$x_{(rt)}\big|^n = R(f) \cdot x_m\big|^n \tag{5.6}$$

The procedure is then repeated for all frequency bands, each having their own $R(f)$, as shown in Table 5.2. In this experiment, the sample rate was set to 8kHz due to computational constraints arising from the large physical volume required to emulate the anechoic measurement space. Thus, to comply with the 2% dispersion limit occurring at $0.186 f_s$ for the IWB scheme (Kowalczyk and van

Table 5.2: Normalised impedance values $\xi_w(f)$, and corresponding reflection factors $R(f)$, used in the multiband boundary experiment.

Band, f	$\xi_w(f)$	$R(f)$
125Hz	19	0.8614
250Hz	7.87	0.6954
500Hz	1.92	0.1529
1kHz	1.57	0.0541
2kHz	2.61	0.2985

Walstijn, 2011), the highest considered frequency band is 1kHz whose upper band limit corresponds to 1420Hz (for visual purposes, the investigation is carried up to the 2kHz frequency band). To generate the frequency dependent impulse response, the individual responses were filtered using an 8^{th} order butterworth filter whose bandwidth corresponds to one octave. Figure 5.11 depicts the theoretical reflection coefficients and the numerically calculated reflectance of the boundary.

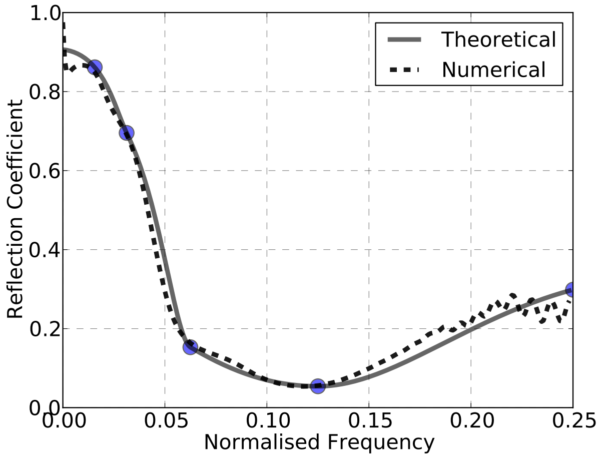


Figure 5.11: Theoretical reflection coefficients (filled circles) plotted against the calculated numerical reflectance (dashed line). The solid line depicts the theoretical reflectance which is calculated by performing cubic interpolation on the theoretical reflection coefficients.

It can be seen that the numerical reflectance curve passes through all theoretical reflection coefficients (filled circles), and intermediate values are interpolated by the corresponding octave band filters (dashed curve). This shows that the reflectance of the boundary between the specified octave bands, is controlled by the post-processing filters. In this example, the numerical curve closely resembles the theoretical curve (solid line) which is obtained by taking the cubic interpolation of the individual reflection coefficients. However, different shapes of intermediate interpolation could be accomplished by choosing different filters.

5.5.2 Source Directivity Model

Whilst the problem of frequency dependent boundaries has been thoroughly covered by many authors, modelling sound sources whose directivity is frequency dependent has only received little attention in the literature. In this section, it is shown how the multiband approach can be applied to such a problem. For simplicity, the most basic form of frequency independent differential sources (Southern and Murphy, 2009) are used in the individual frequency band simulations, however, the approach can be generalised to other directivity models as well.

The simulation setup comprised of a two-dimensional grid of 606x606 nodes. To reduce errors arising from dispersion anisotropy, the grid was solved using the interpolated isotropic scheme, with the simulation parameters set to $b = 1/6$ and $\lambda = \sqrt{1/2}$ (Kowalczyk and van Walstijn, 2010). This scheme features an isotropic behaviour up to $0.2f_s$, meaning that with the sample rate set at 32kHz, reliable results can be obtained upto 6.4kHz, which covers the entire 4kHz octave band. The choice of using a 2D grid was made because of computational constraints, however, generalisation to 3D is straightforward.

The source was situated at the centre of the domain, and 72 receivers were at a radial distance of 178 nodes around the source, corresponding to a full 360° circle sampled at 5° intervals. Six simulations were executed with different directivity functions corresponding to data in a different octave band, as shown in Figure 5.10.

For each simulation, the acoustic pressure was recorded at each of the 72 receivers. Once all data has been obtained, the 8^{th} order Butterworth filters were applied in one octave bandwidth, and results were summed to obtain 72 frequency dependent impulse responses. To investigate the way intermediate directivity functions are interpolated, the resulting set of impulse responses was further filtered with $1/3^{rd}$ octave filters. As seen in Figures 5.13 and 5.14, even though only 6 independent simulations were used, results show a gradual morphing of

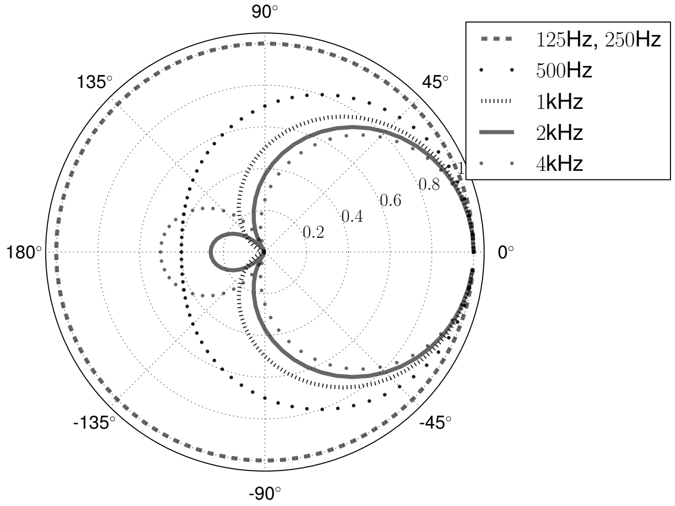


Figure 5.12: Source directivity function as used in each of the six simulations, corresponding to data in single octave bands.

the directivity functions across the different frequency bands.

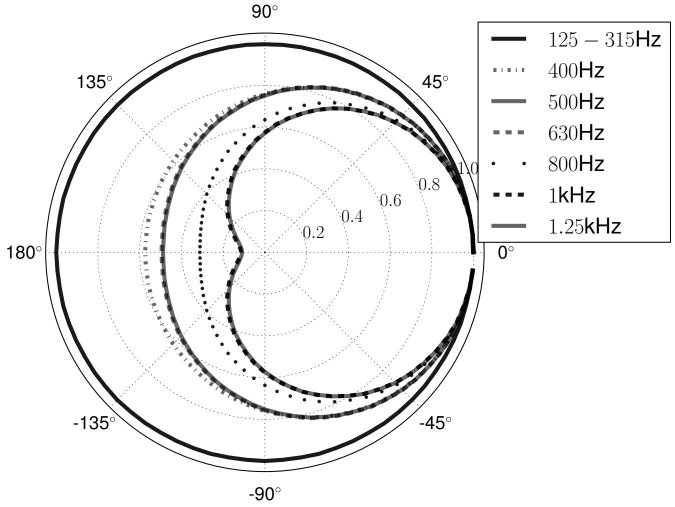


Figure 5.13: Resulting source directivity plotted at $1/3^{rd}$ octave bands, ranging from $125 - 1250$Hz.

5.5.3 Discussion

The results shown in this section indicate that the multiband approach can be applied to problems involving frequency dependent boundaries and sources, and in that sense, provide a general way of turning an FDTD simulation into a frequency dependent one. The simplicity and generality of the method comes at the cost of

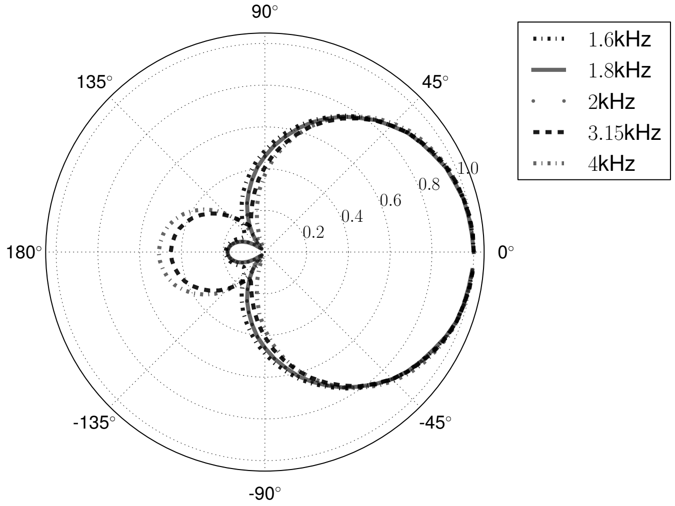


Figure 5.14: Resulting source directivity plotted at $1/3^{rd}$ octave bands, ranging from $1.6 - 4$kHz.

having to run and post-process a number of frequency independent simulations. This, of course, prevents the method from being used in real-time and affects its efficiency, which largely depends on the number of bands required to represent the frequency dependent source or boundaries. Even at a relatively low resolution of a single octave band this may seem as an inefficient method, however, one should bear in mind that when taking the single-pass frequency dependent approach (as discussed in Chapter 3), efficiency would also be dependent on the complexity of the digital impedance filters being used. Additionally, as the grid's sample rate is normally derived from the highest frequency required in the simulation, then one could, in theory, reduce the sample rate for simulations of the lower octave bands, which would result in significant computational savings, as was reviewed in (Sheaffer et al., 2011). However, due to spatial quantisation errors, the resulting impulse responses would have to be efficiently aligned in time or spatially interpolated prior to their summation, which presents an additional challenge.

Another important point concerns the ability to model boundaries of complex impedance. When taking a single-pass approach, then frequency dependency is represented by a digital impedance filter whose magnitude *and* phase response can be controlled by the user. In the case of the multiband approach, the boundary update equation is reduced to a zero-order filter which is in essence, a real valued scalar. Nevertheless, it should be borne in mind that most empirically available material and source directivity data is in the form of absorption coefficients in one or third octave bands. Accordingly, the multiband approach is an acceptable compromise, when trading off efficiency with simplicity and generality.

5.6 Output Visualisation

In context of numerical modelling, the importance of visualisation is twofold. First and foremost, it provides visual feedback allowing the modeller to efficiently inspect time-varying processes across the entire soundfield. Secondly, it is instrumental in tracking down modelling mistakes, such as boundary geometries and source and receiver placements. In WaveCloud, only limited visualisation can be achieved in real time, and more complex visualisation is normally performed offline using ParaView (Ahrens et al., 2005), which is an open-source application aimed at scientific visualisation. Figure 5.15 shows a complete volume rendering of a modelled soundfield captured 20ms after simulation onset, as generated by WaveCloud. As the 3D soundfield is visually complex, a more informative means of visualisation is to slice the volume along one of its principal axis, and accordingly, to plot the 2D pressure on that plane. Figures 5.16 and 5.17 depict the soundfield on the planes normal to the y and z directions, respecitvely.

At this time, communication between WaveCloud and ParaView is performed manually, meaning that the user needs to define the appropriate object pipeline inside ParaView. Since WaveCloud requires only specific visualisation features, a customised ParaView derivative is planned for future work. This would allow WaveCloud to directly communicate with its tailor-made visualisation engine, allowing for an easier visualisation process.

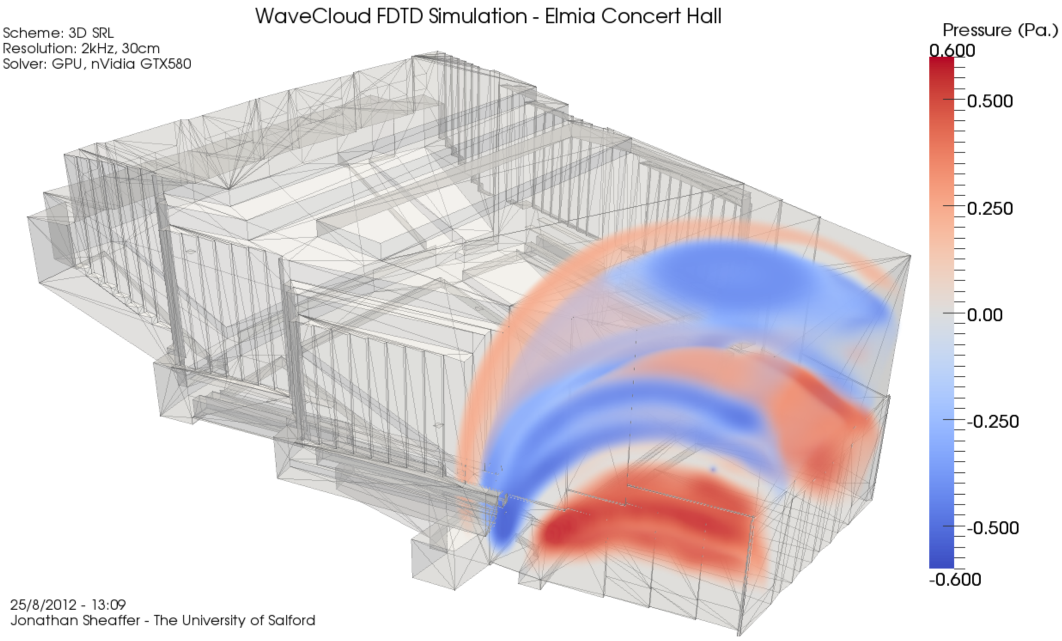


Figure 5.15: Volume rendering of the FDTD model of the Elmia concert hall, for at 20ms from simulation onset.

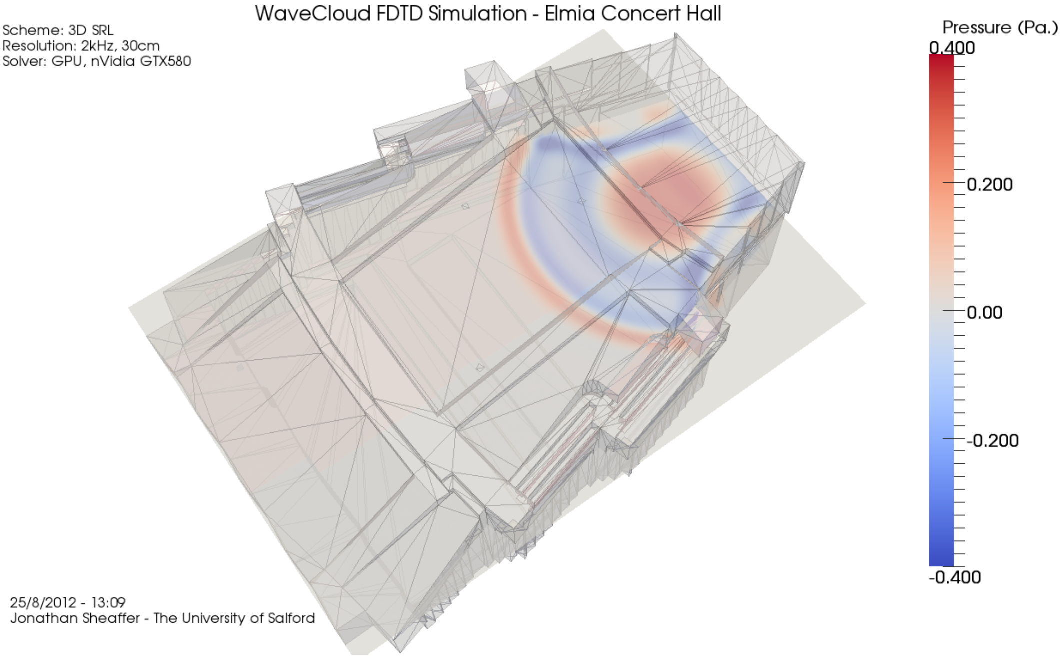


Figure 5.16: Top-slice (Z-normal) visualisation of the FDTD model of the Elmia concert hall, for at 20ms from simulation onset.

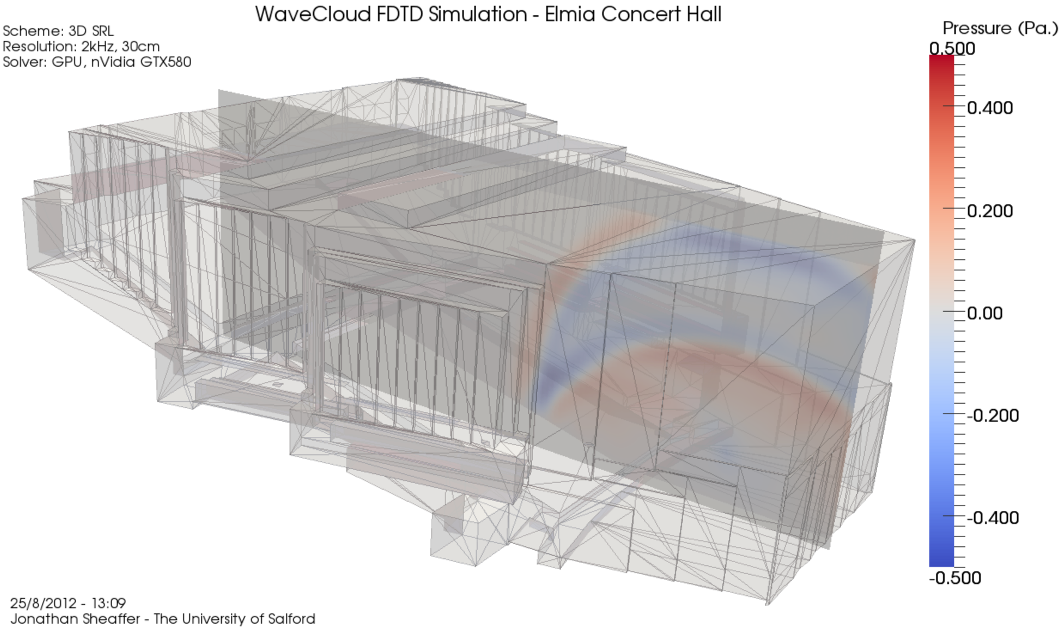


Figure 5.17: Side-slice (Y-normal) visualisation of the FDTD model of the Elmia concert hall, for at 20ms from simulation onset.

—Music is the pleasure the human mind experiences from counting without being aware that it is counting.

Gottfried Leibniz

6

An Integrated Model of Sound Localisation in Rooms Part I: Model Structure

In this chapter, a generalised framework for modelling the perception of sound localisation in rooms is suggested. This framework, which in short shall be called the *Perceptual Model*, is divided into three principal components, namely modelling of the auditory peripheral system, a binaural processor and a final decision stage. In context of this thesis, the room model provides binaural inputs to the perceptual model discussed here. Nevertheless, it is not explicitly required that the inputs are obtained by means of an FDTD simulation, and the model can be used on its own to evaluate measured responses as well. For technical reasons described in section 6.1.1, only binaural cues are generated by the room and listener model. As monaural cues are not calculated, the perceptual model is explicitly designed to localise sound on the horizontal plane and in an angular range of ±90°. The general structure of the perceptual model is shown in Figure 6.1.

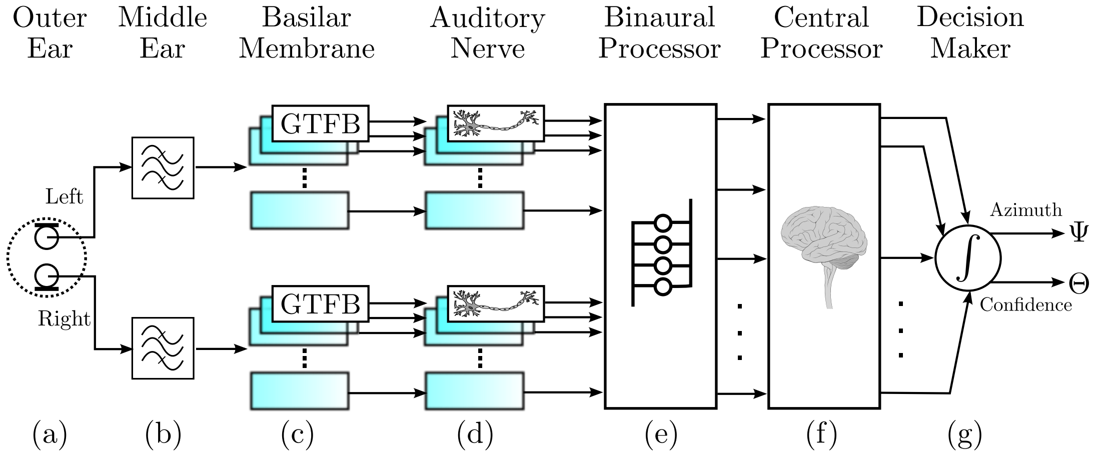


Figure 6.1: Structure of the perceptual model. (a) Output from the room model (or any other binaural signal). (b) Bandpass filter simulating the middle-ear. (c) Gammatone filter bank to model the frequency selectivity of the basilar membrane. (d) Model of the auditory nerve. (e) Binaural processor based on interaural coherence. (f) Central processor based on template matching. (g) Across-frequency integration and final localisation judgement.

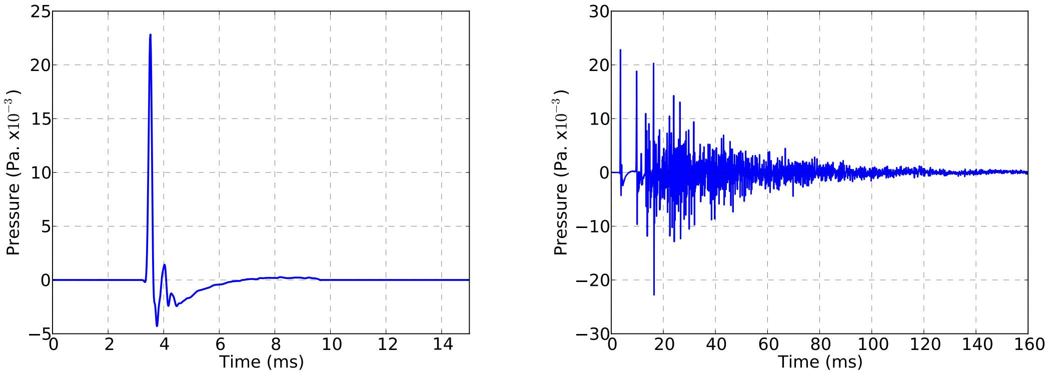
This chapter is structured as follows. Section 6.1 briefly explains the models which are here employed to simulate the function of the auditory periphery, namely the middle-ear (b), the inner-ear (c) and neural transduction (d). Section 6.2 portrays the binaural processor (e) originally introduced in (Faller and Merimaa, 2004b) which is extended here for processing of interaural cues in multiple bands. Section 6.3 complements the binaural processor with central processor (f) and decision-maker (g), both of which are novel contributions of the author. The structure and functionality of entire integrated model is then discussed in Section 6.3.

6.1 Auditory Periphery

6.1.1 Outer and Middle Ear

The signal at the outer ear, as shown in Figure 6.1(a), can be thought of as convolution of the stimulus at the absence of the head with the HRTF at the angle of arrival of the incoming wave. When the perceptual model is used standalone, that is, it is not used in conjunction with the room model, then such a signal can be obtained by convolving a stimulus with a measured HRIR or BRIR. When the entire integrated model is employed, then the output of the room model is equivalent to a measured BRIR. Of course, the degree to which these BRIRs are accurate are strongly dependent on the spatial resolution of the FDTD grid. As the perceptual model presented here only processes binaural cues, obtaining additional monaural cues is not necessary. Thus, the main function of this stage is to account for effects of the head itself. In all examples and experiments reported in this and the following chapters, a *Standard Rectilinear* (SRL) scheme is solved in a sample rate of $f_s = 65000$Hz ($X = 9.15$mm). In order to adhere to the 2% accuracy and dispersion criteria proposed in (Kowalczyk and van Walstijn, 2011), the grid is excited with a *physically constrained source* (PCS) used in conjunction with a maximally flat FIR whose high cutoff frequency is set to $f_c = 0.075 f_s$. The mechanical parameters of the PCS are chosen such that $f_0 = 100$Hz, $M = 25$g and $Q = 0.7$. This means that reliable results can be obtained up to a frequency of 4875Hz, which is slightly above the centre of the $\dot{k} = 28$ auditory filter. As such, all analyses reported in this thesis are performed up to that frequency, unless otherwise stated. Prior to convolving the modelled BRIRs with the desired stimuli, the BRIRs are down-sampled to the standard audio rate of $f_s = 44100$Hz.

Figure 6.2 depicts modelled BRIRs for two extreme cases of listening situations: (a) a source placed at a radial distance of 1m and 30° to the right of a listener in anechoic setting, and (b) similar conditions but at the presence of some reflections (frequency averaged $T_{60} = 275$ms). These examples are further used throughout the chapter to demonstrate the different model stages. Following that, the frequency weighting of the middle ear shown in Figure 6.1(a) is modelled as a bandpass filter with a roll-off of 6dB/oct below 1kHz and above 4kHz, as suggested in (Breebaart et al., 2001).



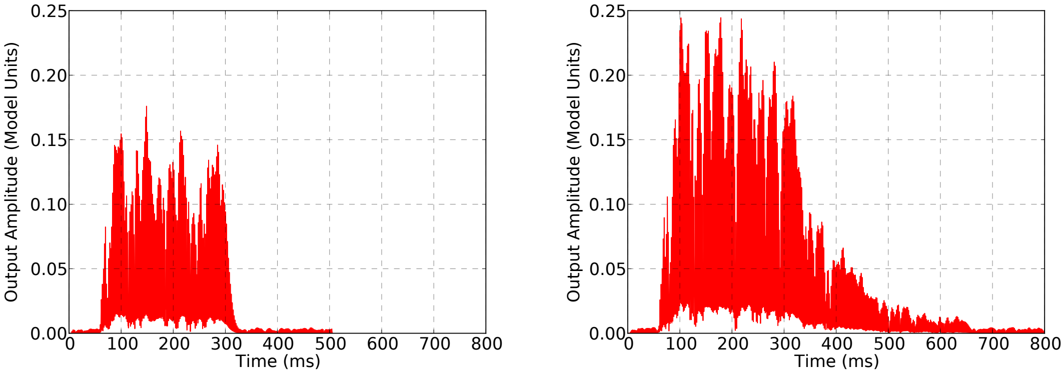
(a) Free Field (b) Enclosed Space

Figure 6.2: Sound pressure at the left ear for a source at a radial distance of 1m and 30° to the right of a listener in (a) anechoic setting and (b) enclosed space .

6.1.2 Inner Ear

The effects of the basilar membrane (Figure 6.1c) are simulated by passing the outputs of the middle ear model through a gammatone filter bank (GTFB), as suggested in (Patterson et al., 1995). The GTFB splits the signal into bands whose centre frequencies are spaced by a single equivalent rectangular bandwidth (ERB) per modelled auditory filter (Moore and Glasberg, 1986). To account for the absolute threshold of the auditory system, statistically-independent Gaussian noise is also passed through the GTFB and then added to the filtered signals. The noise has frequency dependent sound pressure levels which are scaled according to the distributions suggested in (Faller and Merimaa, 2004b).

Each resulting signal is then passed through a model of neural transduction (Bernstein et al., 1999) as suggested in (Faller and Merimaa, 2004b). In this stage, shown in Figure 6.1(d), the signals' envelopes are raised to the power 0.23, half-wave rectified, squared and passed through a fourth order low-pass filter with a cutoff frequency of 425Hz. Below this frequency, the filter has a negligible effect on the output, meaning that losses only occur in the waveform's negative phase (due to the rectification process). Thus, there is no loss of timing information in the remaining positive phase and the fine structure of the waveform is preserved. At higher frequencies residing within the filter's rolloff, there is a gradual loss of phase information, simulating the frequency-dependent decrease of phase locking as known to take place in the auditory nerve (Bernstein and Trahiotis, 1996). This process outputs sets of signals denoting nerve firing densities at different critical bands, one set per ear. An example calculated for the critical band centred at 538Hz ($\dot{k} = 11$) is shown in Figure 6.3. These model stages are implemented using a freely available Matlab toolbox (Faller and Merimaa, 2004a).



(a) Free Field

(b) Enclosed Space

Figure 6.3: Output of the left-ear peripheral processor for a source at a radial distance of 1m and 30° to the right of a listener in (a) anechoic setting and (b) enclosed space. Results are shown for the 538Hz frequency channel.

6.2 Binaural Processor

In the human auditory system, encoding and some processing of spatial cues take place in the superior olivary complex and the inferior colliculus. Rather than attempting to model processes occurring at these areas in the brainstem and midbrain, Faller and Merimaa (2004b) have suggested an analytically motivated signal processing scheme to encode these data. Although it is not likely that such processes are explicitly employed by the auditory system, the model yields psychoacoustically plausible results. In context of the perceptual model presented in this chapter, the Faller-Merimaa approach is used to extract relevant binaural cues from signals arriving from the peripheral processor. Here, the process is performed in all critical bands, as shown in Figure 6.1e. In this section, the general structure of the Faller-Merimaa model is explained and the implication of using their model to evaluate information spectrally is discussed.

6.2.1 Computation of Binaural Cues

Let us consider the signals $x^l_{\dot{k}}[n]$ and $x^r_{\dot{k}}[n]$ to be outputs of the left and right peripheral processors at the auditory band $\dot{k}$ and discrete time n. The normalised cross-correlation function $\gamma_{\dot{k}}[n]$ for each frequency channel $\dot{k}$, is given by

$$\gamma_{\dot{k}}[n,\tau] = \frac{\sum_{m=0}^{M-1} x^l_{\dot{k}}[n-m] x^r_{\dot{k}}[n-m-\tau]}{\sqrt{\sum_{m=0}^{M-1}\left(x^l_{\dot{k}}[n-m]\right)^2}\sqrt{\sum_{m=0}^{M-1}\left(x^r_{\dot{k}}[n-m-\tau]\right)^2}} \tag{6.1}$$

where M is the size of the time window and τ is the time lag, both denoted in samples. Unless otherwise stated, τ is evaluated over the discrete range equivalent to ±2ms, and M corresponds to a single-sided time window with a psychoacoustically plausible time constant of 10ms, as suggested in (Faller and Merimaa, 2004b). The instantaneous, auditory-band dependent ITD $\hat{\tau}_{\dot{k}}[n]$, and interaural coherence $\hat{c}_{\dot{k}}[n]$ are obtained from the cross-correlation function. These are given by

$$\hat{\tau}_{\dot{k}}[n] = \arg\max_{\tau} \gamma_{\dot{k}}[n,\tau] \tag{6.2}$$

$$\hat{c}_{\dot{k}}[n] = \max_{\tau} \gamma_{\dot{k}}[n,\tau] \tag{6.3}$$

The instantaneous, critical-band dependent ILD, is computed as

$$\hat{\alpha}_{\dot{k}}[n] = 10\log_{10}\left(\frac{\sum_{m=0}^{M-1}\left(x^{l}_{\dot{k}}[n-m]\right)^2}{\sum_{m=0}^{M-1}\left(x^{r}_{\dot{k}}[n-m]\right)^2}\right) \tag{6.4}$$

The binaural processor obtains a set of instantaneous cues, $\{\hat{\tau}_{\dot{k}}[n], \hat{\alpha}_{\dot{k}}[n], \hat{c}_{\dot{k}}[n]\}$, in each critical band $\dot{k}$.

To explain the rationale of the Faller-Merimaa approach, let us consider the case of a single stationary source in free field. For simplicity we focus on the analysis at a single critical band, as across-frequency integration only occurs at higher model stages. In this case, the resulting ITD and ILD cues are explicitly encoded in the location of the source. More importantly, from a temporal point of view, these instantaneous cues remain constant over time. On the contrary, when multiple sources or reflections are present, ITD and ILD values fluctuate over time, as these cues are essentially the result of superimposing waves arriving from different directions and therefore at different times and levels. However, reflections only arrive after the direct component of the sound field, thus there is a point in time before which cues generated in a reverberant setting are identical to those of a free-field setting. The underlying assumption of the Faller-Merimaa approach is that in order for the auditory system to resolve the direction of a sound source, it must obtain spatial cues similar to those obtained in a free-field listening situation.

From a signal statistics perspective, it can be said that at times in which presented cues resemble free field cues, the interaural coherence (IC) is high, i.e. $\hat{c}_{\dot{k}}[n] \simeq 1$. At times in which cues fluctuate, the IC decreases, i.e. $\hat{c}_{\dot{k}}[n] < 1$. Thus, it is possible to use IC as an indicator for selecting cues that convey relevant information about the source direction. This can be simply accomplished by defining some threshold c_0, and selecting sets of cues only when the instantaneous IC is above that threshold,

$$\left\{\hat{\tau}_{\dot{k}}[n], \hat{\alpha}_{\dot{k}}[n] \middle| \hat{c}_{\dot{k}}[n] > c_0(\dot{k})\right\} \tag{6.5}$$

As will be further explained, it is useful to define the IC threshold as a frequency-dependent quantity, and therefore c_0 is given as function of the critical band $\dot{k}$. Since c_0 directly controls the proportion of selected cues, it can be deduced that choosing c_0 is a fine balance between the amount of cues being used to make the localisation judgement and their similarity to free-field cues. Thus, lowering c_0 is equivalent to increasing the sensitivity of the system, and raising c_0 is equivalent to increasing the selectivity of the system. For example, if c_0 is set too high, then the effects of incoherent behaviour (such as reflections) are essentially not accounted for. It is therefore desired to select c_0 as low as possible, whilst still maintaining adequate similarity to free-field conditions, as otherwise, the model is not able to accurately localise the source. Some considerations for selecting an appropriate c_0 are given in the following section.

6.2.2 Considerations for Choosing c_0

Since the model does not have a direct physiological correlate, deriving an appropriate c_0 is analytically difficult. Faller and Merimaa (2004b) suggested that the auditory system dynamically adapts c_0 when experiencing a new listening situation. In their study they empirically adapted c_0 in each of their experiments such that the model would yield psychoacoustically plausible results. Whilst this assumption is sensible, it is not exactly clear what changes in independent variables (such as room size, absorption or physical objects) would constitute a new listening situation. The independent variables which define a listening situation should be carefully identified, as one wishes to avoid a situation where the model overfits the data.

It is defined here that a new listening situation corresponds to an entirely new model setup, for example a different room layout. It is not likely, however, that the auditory system re-adapts c_0 whenever a small change in independent variables occur, for example if a window in a room is opened, which effectively means that boundary conditions are slightly altered. Thus, in principle, we try to adapt c_0 as little as possible, and only when it is clear that listening conditions have been significantly changed. As this must be considered on a case-by-case basis, changing c_0 within the same experiment should be supported by appropriate reasoning.

Another important aspect affecting the choice of c_0 relates to the way the auditory system processes information in different auditory bands. Recall from Section 2.3 that the neural transduction mechanism introduces compressive non-linearities which are affected by reduction of phase-locking as frequency rises. In other words, the dynamic range of the IC is much higher at the lower auditory-bands than at the higher ones. This means that at low auditory-bands, c_0 can be more freely chosen whilst still maintaining an effective cue selection procedure. Since it is always desired to choose c_0 as low as possible, the dependence on auditory-band can be seen as a physiological constraint effectively dictating that c_0 should increase with frequency. If, for example, one chooses a low c_0 which is held constant across frequency, then at high critical bands c_0 may fall below the dynamic range of the IC, meaning that selected cues will almost never resemble free field cues. Of course, this constraint is only relevant for non free-field situations, as in a free field all cues can always be accounted for and there is no particular need for cue selection in the first place.

Whilst it is understood that c_0 should generally rise with frequency, it is not exactly clear what shape should the function $c_0(\dot{k})$ take. Based on empirical trials it seems that the *Cumulative Distribution Function* yields good results, however other alternatives may well be possible. This function is given by

$$c_0(\dot{k}) = \left(1 - e^{-\mu \dot{k}}\right) H(\dot{k}) \tag{6.6}$$

where $H(\cdot)$ is the Heaviside step function and μ controls the slope of the function as shown in Figure 6.4

6.2.3 Probability Density Functions

Even after a cue selection procedure has been applied, the computed ITD and ILD functions are still time dependent. It is therefore useful to identify the cues which are most probable to occur within the time series, as these are the

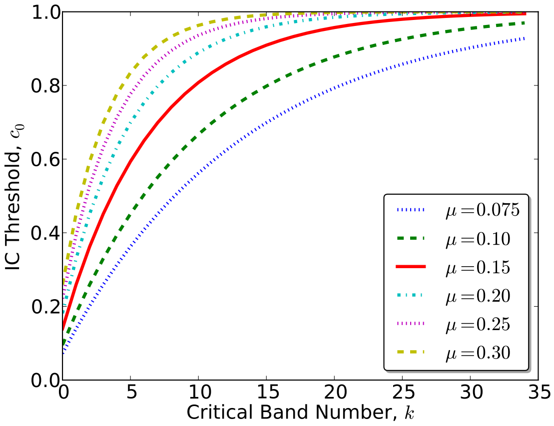
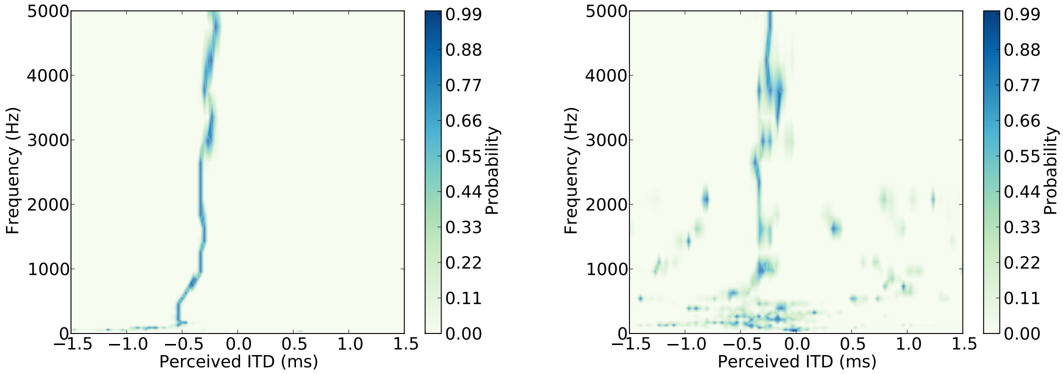
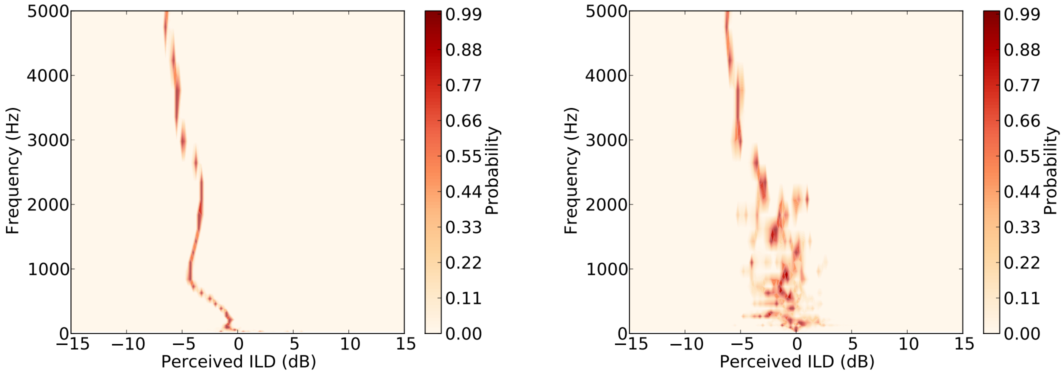


Figure 6.4: Interaural Coherence threshold as function of auditory-band number $c_0(\dot{k})$, shown for different values of μ

cues assumed to be used by the auditory system (Faller and Merimaa, 2004b). This can be accomplished by calculating *Probability Density Functions* (PDFs) for the selected ITD and ILD cues, $P_{\dot{k}}(\hat{\tau})$ and $P_{\dot{k}}(\hat{\alpha})$ respectively. Note that PDFs are normalised such that the most probable cue equals one, and that they are calculated separately for each critical band. It is then possible to plot two dimensional PDFs, which are essentially function of the independent variable as well as of frequency, as shown in Figure 6.5.



(a) Free Field $P_{\dot{k}}(\hat{\tau})$ (b) Enclosed Space $P_{\dot{k}}(\hat{\tau})$



(c) Free Field $P_{\dot{k}}(\hat{\alpha})$ (d) Enclosed Space $P_{\dot{k}}(\hat{\alpha})$

Figure 6.5: Probability density functions for ITD (upper panes) and ILD (lower panes), calculated for a source at a radial distance of 1m and 30° in an anechoic setting (left panes) and enclosed space (right panes). $c_0(\dot{k})$ is calculated with $\mu = 0.2$ in both cases.

6.3 Central Processor

The binaural stage generates a considerable amount of information that needs to be processed in order to arrive at a single localisation judgement. From a neurophysiological perspective, there is growing evidence that this process is accomplished by some means of pattern recognition. More specifically, it is known that the location-specific binaural stimuli generate a unique synchrony pattern activating a specific neural assembly which is characteristic for that location in space. This relationship is assumed to be learned by the brain as the auditory system develops, thus providing reference patterns for comparison, e.g. see (Goodman and Brette, 2010; Fischer, 2003; Arthur, 2002). With this in mind, we design a central processor whose function is to compare and match an unknown set of spectral PDF patterns, with those of previously learned location-specific sets. This is performed on a frequency-band basis, resulting in spectral data which is then weighted and integrated across-frequency to yield a single localisation judgement.

6.3.1 Cue Correlation Diagrams

The first step of the central processor is to resolve the perceived azimuth in each frequency channel. To accomplish this, the model must be able to relate ITD and ILD cues to their respective source location. In other words, the model needs to go through a supervised learning process in which each set of location-specific PDFs are classified and labelled respectively. Let us consider a source placed a known azimuth angle θ' about a listener in a free-field. The room model is executed and obtained signals are passed through the peripheral and binaural model stages resulting in two PDFs for that angle and for each frequency channel $\dot{k}$, namely $P_{\dot{k}}(\hat{\tau}, \theta')$ and $P_{\dot{k}}(\hat{\alpha}, \theta')$. This set of PDFs will be further referred to as a *Cue Probability Pattern* (CPP), and is compactly represented in matrix form as follows

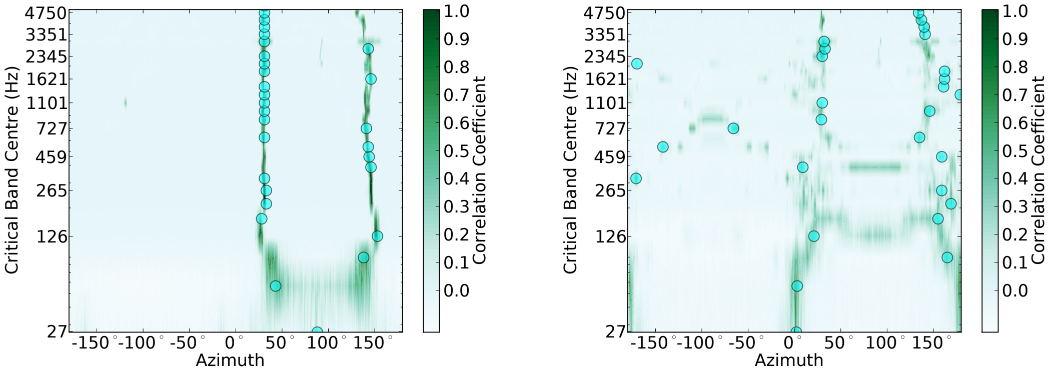
$$A(\dot{k}, \theta') = \begin{bmatrix} P_{\dot{k}}(\hat{\tau}, \theta') \\ P_{\dot{k}}(\hat{\alpha}, \theta') \end{bmatrix} \tag{6.7}$$

The underlying assumption is that these PDF sets are close to unique and thus each source location may be characterised by an individual CPP. The learning process involves generating a database of CPPs for each azimuth by using the room model to obtain HRIRs in a 1° resolution, 360° around the virtual listener. The source is placed at a distance of 1m from the listener, and all stimuli used in the training process are white noise bursts (200ms, 20ms $\cos^2$ ramps). It should be noted that this process is performed only once, and that the model does not attempt to correct or adapt the training-set over time as employed in more sophisticated algorithms simulating the plasticity of the brain.

Once all CPPs have been modelled and correctly labelled, the model is ready to attempt a localisation of unknown data. Given an arbitrary CPP for an unknown target angle θ denoted by $A(\dot{k}, \theta) = [P_{\dot{k}}(\hat{\tau}, \theta) \;\; P_{\dot{k}}(\hat{\alpha}, \theta)]^T$, the model attempts to find its best match in the existing training-set. This is accomplished by calculating the two-dimensional correlation coefficient, $\psi_{\dot{k}}$, between the unknown CPP and each of the previously learnt CPPs in the database. Observe the change of notation used to distinguish between the *target angle* θ which is the azimuth representing the unknown location of the source, and θ' which denotes the azimuth of a CPP within the training set. Similarly, the CPP for a target azimuth and a CPP template (i.e., for an azimuth stored in the model database) are denoted $A(\dot{k}, \theta)$ and $A(\dot{k}, \theta')$, respectively. The process can be expressed mathematically as

$$\psi_{\dot{k}}(\theta) = \frac{\sum\limits_{m}\sum\limits_{n}\left[A_{mn}(\dot{k}, \theta') - \overline{A(\dot{k}, \theta')}\right]\left[A_{mn}(\dot{k}, \theta) - \overline{A(\dot{k}, \theta)}\right]}{\sqrt{\left(\sum\limits_{m}\sum\limits_{n}\left[A_{mn}(\dot{k}, \theta') - \overline{A(\dot{k}, \theta')}\right]^2\right)\left(\sum\limits_{m}\sum\limits_{n}\left[A_{mn}(\dot{k}, \theta) - \overline{A(\dot{k}, \theta)}\right]^2\right)}} \tag{6.8}$$

where $\overline{A(\dot{k}, \theta)}$ and $\overline{A(\dot{k}, \theta')}$ denote the mean values of the target CPP and the CPP template, respectively. At low frequencies, it is assumed that the auditory system does not make use of ILD cues and therefore CPPs containing only ITD-based PDFs are considered. Otherwise, complete CPPs consisting of both ITD and ILD based PDFs are used. The transition band can be controlled by the user, and in this work is set at $\dot{k} = 16$ (approximately 1100Hz). Equation 6.8 is repeated for all azimuth angles in the training-set, which results in a matrix arrangement representing the correlation coefficients of the unknown CPP and in all template CPPs in the database. These data are visualised in a *Cue Correlation Diagram*, effectively showing the perceived location of the source in each critical band. Figure 6.6 depicts cue correlation diagrams for a sound source at 30° in free field and in an enclosed space.



(a) Free Field

(b) Enclosed Space

Figure 6.6: Cue correlation diagrams showing values of ψ_k for a source at a radial distance of 1m and 30° to the right of a listener in (a) anechoic setting and (b) enclosed space. The maximum value in each frequency channel is marked with a circle.

It can be clearly seen in Figure 6.6(a) that high values of ψ_k occur predominantly at response azimuths of 30° and 150°. The former corresponds to the true target azimuth of the source whereas the latter is due to front-back confusion. This is to be expected as the model does not account for monaural cues, and there is a strong resemblance between different CPPs generated for angles within the cone of confusion. For the case on an enclosed space, the front-back confusion is more pronounced and high correlation values are distributed over a wider range of response azimuths. Based on visual inspection it is evident that 30° is the most dominant response angle in the free field case, which is not as clear for the case of an enclosed space. Nevertheless, the model is still able to correctly localise the target azimuth in both cases. This is owing to the weighted integration process explained in the following section.

6.3.2 Across-Frequency Integration

Humans and animals usually refer to a specific point in space when localising sound. This, of course, shows that although cues are processed in different frequency-channels, all resulting information is integrated to yield a single localisation decision. The final stage of the model involves spectral integration of the data presented in the cue correlation diagram, which can be simply accomplished by summing the rows of the cue correlation matrix. However, it is known that the auditory system weights cues differently in different frequency channels (Macpherson and Middlebrooks, 2002), thus it is sensible to include a weighting function $w(f)$ accounting for such behaviour. The integration process yields a *Summary Correlation* function, $S(\theta)$, which can be expressed mathematically as

$$S(\theta) = \sum_{\dot{k}} \psi_{\dot{k}}(\theta) w(f) \tag{6.9}$$

where f is the centre frequency of the auditory band $\dot{k}$ in ERB scale (see Section 2.3). It is emphasised that regardless to the visual representation of cue correlation diagrams in which maximum values of ψ in each frequency channel are marked, in the integration process *all* values of $\psi_{\dot{k}}(\theta)$ are summed.

Following the numerical constraints reviewed in Section 6.1, the integration limits are set to ($3 \leq \dot{k} \leq 28$) for all subsequent experiments reported in this thesis. Deciding on the weighting function $w(f)$ is challenging. Although the issue of across-frequency integration has been investigated both physiologically and psychophysically, there are only a few studies explicitly quantifying the different weights given to integrated information (Raatgever, 1980; Stern et al., 1988; Akeroyd and Summerfield, 2000). These studies are concerned with Jeffress-type models which traditionally use only temporal cues to predict lateralisation. At the absence of ILD-related weighting functions, it is here assumed that high frequency data is uniformly weighted. The weighting function used here is given in equation 6.10 and is directly derived from (Stern et al., 1988).

$$w(f) = 10^{-(b_1 f + b_2 f^2 + b_3 f^3)/10} \tag{6.10}$$

Where f is the frequency in Hz and $b_1 = -9.383 \cdot 10^{-2}, b_2 = 1.125 \cdot 10^{-4}$ and $b_3 = -3.992 \cdot 10^{-8}$. Equation 6.10 is provided for $f \leq 1200$, above which the weighting function is assumed to remain constant, as shown in Figure 6.7.

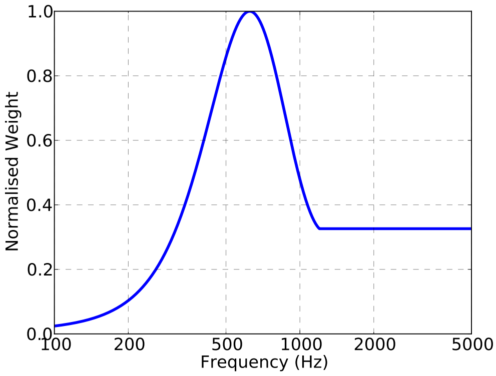
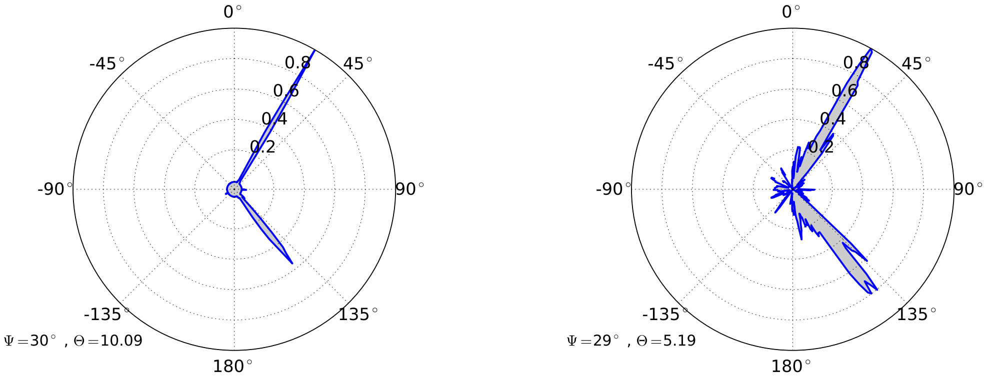


Figure 6.7: Weighting function, $w(f)$, after (Stern et al., 1988), used for across-frequency integration.

Figure 6.8 shows summary correlation diagrams for the two cases of a source in free field and in an enclosed space. Note that for visual clarity plots are shown in polar form and the summary correlation functions are normalised to unity.



(a) Free Field

(b) Enclosed Space

Figure 6.8: Summary correlation plots for a source at a radial distance of 1m and 30° to the right of a listener in (a) anechoic setting and (b) enclosed space. For visual clarity, plots are shown in polar form.

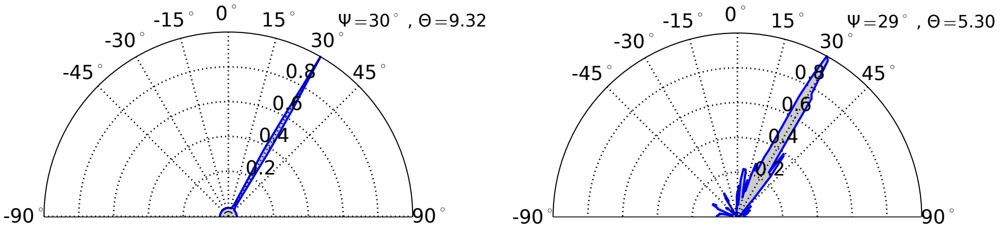
The final response azimuth, Ψ, is obtained from the summary correlation by

$$\Psi = \arg\max_{\theta} S(\theta) \tag{6.11}$$

Visual inspection of Figure 6.8 shows that whilst the model correctly localises the source in both cases, the summary curve for the case of the enclosed space is less defined. This is because its corresponding cue correlation diagram is more ambiguous, or in other words, in comparison to the free field case, there are many auditory bands that are not correctly aligned with the target azimuth. This ambiguity manifests itself in the form of noise and wider peaks in the summary correlation plot. It is therefore sensible to deduce that for the free field case, the model localises the source with greater *confidence*. This quantity, denoted Θ, is calculated as follows

$$\Theta = \frac{\max\ S(\theta)}{\sigma_S} \tag{6.12}$$

where σ_S is the standard deviation of $S(\theta)$. It is interesting that although monaural cues are not considered, the model is still able to distinguish front from back. This is owing to the asymmetrical shape of the human head used in the room simulation (see Section 5.4), and to the fact that the model operates in many frequency channels. However, as shown in Figure 6.8, this ability decreases in reverberant conditions. As such, summary correlation functions and their inherent quantities, Ψ and Θ, are only evaluated over the range $\theta = \pm 90°$, as shown in Figure 6.9. Accordingly, one should ensure that sound sources are presented over that range only, i.e. not only on the horizontal plane but also in front of the listener.



(a) Free Field

(b) Enclosed Space

Figure 6.9: Summary correlation plots, similar to Figure 6.8 only presented over the range $\theta = \pm 90°$.

6.4 Discussion

The perceptual model presented in this chapter can be divided into three principal stages, namely a peripheral processor, a binaural processor and a decision maker. The peripheral stage is predicated on algorithms drawn directly from physiological findings, which are commonly used in auditory modelling as 'off the shelf' solutions to simulate the process of transducing acoustic stimuli to neural responses. As this stage does not present any novelty to the model, it shall not be discussed here.

The core component of the binaural processor is based on a cue selection method introduced by Faller and Merimaa (2004b), which was not based on physiological findings. However, recent studies have suggested that there may be some neurophysiological processes supporting the Faller-Merimaa approach. Goodman and Brette (2010) studied the problem of sound localisation in complex listening situations using a biological spike-timing model simulating the behaviour of neural assemblies in the MSO (see Section 2.3 for details on the MSO). In their model, binaural neurons responded well only when receiving consistent signals, and ignored signals at times when they were noisy. This is equivalent to the underlying assumption of Faller and Merimaa in which binaural cues are selected only at times they resemble free-field cues, or in other words, when the interaural coherence is high. Devore and colleagues (2009) studied the directional sensitivity of inferior colliculus neurons in an anaesthetised cat. They showed that such neurons are more sensitive to the early part of the signal closer to the stimulus onset. These parts of the signal normally correspond to the direct component of a sound field, which is suggestive of a physiological filtering mechanism extracting reliable free-field cues from the complex stimuli, thus supporting the approach adopted here.

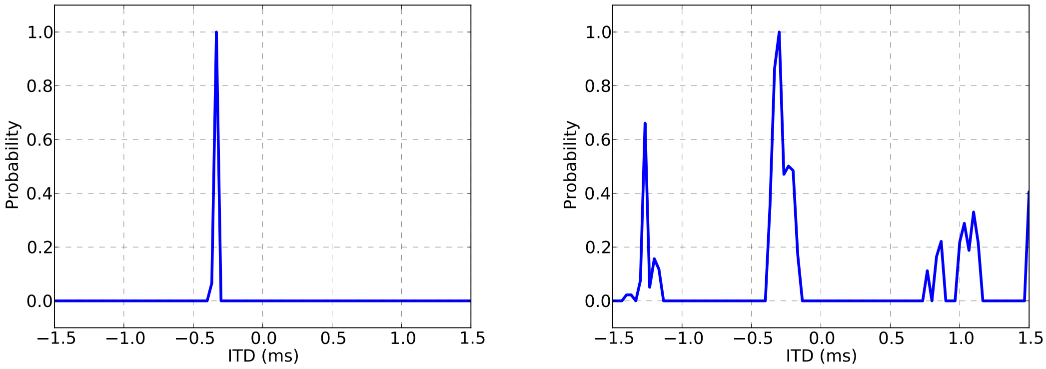
In their study, Faller and Merimaa (2004b) showed how their cue selection algorithm can explain localisation in complex listening situations, however their investigation was restricted to analysis in discrete auditory bands. Here, the novelty in the approach is to extend their model to evaluate binaural cues in multiple auditory bands. To accomplish this the concept of a frequency-dependent IC threshold (c_0) was introduced. Based on physiological processes occurring in the auditory periphery, it was argued that c_0 should rise with frequency. There are some psychophysical data which support employing such an approach in context of this model.

Ihlefeld and Shinn-Cunningham (2011) showed that in reverberant conditions, listeners give a heavier perceptual weight to low frequency cues, even though in such conditions high frequency stimuli are localised more accurately. Due to superposition in the acoustic domain, binaural cues are corrupted at the presence of reflections (Hartmann, 1983; Rakerd and Hartmann, 1985), an effect which intensifies as the direct-to-reverberant ratio decreases (Shinn-Cunningham et al., 2005). As low frequency ITDs are linked to phase information, these temporal fluctuations increase the ITD variability (Ihlefeld and Shinn-Cunningham, 2011), and decrease the consistence and therefore reliability of ITD cues (Shinn-Cunningham and Kawakyu, 2003; Rakerd and Hartmann, 2010). A similar trend is evident in the model presented here. Close observation of Figure 6.6(b) shows that in reverberant conditions, high values of $\psi_{\dot{k}}$ hardly coincide at the same response azimuth across frequency. This in turn means that although low frequencies are indeed given a heavier perceptual weight in the integration process, they hardly affect the position of the summary correlation peak as they are not aligned at the same azimuth (see Figure 6.8). Furthermore, this also justifies the choice of a low c_0 at low frequencies, as lowering c_0 effectively means that the model is less selective and more cues get included in the localisation judgement. As such, the frequency dependence of c_0 (see section 6.2.2) holds even in reverberant conditions. More importantly, if one were to choose a high c_0 at low frequencies, then in reverberant settings resulting cues would coincide better along a single azimuth. In other words, this means that ITD cues would be more perceptually reliable, which would be in contrast to psychophysical findings.

The binaural processor presented here is supplemented with a decision maker based on an efficient pattern recognition algorithm. A similar approach was taken in (Park et al., 2008), with the binaural stage based on a model suggested by Breebaart and colleagues (2001). There, however, the model was employed only in context of stereophonic reproduction in free field, and its performance in a reverberant environment was not investigated. As the Breebaart model produces time varying activity patterns, Park and colleagues relied on time averaged responses as input to their decision maker. Thus, it is not likely that lacking an efficient cue-selection procedure, their model would perform well in a reverberant environment. Nevertheless, the underlying principle of pattern matching is both physiologically reasonable (Stecker and Middlebrooks, 2003) and was shown to be a robust solution for azimuth mapping (Park et al., 2008). As such, it is used here in conjunction with the binaural processor which ensures that only selected cues are mapped.

Another method for cue-to-azimuth mapping employs scalar lookup-tables. In this approach, individual values for ITD and ILD cues in each auditory band are extracted from the stimuli and compared to those stored in a reference look-up table. Reference values are either obtained in closed-form using parametric models such as (Kuhn, 1977; Brown and Duda, 1998; Shinn-Cunningham et al., 2000; Raspaud et al., 2010), or estimated by a self-calibration process, e.g. in (Macpherson, 1991; Braasch, 2005b; Takanen, 2012). Whilst the scalar look-up table approach provides a simple means of mapping cues to their respective azimuth, it only reveals *whether* target and reference cues are similar, and not how similar they are. To exemplify this, consider the ITD-related PDF at the 962Hz frequency channel ($\dot{k} = 15$) for a source at 30° an enclosed space, as shown in Figure 6.10(b). The reference PDF from the corresponding training set is shown in Figure 6.10(a).

In both cases the peak probability occurs at similar ITD values, but clearly the shape of the PDF is significantly different. This is not surprising, as we expect a perceptual difference between the two acoustic conditions. Yet, the two patterns show that the characteristic ITD at 962Hz for a source at 30° is 0.34ms. Therefore, if one employs a scalar look-up table approach, then all the data characterising the *difference* between the two cases is lost. In the approach presented here, the correlation coefficient between the two patterns is measured, effectively showing to what degree the reference curve fits the target data. In this example the



(a) Reference

(b) Target

Figure 6.10: Probability density functions for a source at 30° shown for the 962Hz frequency channel. (a) reference PDF from training-set, (b) target PDF obtained in an enclosed space setting.

correlation coefficient is $\psi_{15}(30°) = 0.4505$ which indicates only a partial positive dependence. Although in this frequency channel this is the highest coefficient (see Figure 6.6b), all other values still participate in the integration process, which is what supplements noise to the summary correlation curve and contributes to the widening its peak (see Figures 6.8 and 6.9). This, in turn, allows us to distinguish between the two cases even when the final response azimuth, Ψ, is similar.

6.5 Conclusion

In this chapter a generalised model of sound localisation in rooms has been introduced. The binaural stage of the model is based on a cue selection method suggested by Faller and Merimaa (2004b), which is employed to analyse cues in the complete range of auditory frequency channels. The binaural processor is supplemented with a decision maker based on pattern matching and across-frequency integration, yielding a single localisation response and a confidence metric. It has been shown how the model can be used both quantitatively and qualitatively to analyse the process of sound localisation in enclosed spaces. In the following chapter, results are shown for typical listening tasks and the applications of the model to room acoustics and spatial audio are demonstrated.

—Each individual fact, taken by itself, can indeed arouse our curiosity or our astonishment, or be useful to us in its practical applications.

Hermann von Helmholtz

7

An Integrated Model of Sound Localisation in Rooms, Part II: Results and Applications

In Chapter 6 a perceptual model of sound localisation was described. When used in conjunction with the source, room and listener models described in Chapters 3 and 5, the integrated framework can be used to study various problems involving human sound localisation in realistic listening environments. The purpose of this chapter is twofold. First, the model is validated in the most simplified listening situation of a single source in free field (Section 7.1), and under the different phases of the precedence effect (Sections 7.2 and 7.3). Once a basic understanding of the model's performance is established, some of its potential applications to room acoustics are demonstrated, namely localisation at the presence of a single reflection (Section 7.4), localisation of occluded sources (Section 7.5) and at the presence of a reverberant field (Section 7.6). This is achieved by conducting a set of virtual listening tests, where in a similar manner to forced-choice subjective tests, the model is presented with a stimulus of unknown spatial position, and outputs a localisation judgement. Ideally, the results of matching real and virtual listening tests would have been compared, however, given the broad scope of the model, conducting listening tests with real human subjects was not feasible. Nevertheless, the fundamental psychoacoustics of sound localisation are in most cases well understood, and subjective data and corresponding statistical analyses obtained by other authors can be found in the literature. Accordingly, the strategy is to simulate, within the numerical constraints of the system, classical listening situations which have already been subjectively studied by other authors. When the experimental setup can be faithfully replicated in the numerical domain, then it is possible to directly compare modelled results to corresponding subjective data. In other cases, the model's output is analysed and compared to general trends observed in psychoacoustic tests.

7.1 A Single Source in Free Field

A single source in free field can be seen as the simplest localisation task, as no competing information from other sources or reflections corrupt the binaural cues available to the auditory system. Accordingly, the goal of this section is to evaluate how the model performs in such a free-field setting, before proceeding to more complex listening situations.

7.1.1 Methodology

Consider a sound source placed at a radial distance of 1m from a listener in free-field, as shown in Figure 7.1. To test the model's performance, 360 virtual listening tests were performed, each corresponding to a different presented azimuth. All simulation and subsequent auditory processing parameters are similar to those presented in Chapter 6. In particular, note that the listener geometry is based on laser-scans of a KEMAR mannikin for the presented target azimuths as well as in the corresponding training-set used by the model. As such, this experiment can be thought of as evaluating the model with individualised HRIRs. Similarly, the effects of using non-individualised HRIRs are studied using the same procedure, only with a laser-scan of an arbitrary individual[1], as shown in Figure 7.2. It should be noted that although the two geometries are significantly

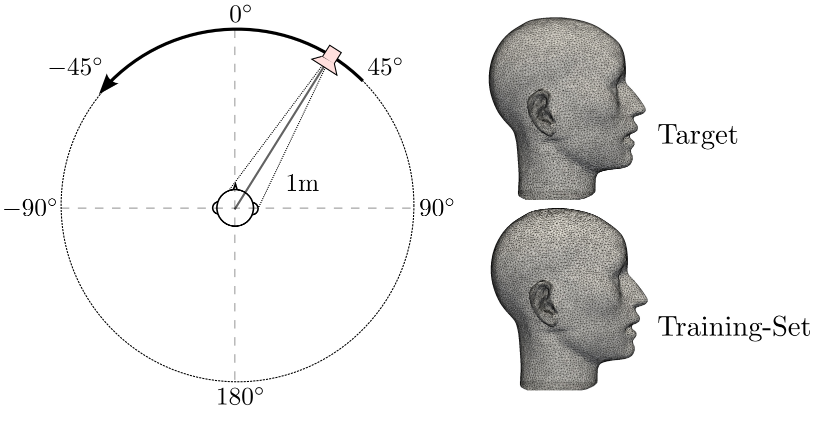


Figure 7.1: Experimental setup (*individualised HRIRs*) - a single source in free field. The listener geometry and all subsequent processing are identical to the target azimuths and the training-set.

---

[1] Laser scans obtained courtesy of James Busby, Ten24.

different, in the numerical domain many of the subtle differences are lost due to a relatively coarse discretisation (see Section 5.4).

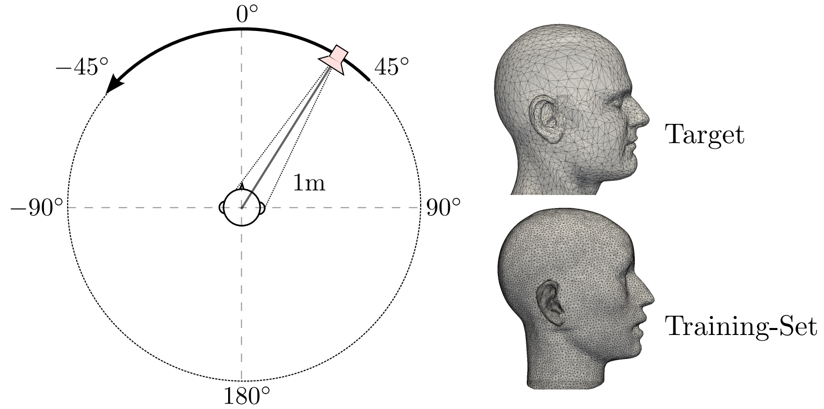


Figure 7.2: Experimental setup (*non-individualised HRIRs*) - a single source in free field. Observe that the listener geometry used for the target is *different* from the one used in the training-set.

To quantify the model's localisation errors, the *Mean Absolute Error* (MAE) between the target and response azimuth is calculated by

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|\theta_i - \Psi_i\right| \tag{7.1}$$

where θ_i and Ψ_i are the target and response azimuths of the i^{th} angle respectively, and N is the total number of angles in the data set.

7.1.2 Results

Figure 7.3 shows the model's response azimuth Ψ (solid line) and confidence metric Θ (dotted line) as function of target azimuth θ, for the case of an individualised HRIR set. In this case the model's response accurately matches the target angles, as evident in the straight line diagonally bisecting the figure. As the target and response azimuths are identical, the MAE is 0°, meaning that (unlike humans) the model performs this task flawlessly.

Results for the non-individualised HRIR case are shown in Figure 7.4 (solid and dotted lines). For reference, the results for the individualised case are also plotted (dashed line). Clearly, the non-individualised response sharply deviates

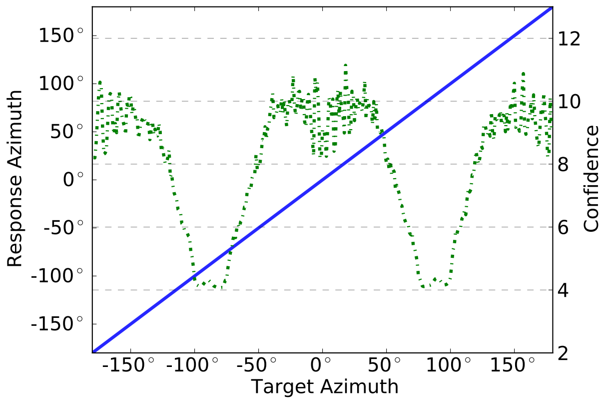


Figure 7.3: Model response as function of presented azimuth for an individualised HRIR set. $MAE = 0°$. Solid line: response azimuth Ψ, Dotted line: model confidence Θ.

from the individualised response, which is especially evident for angles above ±90°. Observe that these deviations are not random, and in fact, correspond to the angle opposite of Ψ, thus indicating front-back confusion. This in turn manifests itself in a significant increase in localisation error, and here $MAE = 45.3°$.

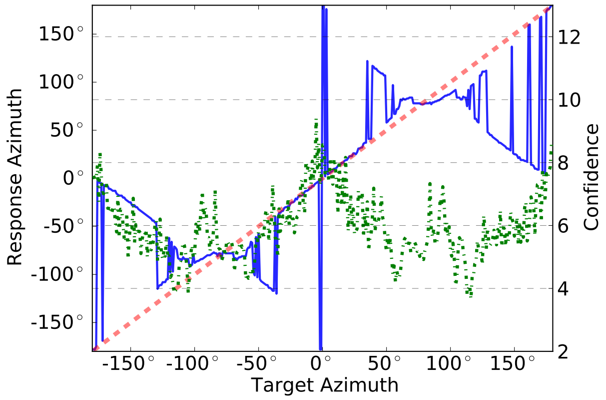


Figure 7.4: Model response as function of presented azimuth for a non-individualised HRIR set. $MAE = 45.3°$. Solid line: response azimuth Ψ, Dashed line: reference response for an individualised HRIR set, Dotted line: model confidence Θ.

In Section 6.3, it was suggested that confining the integration limits of $S(\theta)$ to ±90° could resolve such ambiguities. Figure 7.5 shows results for a similar non-individualised HRIR case, evaluated only over the $\theta = \pm 90°$ range. It can be seen that now the curve describing the non-individualised case almost exactly fits the curve for the individualised case. The localisation error is subsequently reduced to $MAE = 5.6°$.

Figure 7.5: Model response as function of presented azimuth for a non-individualised HRIR set, where the evaluation and corresponding integration limits are confined to ±90°. $MAE = 5.6°$. Solid line: response azimuth Ψ, Dashed line: reference response for an individualised HRIR set, Dotted line: model confidence Θ. Note the difference in the two y-axes.

7.1.3 Discussion

The perceptual model presented in this thesis is designed to imitate sound localisation by human subjects. When individualised HRIRs are used, the model excels at localisation in free field conditions. An MAE of 0° indicates that the model's response is error-free, which certainly exceeds the accuracy to which humans are able to localise sound in similar conditions (Makous and Middlebrooks, 1990). However, it is not uncommon for models to yield idealised results as some independent variables are not always accounted for. For example, it is likely that if a slightly different stimulus, source-listener distance or binaural parameters were used, then some error would have been introduced. Here, the exact same CPPs (see Section 6.3), were used for the targets and the corresponding training-set, thus the operation is similar to evaluating the 2D autocorrelation function of the model. Yet, the fact that no localisation errors are made, verifies the uniqueness of the interaural cues used in the decision making stage of the model, as assumed to be the case in Section 6.3.

The confidence metric Θ can provide some further insight into the model's performance. It can be seen from Figure 7.3 that Θ is high when the target is on the median plane, and gradually decreases as the target azimuth approaches ±90°. This is because the uniqueness of interaural cues, and therefore their respective CPPs, are more compromised around the cone of confusion. The more similar CPPs are, the noisier the summary correlation is, which manifests itself in a lower Θ. Subjective data indicate that the *Minimum Audible Angle* (MAA) is increased as the sound source is moved laterally (Perrott et al., 1990). Considering that MAA is a measure of spatial JND, it is not unreasonable to assume that the higher the JND is, the less confident subjects are in making a final localisation judgement, as such subjective tests are of a forced-choice nature. Thus, since the model's confidence Θ follows the same general trend, it can be loosely seen as a numerical descriptor of human MAA.

One drawback of the model is that it has to go through a supervised-learning process before it can be used to localise unknown data. From a physiological point of view this is a reasonable developmental process. Although humans are born with a crude ability to respond to lateral auditory stimuli (Muir and Field, 1979), the sound localisation system begins to develop only at a later stage (Field et al., 1980). As the model's learning process is time consuming, it is interesting to investigate whether the model can be generalised to many listeners whilst relying on a single pre-calculated training-set. In this case the HRIRs used in the training-set are different from those used for generating the target stimuli, which is equivalent to a localisation task with non-individualised HRIRs. When the model is tasked to localise sounds 360° around the listener, considerable front-back confusions occur, which is consistent with subjective data obtained using non-individualised HRTFs (Wenzel et al., 1993). However, these artefacts can be greatly reduced by confining the model's integration limits to ±90°, in which case localisation errors occur mostly around the cone of confusion.

Another important point is that in both cases the model's confidence is lower (and the MAE is higher) than for the case of an individualised training set.

These findings suggest that the model could be generalised without the need to explicitly perform per-subject training. Considering that the relatively low spatial resolution used to describe the listener's geometry does not allow for subtle details to be calculated, the difference between two listeners in the model is largely dominated by crude details such as the size of the head and the distance between the two ears. As such, it is expected that the model would perform well in localising sounds on the front-horizontal plane, even when using non-individualised training sets.

7.2 Two Coherent Sources - Level Differences

Two coherent sources with an *inter-channel level difference* (ICLD) is a commonly used example of summing localisation. When a listener is positioned equilaterally between the two sources and the inter-channel time difference is strictly 0ms, then interaural phase differences (at low frequencies) and interaural level differences (at high frequencies) are generated due to acoustic crosstalk (Snow, 1953). Accordingly, the locus of the auditory event (often referred to as a *phantom image* in the audio literature) can be controlled by manipulating the ICLD between the two sources. This special case of summing-localisation is the underlying principle of *Intensity-Stereophony*, which has been psychoacoustically studied since the early 1940s, e.g. in (de Boer, 1940; Wendt, 1963) extensively covered in (Blauert, 1997). In this section, a virtual listening test is conducted in order to evaluate the model's performance in the specific case of two coherent sources separated by level differences.

7.2.1 Methodology

Consider a listener positioned at the vertex of an equilateral triangle whose sides are 2m long, as shown in Figure 7.6. The two monopole sources $S1$ and $S2$ radiate coherent signals where their relative amplitudes are controlled parametrically within a ±30dB range, in a similar manner to the subjective tests reported in (Blauert, 1997). A broadband noise burst (200ms, 20ms $\cos^2$ ramps) is used as stimulus, and the interaural coherence threshold is set to $c_0 = 0.99$ which is held constant across the entire frequency range. Otherwise, all simulation parameters are identical to those reported in Chapter 6.

A total of 13 simulations were executed to cover the entire ICLD range of ±30dB and results are compared to the subjective data obtained by Wendt (1963), which is based on listening tests of 10 subjects, as reported in (Blauert, 1997). The modelled listening conditions were made similar to those used by Wendt which used non-directive sources in an anechoic environment. Wendt also immobilised the heads of his subjects to rule out any contribution of head movements to the localisation process. This is of particular importance here, as the localisation model is not designed to account for such cues.

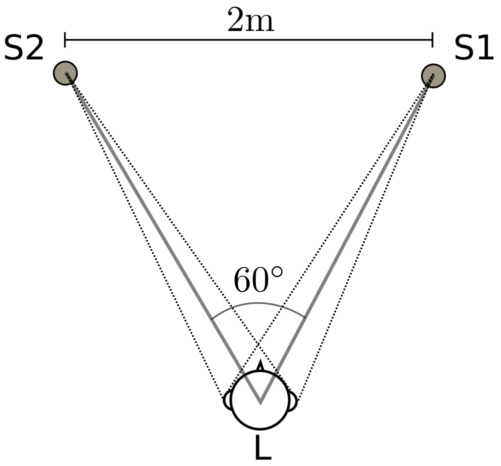


Figure 7.6: Experiment setup for two coherent sources. Listener at point L, two sources at points $S1$ and $S2$ at angles of $\theta_S = \pm 30°$ and radial distance of 2m.

7.2.2 Results and Discussion

The model's response azimuth (solid line) as function of ICLD and corresponding subjective data (dashed line) are shown in Figure 7.7. It can be seen that modelled results are in good agreement with subjective data, with a slight deviation as θ approaches ±30°. However, this deviation is in the order of only a few degrees, which lies within the localisation blur limits reported by Wendt.

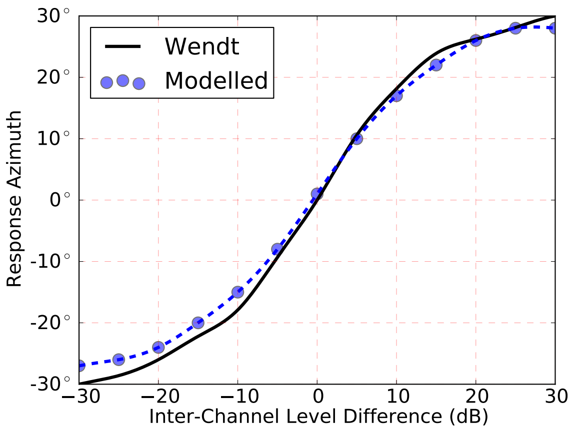
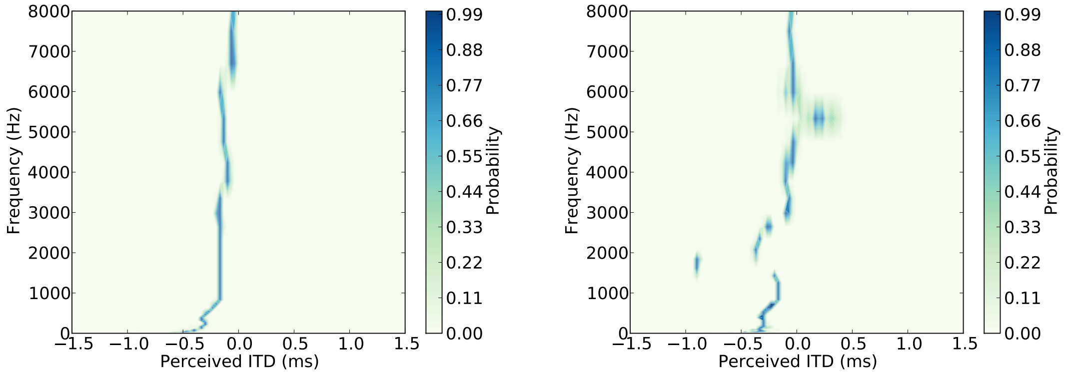


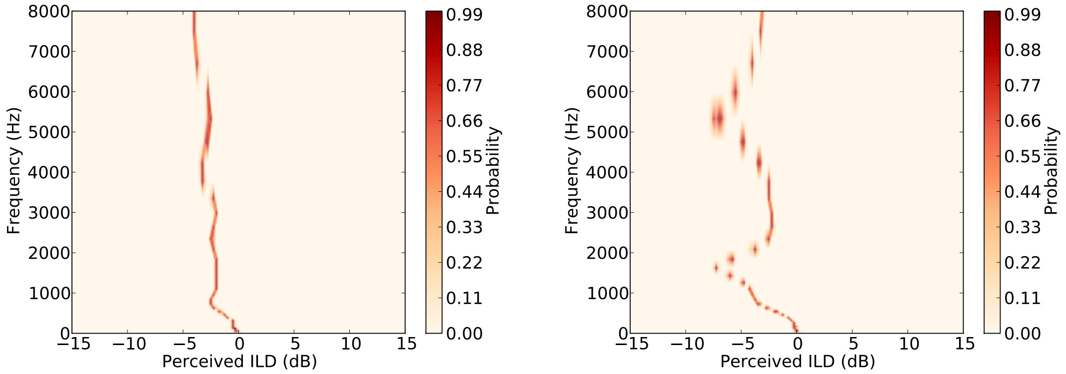
Figure 7.7: Results of modelling summing-localisation using inter-channel level differences. Solid line - fitted data from (Blauert, 1997) as obtained by (Wendt, 1963). Dashed line - modelled results.

It is interesting to compare the interaural cues generated from a real source against those imitated by two coherent sources with an ICLD. For example, at a target ICLD of 10dB the perceived azimuth is 17°, as shown in Figure 7.7. Figure 7.8 shows the output of the binaural processing stage of the model for these two cases. It can be seen that PDFs for the real source and the stereophonically reproduced one are in good agreement only for ITDs and at low frequencies. Since in both cases identical c_0 values were used in the binaural analysis, then this implies that wave superposition in the acoustic domain is largely responsible for the relative accuracy of stereophonic reproduction. In other words, changes in ICLDs cause a phase shift at the ears, and since the auditory system (and accordingly the model) gives a stronger perceptual weight to low frequencies which are dominated by ITDs, the final localisation judgement resembles that of a free-field situation. However, the observed discrepancies at mid and high frequencies imply that the two cases would still be perceived as different (even though the forced choice localisation task yields a correct answer).



(a) Single Source at 17°

(b) ICLD of 10dB



(c) Single Source at 17°

(d) ICLD of 10dB

Figure 7.8: Perceived interaural cues for two sources with ICLD=10dB (right panes) and a corresponding single source at 17° (left panes). Upper panes - ITD, Lower panes - ILD.

7.3 Two Coherent Sources - Time Differences

Two coherent sources with an *inter-channel time difference* (ICTD) is commonly used in the literature to emulate a source and an acoustic reflection in a controlled environment (Litovsky et al., 1999). The listener is presented with two stimuli (normally clicks, noise bursts or speech) from opposing directions separated by some time delay, called the *Inter-Stimulus Interval* (ISI). The source whose wavefront arrives first is termed the *lead*, whereas the other source is called the *lag*. As seen in Section 7.2, when the two sources radiate equal power and the ISI is zero, then the auditory event is perceived exactly in between the sources, which is attributed to *summing localisation*. When the ISI is increased towards 1ms, the auditory event is expected to shift towards the location of the lead (i.e. the 'earlier' source). Above 1ms, the lead is expected to completely dominate the perceived location of the auditory event, which is referred to as *Localisation Dominance*. At a higher ISI, normally ranging between $5-50$ms (depending on the acoustic setting and the temporal and spectral properties of the stimulus), the fusion of the sources break and two separate auditory events are perceived, which is referred to as *acoustic echo* (Litovsky et al., 1999). The purpose of this chapter is to evaluate the model's performance in such ICTD settings, which are referred to as the three phases of the precedence effect.

Methodology

Consider an experimental setup similar to the one detailed in Section 7.2, only with the two sources placed at $\pm 40°$ relative to the listener. In this experiment, the two sources are used to model click playback, by simulating a BRIR and convolving it with an anechoic recording of a click stimulus. Whilst the two sources radiate coherent signals, they are temporally separated by an ICTD, which for the click stimulus is simply regarded to as the *inter click interval* (ICI). It is important to distinguish between the ICTD and the ITD at the ears, as unlike lateralisation tests over headphones, here the ITD is altered due to acoustic crosstalk between the two sources. Thus, in this experiment the ICI corresponds to ICTD rather than ITD.

The experiment was repeated for ICIs ranging from 0 − 20ms. This range effectively covers all three phases of precedence related effects, namely summing localisation, localisation dominance and echo. For the fine range of summing-localisation (0 − 1ms) a sampling interval of 0.05ms was used, whereas above that increments of 0.5ms were used. The same frequency-dependent coherence threshold $c_0(\dot{k})$ was used in all experiments, with the slope set to $\mu = 0.2$.

Results

The model's response as function of ICI is shown in Figure 7.9. It can be clearly seen that the model is able to predict all three precedence related effects. At ICI=0ms the stimulus is localised exactly between the two sources, and as ICI is increased to 0.6ms, the model's response shifts toward the location of the leading source, which corresponds to summing localisation. Between 0.6 − 1.0ms, the location of the auditory event is perceived to exceed the actual azimuth of the leading source, and at about 1−3ms the localisation judgement fluctuates around the azimuth of the lead. Between 3 − 12.5ms the model localises the auditory event at the lead, which corresponds to localisation dominance. Above 12.5ms the model's response fluctuates between the lead and the lag, effectively indicating that two auditory events are perceived. This is because the model operates in a manner similar to a *forced-choice* listening experiment, thus at its final decision stage it must make a *single* localisation judgement regardless of the number of presented coherent sources (the decision is the maximum value of the summary-correlation plot).

To gain a better understanding of these results, it is useful to qualitatively analyse the outputs of the different intermediate components of the model. Figures 7.10 and 7.11 depict the output of the binaural processor and the summary-correlation diagram for ICIs within the localisation dominance and acoustic echo regions. At an ICI of 4.5ms, ITDs are inconsistent across frequency which is expected as the lagging wavefront corrupts the timing information carried by the lead. The ILDs, however, are split into two ranges. At frequencies up to about 2.5kHz they fluctuate slightly around 0dB, and above 3kHz they transition from

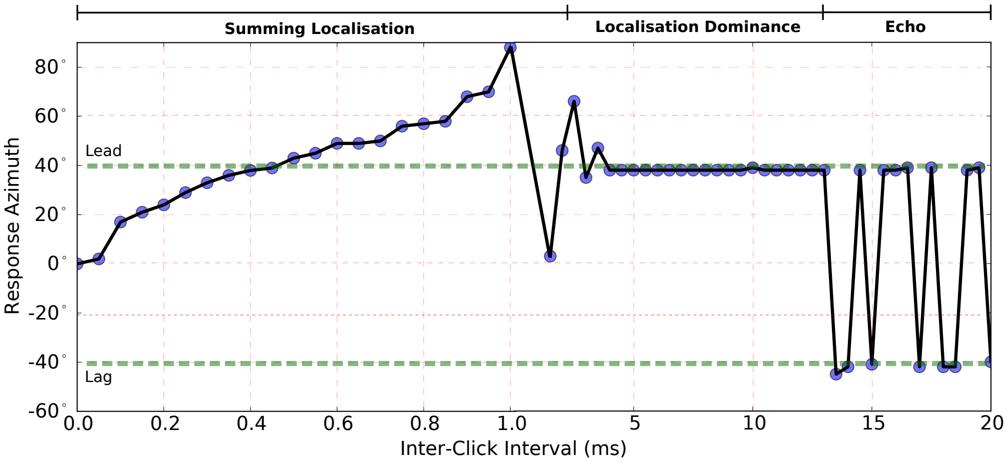
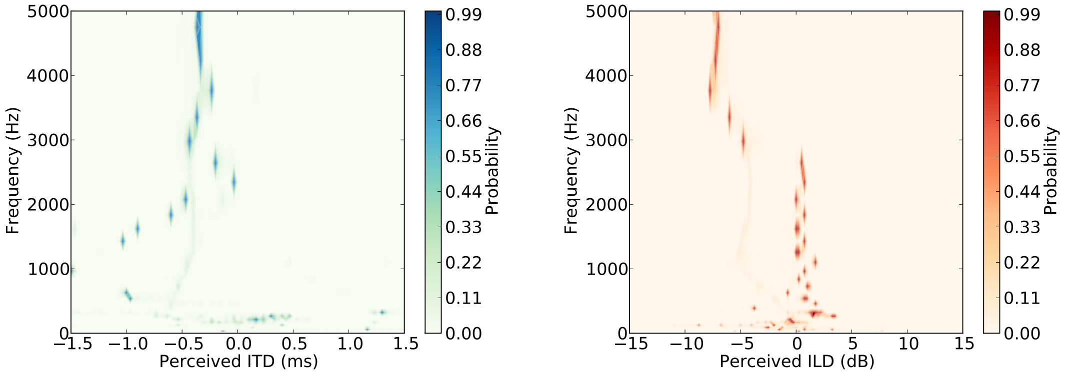


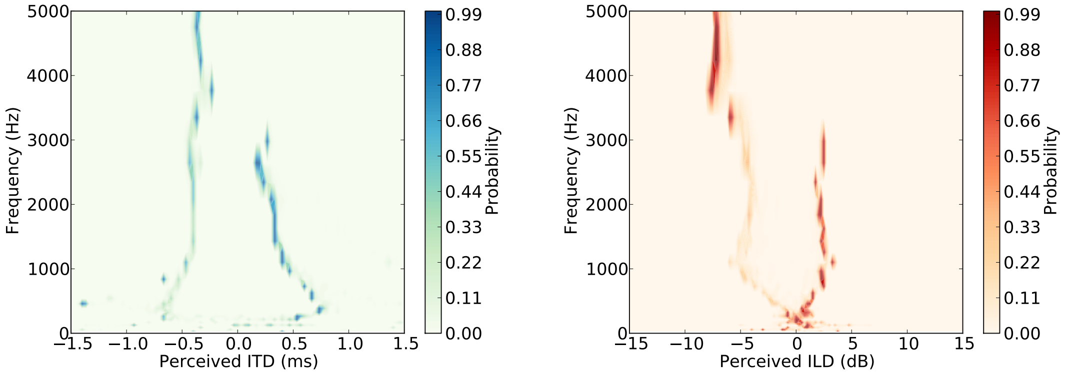
Figure 7.9: Model response as function of ICI. Dots indicate the exact response azimuth. The real azimuths of the lead and lag are marked in dashed lines. Note that the left and right panes use different scales for the time axis.

−5dB to about −7.5dB at 4.5kHz, which is consistent with ILDs of a real source at the position of the lead. As in this case ITDs feature a high variability, they only have a small contribution to the summary correlation, which is evident in the 0° component in Figure 7.11(a). ILDs, however, are more consistent around a single value, which results in a stronger component in the summary correlation. Accordingly, the final across-frequency integration yields a response of 38° which is only −2° from the true location of the lead.

At an ICI of 14ms, two distinct sets of cues are visible in both ITD and ILD plots. The darker image corresponds to the source which generates an interaural coherence considerably higher than c_0, whereas the lighter image implies that the source's interaural coherence only borders c_0. At frequencies above about 3kHz only a single image is visible, which is a result of c_0 increasing with frequency. Figure 7.11(b) clearly shows two strong spatial components, at −42° and +38°, which implies audibility of two separate sources. Because the integration process gives a higher weight to frequencies in the 600Hz range, in which the negative source generates a higher interaural coherence, then the −42° component is slightly stronger in the summary correlation, and the model decides in favour of that source. This exemplifies the importance of qualitatively analysing the intermediate model outputs, rather than strictly relying on a final localisation judgement.

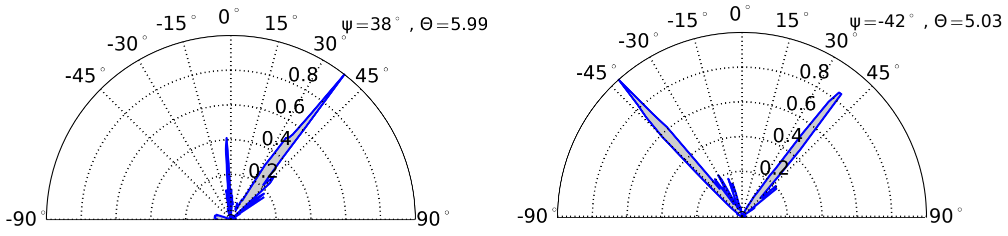


(a) ICI=4.5ms (b) ICI=4.5ms



(c) ICI=14ms (d) ICI=14ms

Figure 7.10: Output of the binaural processor for ICIs of 4.5ms (top panes) and 14ms (bottom panes), for two coherent sources at ±40°. Left - ITDs, Right - ILDs. .



(a) ICI=4.5ms (b) ICI=14ms

Figure 7.11: Summary correlation for and ICIs of 4.5ms and 14ms, for two coherent sources at ±40°.

At the summing localisation range, above ICI=0.5ms, the response azimuth exceeds the position of the lead and continues to rise as the ICI approaches 1.0ms. To gain further insight into this peculiar behaviour, it is useful to conduct a more specialised experiment focusing on this phase of the precedence effect. For this, the same experimental setup described in Section 7.2 is used, only with ICTD instead of ICLD as the independent variable. Figure 7.12 depicts the model's response azimuth as function of ICTD, which plotted over the subjective data of de Boer (1940) and Wendt (1963).

There appears to be good agreement between the model and de Boer's data within the range of ±0.4ms, above which the model seems to extrapolate the subjective data (which was not available in this range). The curve generated from Wendt's data is less linear and less steep than the modelled results, which is interesting, as the main (known) difference between Wendt and de Boer's experiments is the ability of subjects to move their heads. Thus, in a similar manner to the experiment reported in Section 7.2, one would expect that modelled results would resemble Wendt's data rather than de Boer's data.

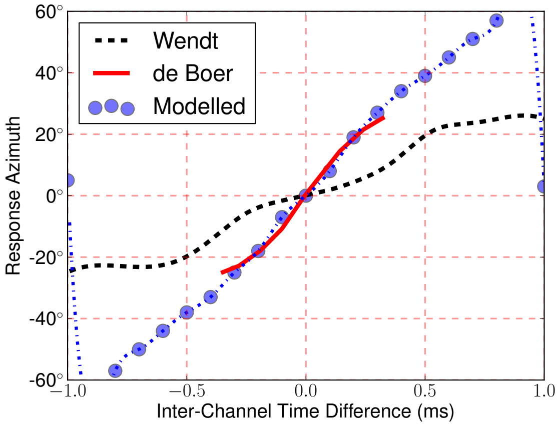


Figure 7.12: Results of modelling summing-localisation using inter-channel time differences, for two coherent sources at ±40°. Solid line - data fitted from (de Boer, 1940), dashed line - data fitted from (Wendt, 1963). Dotted line - modelled results.

Another interesting point is that at an ICTD of about 0.9ms, the response azimuth changes direction and returns towards 0°. The same effect is observed in Figure 7.9 at an ICI of 1.5ms, above which cues begin to fluctuate around the location of the lead. Possible reasons for this are discussed in the following section.

7.3.1 Discussion

The experiment reported in this section effectively validates that the model can account for the different phases of the precedence effect, which provides a basis for generalisation to more complex listening situations. Within the summing localisation range, the model correctly predicts the trend by which increasing the ICI results in a shift of the auditory event towards the location of the lead. Above $ICI = 0.5$ms, the model's response appears to be exaggerated in comparison to subjective data, however, it was established in (Blauert, 1997) that for the case of two coherent sources with an ICTD, *"auditory events may in some cases appear outside the range of angles bounded by the loudspeakers"* (p.207). Unfortunately, the extent of this phenomena is still not well understood, as most loudspeaker-based localisation tests reported in the literature are focused at a significantly higher ICTD range, e.g. in (Haas, 1951; Meyer and Schodder, 1952; Seraphim, 1961) as reported in (Blauert, 1997).

At a range of ICI=$1.0 - 1.5$ms the model misjudges the location of the source, and places it at around 0° rather than at the location of the lead. A possible explanation of this is related to the internal structure of the model. When the model is presented with a target set of interaural cues, it attempts to localise the source by finding a similar cue pattern from the previously-learned templates. ICIs at the range of $1.0 - 1.5$ms are too low for the model's precedence mechanism to be activated (the variability of low frequency ITD values is still not high enough such that they will not affect the integration process), thus summing-localisation is inherently performed. For example, an ICI of 1.5ms will result in an ITD of 1.5ms across most of the frequency range, but an ILD of 0dB at high frequencies (since no ICLD is applied). As there is no such naturally occurring combination of ITD/ILD values, then when the model performs the 2D template matching procedure only ILD values will contribute to finding an appropriate template, which in this case, would most likely resemble the 0° template (for which the ILD is indeed, 0dB).

At ICIs above 1.5ms up to about 12ms, the model predicts that the auditory event is localised at the lead, which is consistent with results reported in the literature (Litovsky et al., 1999). As the ICI is increased, the interaural coherence generated by the two sources approaches the value of c_0, above which, the model no longer suppresses cues from the lagging source. Thus, when this occurs in a sufficient number of auditory bands (or in those which are more strongly weighted in the integration process), then the model localises two separate auditory events. In the example used in this section, such an echo threshold occurs above 12ms, however this may be controlled by using a different value of μ. In the literature, various echo thresholds are reported, which for click stimuli range from $5 - 10$ms (Freyman et al., 1991; Yang and Grantham, 1997) and even up to 20ms in some cases (Litovsky and Godar, 2010)[2].

In (Faller and Merimaa, 2004b), a similar experiment was conducted, which effectively showed that the binaural processor used in the model can account for the three phases of the precedence effect. In their experiment, however, Faller and Merimaa assumed that changing the ICI (or ICTD, in this case) is perceptually equivalent to a new listening situation, and as such, the value of c_0 needs to be adapted accordingly. Additionally, they focused on probability density functions at the 500Hz band alone. Here, the entire frequency range was evaluated using a new approach for choosing c_0 as function of critical band. The same $c_0(\dot{k})$ slope was used (with $\mu = 0.2$) in *all* ICI combinations, effectively showing that c_0 does not need to be altered for different ICIs, which is an improvement to Faller and Merimaa's original model. This is also sensible from a psychophysical point of view, as changing c_0 corresponds to the auditory system being adapted to a new listening situation. In an everyday room, reflections normally arrive at the ear in different delay times (which here would correspond to different ICIs), however, all reflections can be seen as part of the same listening situation[3].

---

[2] Results from the cited paper refer to experiments conducted in adults.

[3] This of course depends on the (somewhat fluid) definition of *listening situation.*

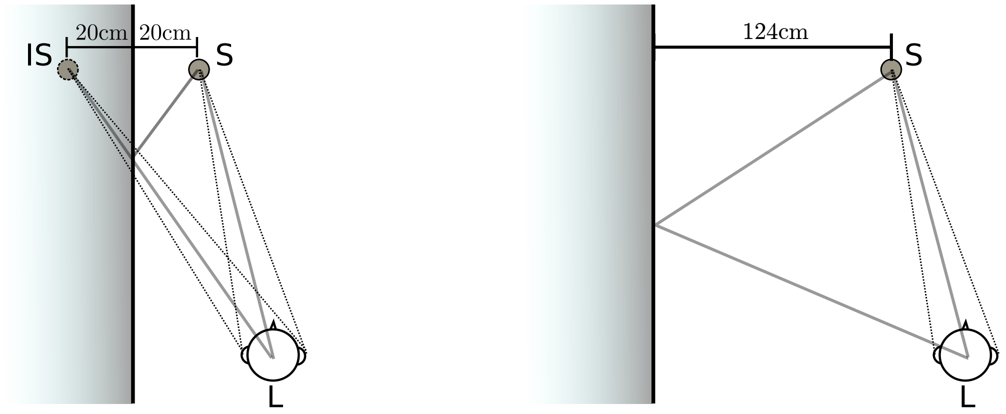
7.4 Effects of a Single Reflection

When a sound source is placed close to a boundary, a mirror image is formed. If the boundary is reflective and of smooth geometry, then it reflects a coherent wave with a time delay and amplitude difference, which are proportional to the difference in path lengths between the real and mirrored sources. Accordingly, such source-boundary interaction can be seen as a more composite localisation task, involving both time and intensity differences, as opposed to the more idealised tests reported in Sections 7.2 and 7.3. The goal of the experiment reported in this section, is to exemplify how the integrated model can be applied to some fundamental problems in room acoustics involving sound localisation at the presence of real reflections.

7.4.1 Methodology

The experiment involves a sound source placed at an angle of $-10°$ and a radial distance of 1.5m from a listener, as shown in Figure 7.13. The room model was designed such that a single reflecting surface ($\xi_w = 192$, at all frequencies) was placed to the left of the virtual listener. Two scenarios are studied involving source-boundary distances of 20cm and 124cm. For these two cases, the sources generate mirror-images at $-24°$ and $-61.7°$ from the listener, which would result in direct-to-reflected time differences of 0.35ms (summing-localisation region) and 4.86ms (localisation dominance region), respectively. All used stimuli were white noise bursts (200ms, 20ms $\cos^2$ ramps), and the interaural coherence threshold was set such that $\mu = 0.2$. Other than that, all parameters were as described in preceding sections.

To evaluate the effects of the boundary on localisation, two simulations were carried out according to the experimental setup shown in Figure 7.13. An additional simulation, in which the boundary is completely removed, was used as reference. The resulting conditions are compared in terms of generated binaural cues and cue correlation diagrams.



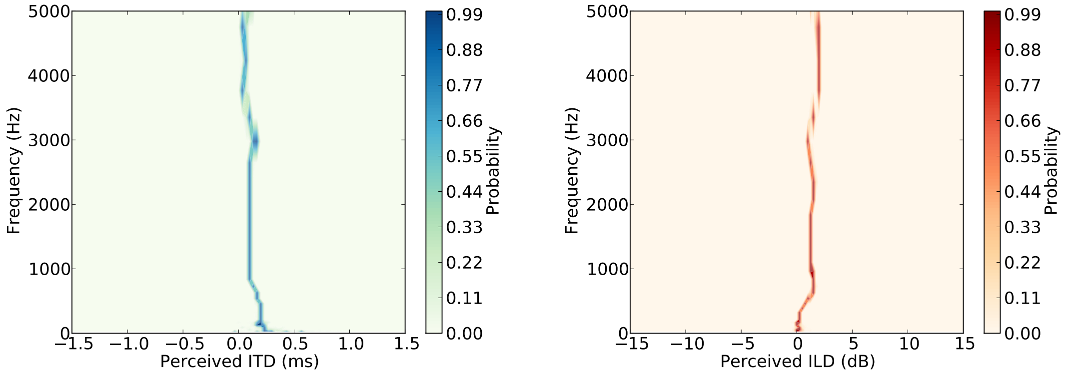
(a) Source near boundary (b) Source away from boundary

Figure 7.13: Experimental setup for two cases of localisation at the presence of a single reflection. (a) source positioned 20cm from wall, generating a path difference of 12cm ($\Delta t = 0.35$ms). (b) source positioned 124cm from wall, generating a path difference of 161cm ($\Delta t = 4.68$ms). Both sources incident at $-10°$ relative to listener. The image-source at (b) is removed for clarity.

7.4.2 Results and Discussion

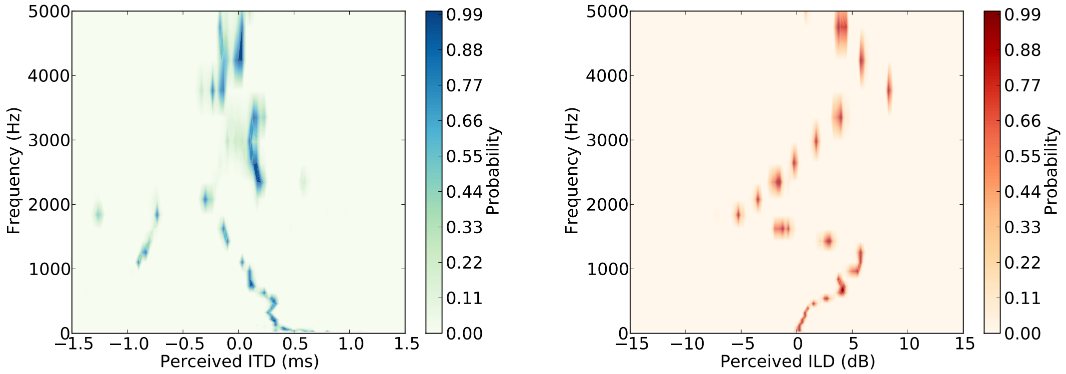
Figure 7.14 shows CPPs for the two cases of a single reflection and for the reference condition. It can be seen that the reference simulation yields clearly-defined cues, with narrow probability functions, as expected in a free field situation. For the case of the source in close proximity to the boundary, there is a clear shift towards the positive ITD and ILD range at low frequencies, suggesting a locus shift towards a negative azimuth. For the case of a source away from the boundary, the binaural cues much better resemble the free field condition across the entire frequency range. The probability functions, however, are much wider and some discrepancies are evident at low frequencies suggesting a localisation judgement with decreased confidence.

These effects are further demonstrated in the cue correlation diagrams, as shown in Figure 7.15. For the case of the source near a boundary, the positions of high correlation values are farther left in comparison to the reference condition, which is suggestive of a locus shift. This is to be expected as the reflection arrives within the summing-localisation time region. It is mostly evident at lower frequencies where ITDs have a stronger effect on the integration processes, whereas



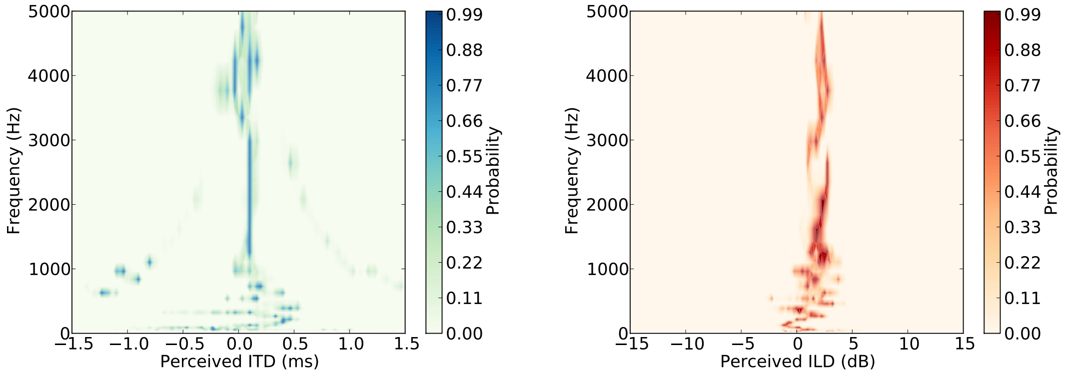
(a) No boundary

(b) No boundary



(c) Source 20cm from boundary

(d) Source 20cm from boundary



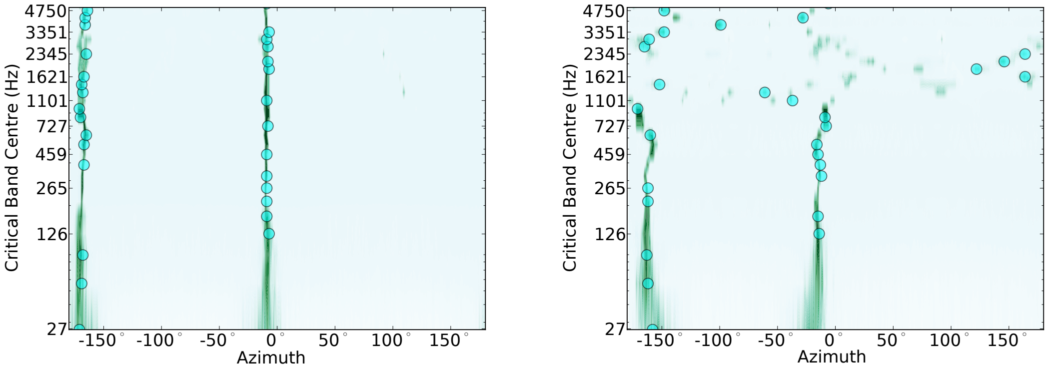
(e) Source 124cm from boundary

(f) Source 124cm from boundary

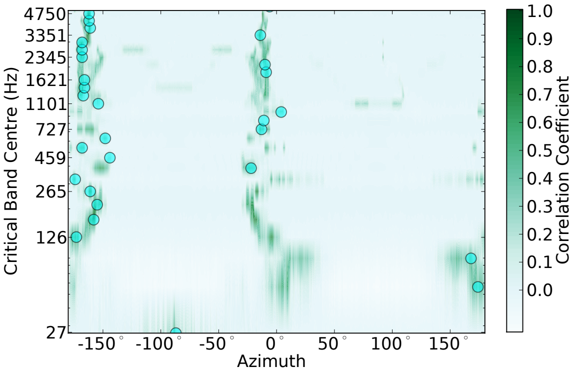
Figure 7.14: Output of binaural processor for a source at −10° and 1.5m away from a listener in (a)(b) free field, (c)(d) close to a reflecting boundary and (e)(f) away from a reflecting boundary .

at higher frequencies, the resulting binaural cues seem to correlate well with those of other possible azimuths, suggesting a less confident localisation judgement.

For the case of the source away from the boundary, high correlation values are less scattered, however front to back confusion is much more pronounced than for the reference condition, which is evident in the high correlation values around −170°. Furthermore, in comparison to the reference condition, high correlation points are less aligned along a single azimuth, which also suggests that the final localisation judgement is made with less confidence.



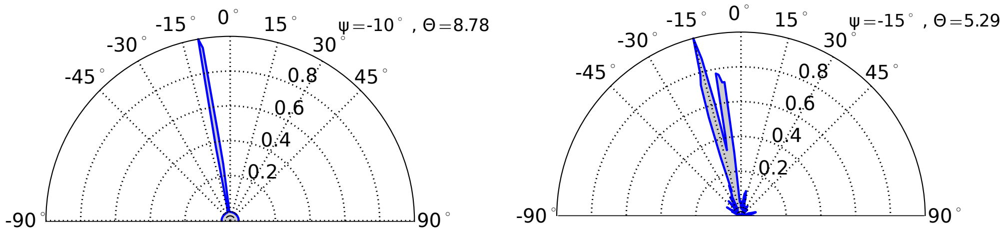
(a) No boundary (b) Source 20cm from boundary



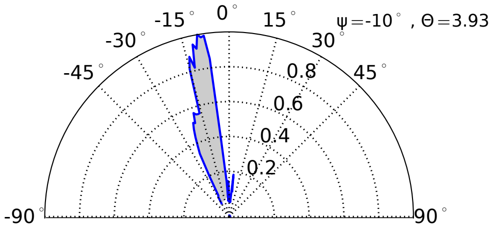
(c) Source 124cm from boundary

Figure 7.15: Template correlation response for source at −10° and 1.5m away from a listener in (a) free field, (b) close to a reflecting boundary and (c) away from a reflecting boundary. The same correlation coefficient scale is used in all three subfigures.

Figure 7.16 shows the summary correlation plots for the three cases. For the reference condition, the model correctly localises the source at −10°. With a reflecting surface at close proximity, an image-shift of −5° is introduced and the



(a) No boundary (b) Source 20cm from boundary



(c) Source 124cm from boundary

Figure 7.16: Summary correlation plots for source at $-10°$ and 1.5m away from a listener in (a) free field, $\Psi = -10°$, $\Theta = 8.78$; (b) close to a reflecting boundary, $\Psi = -15°$, $\Theta = 5.29$ and (c) away from a reflecting boundary, $\Psi = -10°$, $\Theta = 3.93$.

model's response is $\psi = -15°$. Considering that the surface is nearly fully reflective, then the amplitude of the image source is nearly identical to the real source, and as both wavefronts arrive within the summing localisation time range, the auditory event is 'pulled' towards the image source. In a similar manner, the path difference between the direct and reflected waves is 12cm, meaning that the reflected wave arrives at a time lag of 0.35ms, which causes the auditory event to be 'pushed' in the opposite direction (i.e. towards the 'earlier' source). Thus, the final localisation judgement can be seen as a case of time/intensity trading (Franssen, 1964; Blauert, 1997) where intensity trumps time. Observation of Figure 7.15(b) shows that at low frequencies high correlation values are much better aligned than at high frequencies. The reason for this is that at low frequencies, the 2D correlation algorithm operates only on ITDs, which get defined by the sources' level differences rather than time differences. Accordingly, when the model performs across-frequency integration of the correlation response, it gives a higher weight to low-frequency bands which are dominated by ITD cues. This explains the model's final decision to bias localisation in the negative direction, i.e. towards the image source. Also observe that when the reflecting boundary is present the summary correlation is more noisy. This manifests itself in a reduced confidence, $\Theta = 5.29$ compared to $\Theta = 8.78$ for the reference condition, which implies an increase in localisation blur.

For the case of a source 124cm away from the boundary, the model's response is $\psi = -10^\circ$ indicating no shift in the locus of the source. This is to be expected as in this case the time delay between the direct and reflected components of the soundfield is 4.68ms, which is within the precedence region. Although this response is similar to the reference condition, it can be seen from the summary correlation plot that the curve is much wider, meaning that the decision is made with less confidence ($\Theta = 3.93$) in comparison to the reference condition as well as to the close-proximity situation.

7.5 Effects of Occlusion and Diffraction

All of the experiments reported so far were conducted under relatively simple acoustic settings. In other words, binaural responses used in the perceptual model could have been easily obtained analytically or by making use of an image-source model. One advantage of the FDTD method is that wave effects such as diffraction and superposition are inherent in the method. Thus, to exemplify where the integrated model would really benefit from using the FDTD method in the room simulation stage, a listening situation involving sound diffraction and occlusion is studied in this section.

In everyday listening situations sound sources are not always in direct line-of-sight to the listener, as they may be present behind interfering obstacles, such as furniture, room corners and other listeners. Wave diffraction around such objects alters the original wavefront's direction of arrival, and consequently the localisation cues available to the auditory system. When diffraction around an entire object occurs, two separate propagation paths are generated from the two sides of the object. As such, the auditory system is presented with two sets of cues, in a similar manner to a two source situation. Thus, depending on the physical size of the object, the two wavefronts may arrive at the ears at different delay times, which would result in different localisation judgements.

7.5.1 Methodology

The experimental setup involves a virtual room corresponding to physical dimensions of 4x4x3m with a rigid barrier placed at the centre of the domain, and a receiver placed at a distance of 70cm from the rear wall, as shown in Figure 7.17. Five simulations were executed, with the sound source placed at the positions $P0$, $P1$, $P2$, $P3$, $P4$, corresponding to azimuths of $0°$, $-5°$, $-15°$, $-25°$ and $-35°$ respectively. Due to computational constraints and the additional propagation time caused by diffraction, it is not possible to make the domain large enough to exclude reflections from the outer boundaries. Accordingly, the outer boundaries were made absorptive by setting $\xi_w = 1$. As employed the absorption model does not work very well at some oblique angles of incidence (Kowalczyk and van

Walstijn, 2008), an additional reference experiment was executed, in which the occluding barrier was removed. This is to ensure that reflections arriving from the outer boundaries are of small enough magnitude, such that would do not mislead the model's localisation judgement.

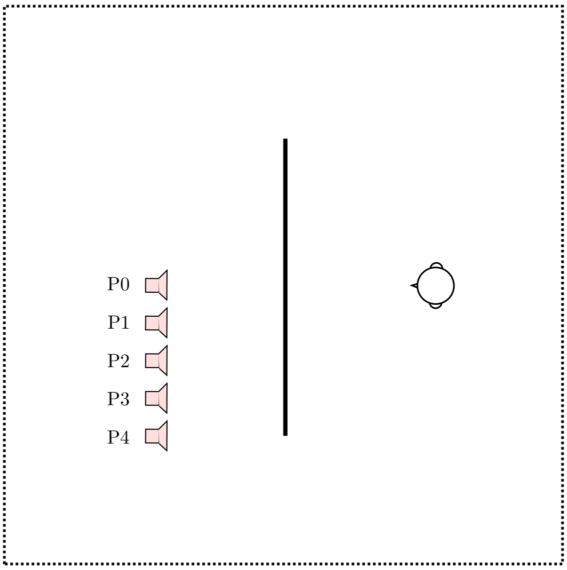
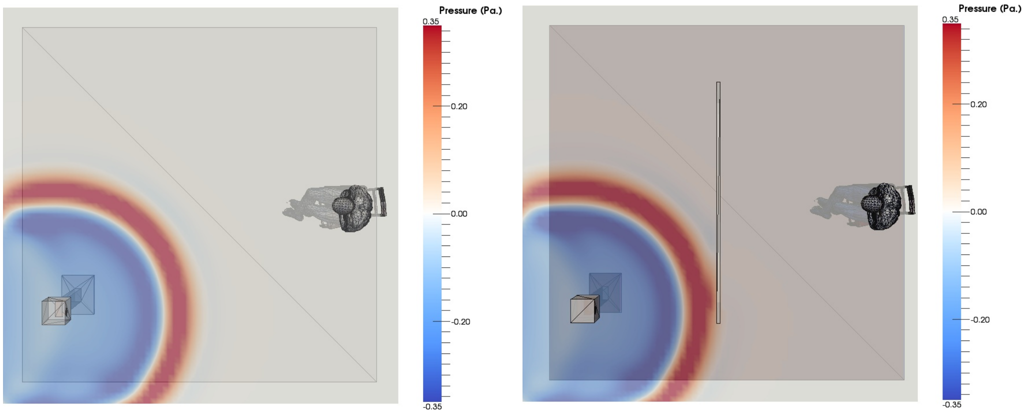


Figure 7.17: Experimental setup used in the occluded source investigations. The source positions correspond to azimuths of $P0$ 0°, $P1$ $-5°$, $P2$ $-15°$, $P3$ $-25°$, $P4$ $-35°$.

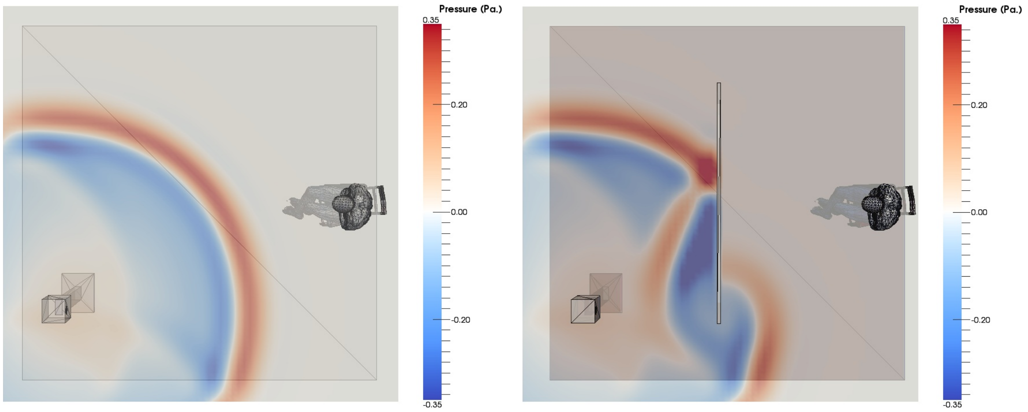
7.5.2 Results and Discussion

To exemplify the physical effects of source occlusion, Figure 7.18 shows the output of the room model at three different time steps. Observe that due to diffraction around the barrier, two wavefronts arrive at the listener from the two edges of the barrier. In this example, the source is close to the left side of the barrier, thus the lead wavefront arrives from the left of the listener (solid arrow), which is then followed by a lag wavefront from the right (dashed arrow).

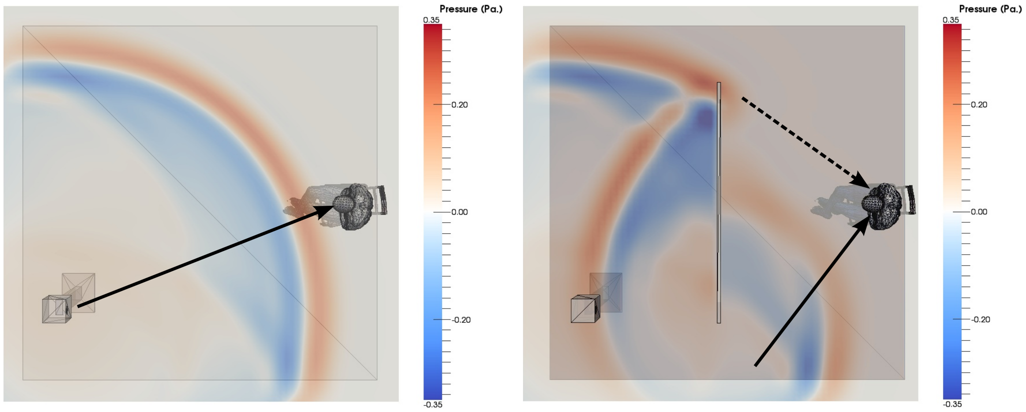
Figure 7.19 shows the model's response as function of target azimuth for the five tested source positions. At $P0$, even when the barrier is present, the two wavefronts arrive at the listener's ears at the same time as the two propagation paths around the barrier are similar. Accordingly, the model localises the source at 0° for both the occluded and reference conditions. At other source positions, the wavefront arriving from the left of the listener trigger a precedence mechanism



(a) 0.76ms (b) 0.76ms



(c) 1.06ms (d) 1.06ms



(e) 1.38ms (f) 1.38ms

Figure 7.18: Simulation of a sound source in free field (left panes) and behind an obstacle (right panes), captured at three different time steps. Loudspeaker and full-size listener are only for visualisation purposes and are not included in the calculation. Solid and dashed arrows denote the wavefronts' directions of arrival.

which causes directional information from the lagging wavefront to be suppressed, and the model localises the source approximately at the left edge of the barrier.

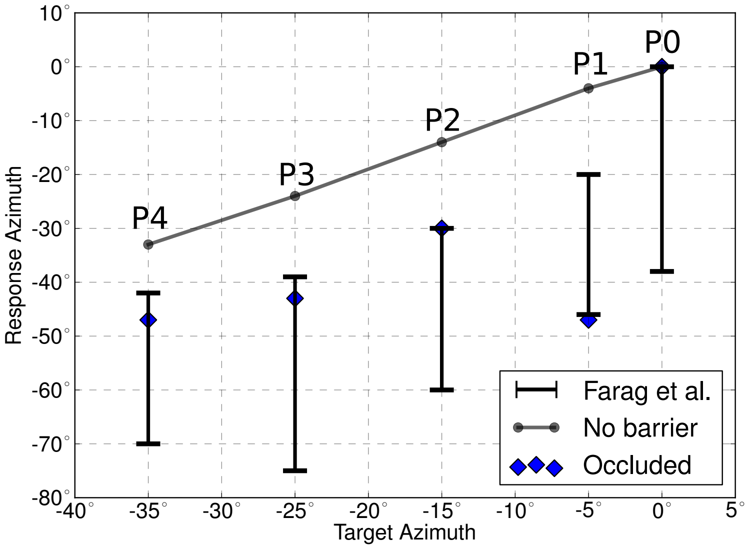


Figure 7.19: Results for an occluded source simulation. Solid line - reference condition, Diamond markers - occluded source condition, Errorbars - quartile edges of human responses from (Farag et al., 2003).

Farag and colleagues (2003) studied the subjective effects of sound occlusion, with two different barrier sizes. Unfortunately, replicating their exact experimental setup was not possible because of computational constraints as well as due to missing information on the precise distances of the sources and listener relative to the barrier in the reported study. However, as the source angles and barrier dimensions used here are similar to Farag et al., it is useful to compare modelled results to their subjective data. Accordingly, the upper and lower quartiles of their data are also shown in Figure 7.19. It can be seen that all modelled responses are within the interquartile range, excluding $P1$ which is slightly lower. Although interquartile ranges are quite wide, in essentially all cases the locus is exaggerated in comparison to the reference condition, which is expected as the occluding object affects the direction of arrival of the leading wavefront. This indicates that for the case of an occluded source, the model responds in a manner similar to human subjects.

7.6 Effects of Reverberation

In order to localise sound in a reverberant field, the auditory system needs to distinguish between components of the soundfield which carry meaningful spatial information and those who don't. It was already shown in Chapter 6 how the model accomplishes this by making use of interaural coherence. In this section, a more detailed investigation of the model's performance in a reverberant field is presented. The goal is to evaluate how the model's predictions correlate with human behaviour, in terms of localisation error and localisation confidence under reverberation.

7.6.1 Methodology

A small rectilinear room of 3.8x3.0x2.7m was simulated with a source placed at a radial distance of 2m from the receiver, as illustrated in Figure 7.20. As localisation accuracy is expected to vary with source azimuth, a BRIR was recorded at four different source positions corresponding to azimuths of 0°, 30°, 45°, and 60°.

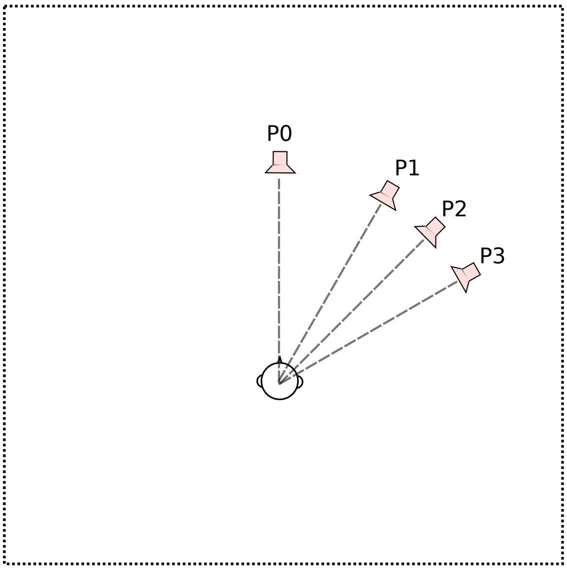


Figure 7.20: Experimental setup used for investigating the effects of reverberation. Source positions correspond to azimuths of $P0$ 0°, $P1$ 30°, $P2$ 45°, and $P3$ 60°, with a radial distance of 2m from the receiver.

The dependent variable in this experiment is the absorption coefficient of the room's surfaces, which in turn affects the room's direct-to-reverberant ratio

Table 7.1: Absorption coefficients α, normalised surface impedances ξ_w and total absorption $S\hat{\alpha}$, used in the reverberation experiment.

α	0.15	0.30	0.45	0.60	0.75	0.90
ξ_w	24.62	11.24	6.74	4.44	3.00	1.92
$S\hat{\alpha}$ [m^2sabin]	8.92	17.85	26.78	35.751	44.64	53.56

(D/R) and reverberation time (T_{60}). In small rooms, particularly when significant absorption is involved, such statistical measures do not characterise the soundfield very well. Accordingly, focus is given here to the total amount of absorption in the room, which is measured by

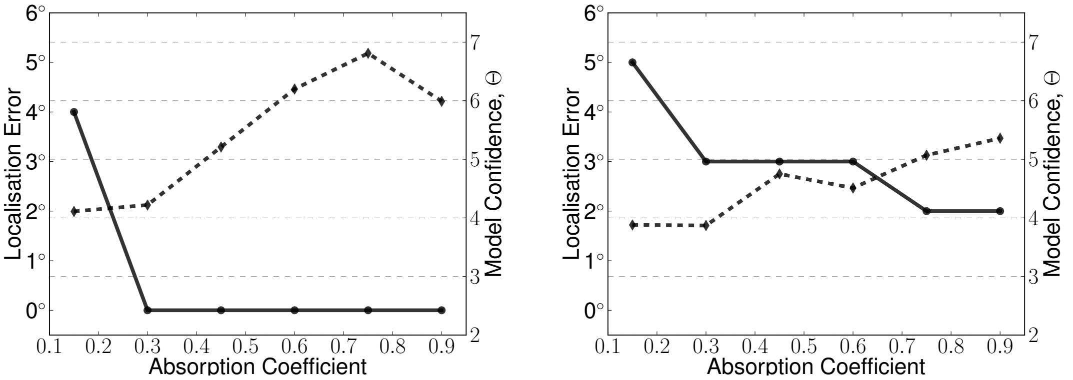
$$S\hat{\alpha} = \sum_i S_i \alpha_i \tag{7.2}$$

where α_i and S_i are the acoustic absorption coefficient and total area of the i^{th} surface, and $S\hat{\alpha}$ has the dimension of m^2*sabin*. As shown in Table 7.1, the experiment was repeated with six absorption values for each source position, totalling in 24 individual simulations. All room walls were allocated the same α, which was held constant across frequency. The stimulus used for localisation was a broadband noise burst (200ms, 20ms $\cos^2$ ramps), and simulations were executed to obtain impulse responses of 600ms duration. Otherwise all model parameters were as described in Section 7.3.

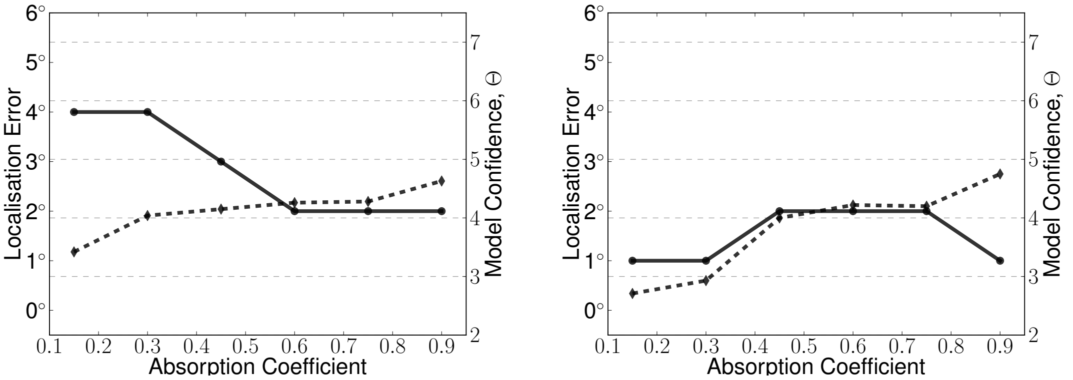
7.6.2 Results and Discussion

Figure 7.21 shows the model's localisation error $|\theta - \Psi|$, and confidence, Θ, as function of absorption coefficient, for the four different angles of incidence. It is evident that, in general, the model's localisation error decreases as the absorption coefficient increases. Conversely, the model's confidence tends to increase as the absorption coefficient becomes higher. The only exception to this is at the source azimuth of 60°, where localisation error slightly increases (and then decreases) as the absorption coefficient is increased.

Figure 7.22 shows the mean localisation error and confidence, as function of absorption coefficient and source azimuth. It is evident that as the room becomes



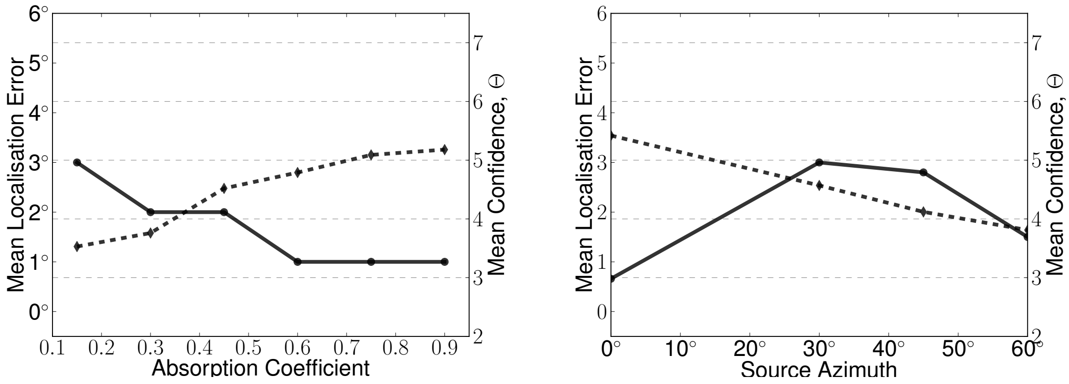
(a) Source at 0°　　(b) Source at 30°



(c) Source at 45°　　(d) Source at 60°

Figure 7.21: Localisation error (solid lines) and model confidence Θ (dashed lines) for simulations executed at four different source positions.

less reverberant, the model localises the source with more accuracy and with more confidence, which is in agreement with results reported in the literature (Giguère and Abel, 1993). Using octave-band filtered noise, Giguère and Abel studied localisation in a room of similar volume whose surface absorption was adjusted to yield reverberation times in the range $0.2 \leq T_{60} \leq 1.0$. They discussed that localisation scores of their subjects were consistently lower in a reverberant environment, and that this was independent of frequency or the lateral position of the source. A similar trend was observed by Hartmann (1983), who used broadband noise stimulus, however in larger rooms having significantly higher reverberation times ($1.0 \leq T_{60} \leq 5.5$).



(a) (b)

Figure 7.22: Mean localisation error (solid lines) and model confidence (dashed lines), as function of (a) absorption coefficient and (b) source azimuth.

Another interesting point is the performance of the model in terms of source azimuth. It can be seen in Figure 7.22(b) that the mean localisation error appears to be higher at more lateral source positions, however this does not follow a consistent trend. However, in a similar manner to a free field situation (see Section 7.1), the model's confidence is decreased as the source moves laterally, as much as can be concluded from the four sampled azimuths. This trend also resembles the results of Giguère and Abel (1993), who showed that their subjects' performance was poorer as they moved the source laterally.

7.7 Summary

In this chapter, the performance of the integrated model was evaluated in different listening situations. It was shown that the model performs flawlessly in a free-field, when using individualised HRIRs. Although this outperforms human behaviour, it was shown that the model's confidence effectively indicates a localisation trend similar to that of human subjects. Accordingly, it was argued that the confidence metric can be used to complement the model's final localisation judgement. When testing the model with non-individualised HRIRs, it was shown that considerable front-back confusions are demonstrated, which is consistent with subjective findings.

The model was then evaluated in a controlled environment including two coherent sources whose time and level differences were parametrically modified. When level differences were introduced, the model was able to precisely replicate the subjective results of Wendt (1963). When time differences were introduced, the model was able to predict the three phases of the precedence effect, namely summing localisation, localisation dominance and echo. Within the time range of $0.5 - 1.0$ms, there was some disagreement between the model's predictions and subjective data, however, some correlation was found with the results of de Boer (1940).

Once the model has been validated in idealised listening situations, its applications to room acoustics research were further explored. It was shown how the model can be applied to study the effects of a single reflection, which sometimes involve conflicting ITD and ILD cues. To exemplify the advantages of using an FDTD algorithm for the room, the model was presented with a localisation task involving sound diffraction around a barrier, the results of which have been shown to be in agreement with a subjective study by Farag and colleagues (2003). Finally, the model was tested in a listening condition involving multiple reflections, and was shown that as expected from human behaviour, the model's performance is decreased as a room becomes more reverberant.

—I love deadlines. I like the whooshing sound they make as they fly by.

Douglas Adams

8

Conclusion

8.1 General Conclusions

The goal of this thesis was to develop a generalised model of human sound localisation in rooms on the horizontal plane. Unlike existing models, the effort here was to include modelling components in both the acoustic and psychoacoustic domains, thus allowing for complex listening situations to be evaluated entirely using virtual technologies. Furthermore, emphasis was given to generality and applicability to problems in room acoustics and spatial audio. This goal was accomplished by contributing to two independent aspects, which together constitute an integrated model of sound localisation:

1. **Finite Difference Modelling of Room Acoustics.** An investigation of grid excitation methods in FDTD (Chapter 3) have resulted in new knowledge regarding physical and numerical constraints in sound source mod-

elling. Accordingly, existing FDTD excitation methods were reviewed in light of this knowledge, which has shown that all known inherent problems (e.g. source scattering, numerical consistency and solution growth) can be avoided altogether when adhering to certain constraints, depending on the numerical scheme being used. To exemplify how these problems can be avoided, a new method for exciting an FDTD grid was developed and validated. Additionally, a novel approach to generalise source modelling using cascaded filters was suggested, and existing source methods were shown to be special cases of this approach. The main implication of this contribution on the goal of the thesis is that it was made possible to design sources that generate a prescribed pressure field, thus allowing for the sound pressure level at the virtual listener's ears to be controlled parametrically.

To be able to model problems whose bandwidth sufficiently cover the frequency range in which interaural cues dominate, the FDTD method was adapted for execution on general purpose graphics hardware (GPGPU). In Section 5.2 an accelerated parallel FDTD implementation was described, and was shown to yield a significant (up to x80) computation speed improvement over a traditional CPU implementation.

In order to obtain interaural cues required for the perceptual modelling stage, a method for embedding binaural receivers in an FDTD grid was suggested and evaluated. The concept is to use a laser-scanned 3D head geometry, and in itself, is not a novel one. However, this is the first time where this approach was investigated in context of room acoustics modelling using the FDTD method, and it was verified that correct interaural cues can indeed be generated by means of directly embedding objects approximating the head in the grid.

In Chapter 5, computer software for room acoustics modelling using FDTD was introduced and named *WaveCloud*. The software itself has been the underlying engine for obtaining room responses used throughout the thesis, yet, it was designed with future applications in mind. It is intended to be the foundation of an open-source collaborative project, and has already been used outside of this research by faculty members at Salford University. Although not directly used in the perceptual evaluations reported in the thesis, a novel method for modelling frequency dependent sources and boundaries (Section 5.5) was suggested and implemented in WaveCloud, thus allowing for more realistic conditions to be modelled in the future.

2. **Modelling Human Sound Localisation in Rooms.** The interaural coherence approach suggested by Faller and Merimaa (2004b) was used as the core of the binaural processor described in Chapter 6. The processor was complemented with a central processor based on supervised learning and a decision maker based on across-frequency integration, which together form a unified model emulating a listener in a forced-choice test. The concept of a frequency dependent interaural coherence threshold was introduced, and was shown to be instrumental in the generality of the model. Accordingly, it is now possible to model the entire process of human sound localisation, from the peripheral auditory system to the final decision in the brain, in a manner general enough to cover a wide range of problems in room acoustics and spatial audio.

   The source, room and listener models discussed in Chapters 4 and 5 were integrated with the perceptual model introduced in Chapter 6, resulting in a single scientific framework for studying sound localisation in complex listening environments. In Chapter 7, this model was tested in idealised conditions involving localisation in free field, and with two coherent sources emulating the direct and reflected components of a soundfield. Modelled results have been shown to be in good agreement with subjective data found in the literature, and accordingly, it was concluded that the model is psychoacoustically plausible. Some disagreement was found in the special case of summing localisation where interchannel time differences are in the range of 0.5-1.0ms, and possible explanations for this were given. This discrepancy opens the door for future research, which could either improve the model but may also shed more light on human behaviour in such special cases.

> To demonstrate how the integrated model could be applied to problems in room acoustics, three problems were studied. The effects of a single reflecting surface was modelled and it was shown how a qualitative analysis of the model's output could be used in a perceptual investigation. To exemplify the benefits of using a wave-based method in the acoustic modelling stage, the integrated model was presented with an occluded source listening situation, involving wave diffraction. Lastly, the integrated model was employed in a reverberant environment, which naturally involves multiple reflections. In all cases it was shown that modelled results are consistent with human behaviour.

Accordingly, it can be concluded that the aim and objectives of the thesis, as described in Section 1.2, have been met. Early room models were aimed at a prediction of objective metrics, such as reverberation time or frequency response. These models were then complemented with the ability to generate binaural room responses for auralisation. In this work, the process of auralisation was replaced by a perceptual model, effectively allowing a machine to evaluate sound localisation in rooms, which is a single yet important subjective property. Thus, the general implication of the thesis is the ability to model the entire process of sound localisation, from physics to perception, which may be considered as a step forward in the evolution of computational room acoustics.

8.2 Future Challenges

The work on this thesis has opened doors for some interesting topics for future research. The physically-constrained source model developed in the thesis is limited to a point source which is mechanically uncoupled from the grid. It would be useful to develop mechanical filters simulating the dynamics of more complex sources, such as a loudspeaker driver, which would also be coupled with the finite difference scheme for the room.

In context of FDTD usability, GPU technology has been proven to be effective in acceleration of a standard explicit FDTD scheme. However, there may be numerical methods which would benefit more from the GPU programming paradigm. In particular, spectral methods (Liu, 1998; Spa et al., 2010) make an extensive use of Fourier transforms which largely benefit from parallel implementations.

Modelling binaural receivers in FDTD was solved in this work by directly embedding a listener geometry in the grid. Whilst in context of this work this is an acceptable solution, to the author's opinion this approach is challenging to generalise. The main reason for this is that the head geometry essentially changes with the spatial sample period, and the method is thus numerically inconsistent across different sample rates. Thus, it is clear that the FDTD method would benefit from a more robust approach to model binaural receivers. One of the advantages of the FDTD method, is that the particle velocity field is either inherent or straightforward to solve for. Thus, an option would be to synthesise a binaural response based on directional information encoded in the particle velocity vector at the receiver, for example by making use of a spatial decomposition method (Tervo et al., 2013).

The WaveCloud project would also benefit from future work, which would ideally be based on a collaboration between scientists studying rooms using the FDTD method. Some future challenges include the implementation of sources whose directivity corresponds to measured data, further optimisation of the computational engine, an option to embed multiple receivers, a more comprehensive GUI and detailed documentation. Additionally, it is the author's intention to complete the work on a Scientific Python based API, which would allow users to run WaveCloud on a completely open-source framework.

There are a few limitations to the sound localisation model described in the thesis. From a technical perspective, the usable bandwidth is dependent on the computation power of the employed system. Thus, it would be interesting to put the model through a supervised learning process with measured HRTFs, and use it with measured (rather than modelled) binaural room responses. Additionally, the model was designed to localise sources on the horizontal plane only, and in most cases, in front of the listener. It would be useful to extend the model for localisation in elevation as well. This could be accomplished by adding a monaural processor aside to the existing binaural processor, and adapting the central processor accordingly. An additional challenge is to thoroughly investigate the effects of perceptual weights used for across-frequency integration by the decision maker, which could contribute to the accuracy of the model and possibly reveal more information about how humans perform this process.

In this work, the integrated model was used to exemplify only few listening situations, however it has the potential to be used in further research. As the model can account for numerous sources, it could be employed to investigate localisation in different sound reproduction paradigms, such as intensity stereo, ambisonics or wavefield synthesis. It would also be interesting to evaluate these paradigms with different room geometries, wall absorption and source directivity, and such an investigation could lead to a better understanding of the acoustic design of critical listening rooms. One could also use the model to evaluate acoustic treatment in auditoria, for example, to analyse whether distances and angles of reflecting surfaces would result in acoustic echo, or will go unnoticed as far as sound localisation is concerned.

Another interesting application of the sound localisation model is to evaluate the perceptual effects of different numerical modelling artefacts. For example, within the field of FDTD modelling, one could use the perceptual model to assess different source, boundary and receiver techniques. In the wider area of numerical simulation, it could be used to study differences between various methods for room modelling and auralisation.

A

Sample WaveCloud Script

In this appendix, a sample MATLAB script for solving a room using WaveCloud is given. Its purpose is to demonstrate how a WaveCloud model can be programmed, and to provide a quickstart guide which could be used until more thorough documentation is published. First, we initialise a new MATLAB struct `myModel`:

Listing A.1: Define new model

```
myModel=newModel;              % Construct new model
```

Next, we define some global modelling parameters:

Listing A.2: Global parameters

```
myModel=setSLF(myModel);       % Set numerical scheme
myModel.samplingFreq=40000;    % Set sample rate
myModel.runTime=1;             % Simulation length (s)
```

Note that one can replace `setSLF` with any other scheme code, e.g. `setIWB`, to change the numerical scheme being used. The next step is to define calculation parameters:

Listing A.3: Calculation parameters

```
myModel.gpuOptions.isSP=true; % true if single precision
myModel.onGPU=true;           % true for GPU kernel
myModel.compOptions.standalone=true;
myModel.showAnimate=false;        % true for visualisation
```

The `standalone` parameter is set to `false` only for debugging purposes which require that a larger portion of the code runs from within MATLAB. The `showAnimate` parameter is for coarse visualisation inside MATLAB, whereas generation of VTK files for offline visualisation is accomplished by using the `saveAnimate` parameters. We define the dimensions of the model by setting:

Listing A.4: Model extents (meters)

```
myModel.room.length = 4;
myModel.room.width = 4;
myModel.room.height = 3;
```

The next step is to define the type and position for the receiver:

Listing A.5: Receiver settings

```
myModel.elements.recX = 2;
myModel.elements.recY = 2;
myModel.elements.recZ = 1.5;
myModel = setRecBinaural(myModel, 0);
myModel.elements.headImpedance = 90;
```

The `setRecBinaural` function places two spaced pressure probes at the positions of the ears (a single pressure/velocity receiver can be defined instead of a binaural receiver by using `setRecMono` instead), and `headImpedance` sets the specific boundary impedance of the head object. Next the source parameters are defined:

Listing A.6: Source settings

```
myModel.elements.srcShape='maxflat';
myModel.elements.sourceAmplitude = 250e-6;
myModel.elements.srcFc=0.186;
```

The `srcShape` flag tells WaveCloud how to design the excitation function (e.g. `'Gaussian'` could also be used), `sourceAmplitude` is the peak amplitude, and `srcFc` is the normalised high cutoff frequency of the function. To design a physically-constrained source, one may use a dedicated function `designPCS` instead. The next step is to tell WaveCloud to calculate some intermediate variables which are necessary for subsequent functions:

Listing A.7: Intermediate calculation

```
myModel=calcVars(myModel);
```

Next, the mesh is created for the domain:

Listing A.8: Create the grid

```
myModel=meshRect(myModel, 415.6);
```

The second parameter represents the characteristic impedance of the boundary truncating the grid, which here is set to 415.6 to imitate absorbing boundaries at the edges of the domain. To overlay an arbitrary room geometry on the grid, we call the function:

Listing A.9: Load geometry file

```
myModel = loadGeom_stl(myModel,'geometry.stl',8000);
```

where `geometry.stl` is the 3D geometry file in STL format, and the last parameter denotes the characteristic acoustic impedance of the geometry's boundaries, in this case 8000. Thus to import a composite model which includes multiple materials, one would export each material as a different layer in a corresponding STL file, and multiple instances of `loadGeom` would be used. In a similar manner, we import the geometry of the head by calling the function:

Listing A.10: Load head geometry

```
myModel=addHead(myModel);
```

The head geometry is automatically decimated according to the chosen sample rate. Next, we add sources:

Listing A.11: Add sources

```
%addSource(myModel, x, y, z, amplitude, type, timelimit)
myModel=addSource(myModel, 1, 1, 1.5, 1, 'phys', false);
myModel=addSource(myModel, 0.5, 1.2, 1.5, 1,'phys',false);
```

The values corresponding to `x,y,z` denote the source position, `amplitude` is the *relative* amplitude of the source (e.g., 1=unchanged, 0.5=half of the defined amplitude), `type` defines the injection filter (`'phys'` is for PCS injection, `'soft'` and `'hard'` for soft and hard injection, respectively), and `timelimit` is used for debugging purposes only. It is possible to add as many sources as needed. Finally, simulation is executed by calling the function:

Listing A.12: Execute simulation

```
myModel=simulateModel(myModel);
```

For a binaural receiver, resulting impulse responses for the left and right ears are stored in the variables:

Listing A.13: Results

```
m.results.irl
m.results.irr
```

For a monophonic receiver, resulting impulse responses for pressure and particle velocity are stored in the variables:

Listing A.14: Results

```
m.results.irl
m.results.ux
m.results.uy
m.results.uz
```

Bibliography

J. Ahrens, B. Geveci, and C. Law. Paraview: An end user tool for large data visualization. In *the Visualization Handbook. Edited by CD Hansen and CR Johnson. Elsevier*, 2005.

M.A. Akeroyd and A.Q. Summerfield. A fully-temporal account of the perception of dichotic pitches. In *Br J Audiol*, volume 33, pp. 106–107, 2000.

V.R. Algazi, R.O. Duda, R. Duraiswami, N.A. Gumerov, and Z. Tang. Approximating the head-related transfer function using simple geometric models of the head and torso. In *The Journal of the Acoustical Society of America*, volume 112, p. 2053, 2002.

J.B. Allen and D.A. Berkley. Image method for efficiently simulating small-room acoustics. In *The Journal of the Acoustical Society of America*, volume 65, p. 943, 1979.

J.R. Angell and W. Fite. From the psychological laboratory of the university of chicago: The monaural localization of sound. In *Psychological Review*, volume 8, p. 225. The Macmillan Company, 1901.

M. Aretz. *Combined Wave And Ray Based Room Acoustic Simulations Of Small Rooms.* PhD thesis, RWTH Aachen University, 2012.

B.J. Arthur. *Neural computations leading to space-specific auditory responses in the barn owl.* PhD thesis, California Institute of Technology, 2002.

M. Barron. The subjective effects of first reflections in concert hallsthe need for lateral reflections. In *Journal of sound and vibration*, volume 15, pp. 475–494. Elsevier, 1971.

D.W. Batteau. The role of the pinna in human localization. In *Proceedings of the Royal Society of London. Series B. Biological Sciences*, volume 168, pp. 158–180. The Royal Society, 1967.

S. Bech. Spatial aspects of reproduced sound in small rooms. In *The Journal of the Acoustical Society of America*, volume 103, p. 434, 1998.

M.J. Beeson and D.T. Murphy. Roomweaver: A digital waveguide mesh based room acoustics research tool. In *Int. Conf. DAFx04*, pp. 268–273, 2004.

J.P. Berenger. A perfectly matched layer for the absorption of electromagnetic waves. In *Journal of computational physics*, volume 114, pp. 185–200. Elsevier, 1994.

LR Bernstein and C. Trahiotis. The normalized correlation: accounting for binaural detection across center frequency. In *The Journal of the Acoustical Society of America*, volume 100, p. 3774, 1996.

L.R. Bernstein, S. van de Par, and C. Trahiotis. The normalized interaural correlation: Accounting for nosπ thresholds obtained with gaussian and low-noise masking noise. In *The Journal of the Acoustical Society of America*, volume 106, p. 870, 1999.

S. Bilbao. *Wave and scattering methods for numerical simulation.* Wiley, 2004.

S. Bilbao. *Numerical Sound Synthesis: Finite Difference Schemes and Simulation in Musical Acoustics.* Wiley, 1 edition, December 2009. ISBN 0470510463.

S. Bilbao. Optimized FDTD schemes for 3D acoustic wave propagation. In *IEEE Transactions on Audio, Speech, and Language Processing*, number 99, pp. 1–1. IEEE, 2012.

J. Blauert. Sound localization in the median plane. In *Acustica*, volume 22, pp. 205–213, 1970.

J. Blauert. *Spatial Hearing (revised edition)*. Cambridge, MA: Massachusetts Institute of Technology, 1997.

J. Blauert. Conceptual aspects regarding the qualification of spaces for aural performances. In *Acta Acustica united with Acustica*, volume 99, pp. 1–13. S. Hirzel Verlag, 2013.

J. Blauert and J. Braasch. Binaural signal processing. In *IEEE 17th International Conference on Digital Signal Processing (DSP), 2011*, pp. 1–11. IEEE, 2011.

J. Blauert, J. Braasch, J. Buchholz, H.S. Colburn, U. Jekosch, A. Kohlrausch, J. Mourjopoulos, V. Pulkki, and A. Raake. Aural assessment by means of binaural algorithms–the AABBA project. In *Proc. Int. Symp. Auditory Audiolog. Res.- ISAAR*, volume 9, 2009.

I. Bork. A comparison of room simulation software-the 2nd round robin on room acoustical computer simulation. In *Acta Acustica u/w Acustica*, volume 86, pp. 943–956. S. Hirzel Verlag, 2000.

D. Botteldooren. Finite-difference time-domain simulation of low-frequency room acoustic problems. In *The Journal of the Acoustical Society of America*, volume 98, p. 3302, 1995.

J. Botts, A. Bockman, and N. Xiang. On the selection and implementation of sources for finite-difference methods. In *Proceedings of 20th International Congress on Acoustics*, 2010.

J. Braasch. A binaural model to predict position and extension of spatial images created with standard sound recording techniques. In *Audio Engineering Society Convention 119*, 10 2005a.

J. Braasch. *Communication Acoustics*, chapter 4 Modelling of Binaural Hearing, p. 75. Springer, 2005b.

J. Breebaart, S. Van De Par, and A. Kohlrausch. Binaural processing model based on contralateral inhibition. i. model structure. In *The Journal of the Acoustical Society of America*, volume 110, p. 1074, 2001.

C.P. Brown and R.O. Duda. A structural model for binaural sound synthesis. In *Speech and Audio Processing, IEEE Transactions on*, volume 6, pp. 476–488. IEEE, 1998.

D.S. Brungart and W.M. Rabinowitz. Auditory localization of nearby sources. head-related transfer functions. In *The Journal of the Acoustical Society of America*, volume 106, p. 1465, 1999.

D.N. Buechler, D.H. Roper, C.H. Durney, and D.A. Christensen. Modeling sources in the FDTD formulation and their use in quantifying source and boundary condition errors. In *Microwave Theory and Techniques, IEEE Transactions on*, volume 43, pp. 810–814. IEEE, 1995.

L. Chittka and A. Brockmann. Perception space the final frontier. In *PLoS Biology*, volume 3, p. e137, 04 2005.

L.W. Christopher and J.B. Schneider. An acoustic finite-difference time-domain algorithm with isotropic dispersion. In *Journal of Computational Acoustics*, volume 13, pp. 365–384. World Scientific, 2005.

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

R.K. Clifton. Breakdown of echo suppression in the precedence effect. In *The Journal of the Acoustical Society of America*, volume 82, p. 1834, 1987.

R.K. Clifton, R.L. Freyman, R.Y. Litovsky, and D. McCall. Listeners expectations about echoes can raise or lower echo threshold. In *The Journal of the Acoustical Society of America*, volume 95, p. 1525, 1994.

R.E. Collecchia, M.A. Kolar, and J.S. Abel. A computational acoustic model of the coupled interior architecture of ancient chavín. In *Audio Eng. Soc. Convention 133*, 2012.

R. Courant. Variational methods for the solution of problems of equilibrium and vibrations. In *Bull. Amer. Math. Soc*, volume 49, p. 23, 1943.

K. de Boer. Plastische klangwiedergabe. In *Philips techn. Rdsch*, volume 5, pp. 107–115, 1940.

K. de Boer. The formation of stereophonic images. In *Philips Tech. Rev*, volume 8, pp. 51–56, 1946.

S. Devore, A. Ihlefeld, K. Hancock, B. Shinn-Cunningham, and B. Delgutte. Accurate sound localization in reverberant environments is mediated by robust encoding of spatial cues in the auditory midbrain. In *Neuron*, volume 62, pp. 123–134. Elsevier, 2009.

S. Dey and R. Mittra. A locally conformal finite-difference time-domain (FDTD) algorithm for modeling three-dimensional perfectly conducting objects. In *Microwave and Guided Wave Letters, IEEE*, volume 7, pp. 273–275. IEEE, 1997.

T. Djelani and J. Blauert. Investigations into the build-up and breakdown of the precedence effect. In *Acta Acustica united with Acustica*, volume 87, pp. 253–261. S. Hirzel Verlag, 2001.

I. Drumm. A hybrid finite element/finite difference time domain technique for modelling the acoustics of surfaces within a medium. In *Acta acustica united with acustica*, volume 93, pp. 804–809. S. Hirzel Verlag, 2007.

I. Drumm, J. Hirst, and R. Oldfield. A finite difference time domain approach to analysing room effects on wave field synthesis reproduction. In *Audio Engineering Society Convention 124*, 5 2008.

IA Drumm and YW Lam. Development and assessment of a finite difference time domain room acoustic prediction model that uses hall data in popular formats. In *INCE Conference Proceedings*, volume 209, pp. 211–220, 2007.

N.I. Durlach. Equalization and cancellation theory of binaural masking-level differences. In *The Journal of the Acoustical Society of America*, volume 35, pp. 1206–1218. Acoustical Society of America, 1963.

B. Enquist and A. Majda. Absorbing boundary conditions for the numerical evaluation of waves. In *Mathematics of Computation*, volume 31(139), pp. 629–651, 1977.

J. Escolano, J.J. Lopez, and B. Pueo. Directive sources in acoustic discrete-time domain simulations based on directivity diagrams. In *The Journal of the Acoustical Society of America*, volume 121, pp. EL256–EL262. Acoustical Society of America, 2007.

J. Escolano, J.J. Lopez, and B. Pueo. Broadband directive sources for acoustic discrete-time simulations. In *The Journal of the Acoustical Society of America*, volume 126, p. 2856, 2009.

C. Faller and J. Merimaa. Binaural cue selection toolbox, 2004a. URL `http://www.acoustics.hut.fi/software/cueselection/`.

C. Faller and J. Merimaa. Source localization in complex listening situations: Selection of binaural cues based on interaural coherence. In *The Journal of the Acoustical Society of America*, volume 116, p. 3075, 2004b.

H. Farag, J. Blauert, and O. Abdel-Alim. Psychoacoustic investigations on sound-source occlusion. In *Journal of the Audio Engineering Society*, volume 51, pp. 635–646. Audio Engineering Society, 2003.

B.M. Fazenda and J. Sheaffer. Listening to Stonehenge - from measurement to auralisation. In *Submitted to Proceedings of the 42nd International Congress and Exposition on Noise Control Engineering (Inter-Noise)*, 2013.

J. Field, D. Muir, R. Pilon, M. Sinclair, and P. Dodwell. Infants' orientation to lateral sounds from birth to three months. In *Child Development*, pp. 295–298. JSTOR, 1980.

B.J. Fischer. Optimal models of sound localization by barn owls. In *Advances in Neural Information Processing Systems*, volume 20. Citeseer, 2003.

H.G. Fisher and S.J. Freedman. The role of the pinna in auditory localization. In *Journal of Auditory research*. CW Shilling Auditory Research Center, Inc., 1968.

F. Fontana and D. Rocchesso. Physical modeling of membranes for percussion instruments. In *Acta Acustica united with Acustica*, volume 84, pp. 529–542. S. Hirzel Verlag, 1998.

N.V. Franssen. *Some considerations on the mechanism of directional hearing*. PhD thesis, Technische Hogeschool Delft, 1960.

N.V. Franssen. *Stereophony*. Philips' Gloeilampenfabrieken, Distributors Cleaver-Hume Press, London, 1964.

R.L. Freyman, R.K. Clifton, and R.Y. Litovsky. Dynamic processes in the precedence effect. In *The Journal of the Acoustical Society of America*, volume 90, p. 874, 1991.

W. Gaik. Combined evaluation of interaural time and intensity differences: Psychoacoustic results and computer modeling. In *The Journal of the Acoustical Society of America*, volume 94, p. 98, 1993.

B. Gardner, K. Martin, et al. HRTF measurements of a kemar dummy-head microphone. Vision and Modeling Group, Media Laboratory, Massachusetts Institute of Technology, 1994.

H. Gaskell. The precedence effect. In *Hearing research*, volume 12, pp. 277–303. Elsevier, 1983.

E.R. Geddes. *An analysis of the low frequency sound field in non-rectangular enclosures using the finite element method*. PhD thesis, Penn State University, 1984.

S.D. Gedney. *Introduction to the Finite-difference Time-domain (FDTD) Method for Electromagnetics*, volume 6. Morgan & Claypool Publishers, 2011.

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

C. Giguère and P.C Woodland. A computational model of the auditory periphery for speech and hearing research. i. ascending path. In *The Journal of the Acoustical Society of America*, volume 95, p. 331, 1994.

R.R. Glasberg and B.C.J. Moore. Derivation of auditory filter shapes from notched-noise data. In *Hearing research*, volume 47, pp. 103–138. Elsevier, 1990.

D.F.M. Goodman and R. Brette. Spike-timing-based computation in sound localization. In *PLoS computational biology*, volume 6, p. e1000993. Public Library of Science, 2010.

M.J. Goupell, G. Yu, and R.Y. Litovsky. The effect of an additional reflection in a precedence effect experiment. In *The Journal of the Acoustical Society of America*, volume 131, p. 2958, 2012.

W.L. Gulick. Hearing: Physiology and psychophysics. Oxford U. Press, 1971.

H. Haas. Uber den einfluss eines einfachechos auf die horsamkeit von sprache. In *Acustica*, volume 1, pp. 49–58, 1951.

H. Hacihabiboglu, B. Gunel, and A.M. Kondoz. Time-domain simulation of directive sources in 3-D digital waveguide mesh-based acoustical models. In *IEEE Transactions on Audio, Speech, and Language Processing*, volume 16, pp. 934–946. IEEE, 2008.

J.A. Hargreaves and T.J. Cox. A transient boundary element method model of schroeder diffuser scattering using well mouth impedance. In *The Journal of the Acoustical Society of America*, volume 124, p. 2942, 2008.

William M Hartmann. Localization of sound in rooms. In *The Journal of the Acoustical Society of America*, volume 74, p. 1380, 1983.

W.M. Hartmann. Listening in a room and the precedence effect. In *Binaural and spatial hearing in real and virtual environments*, pp. 191–210. New York: Erlbaum, 1997.

W.M. Hartmann and B. Rakerd. Localization of sound in rooms iv: The franssen effect. In *The Journal of the Acoustical Society of America*, volume 86, p. 1366, 1989.

O. Herrmann. On the approximation problem in nonrecursive digital filter design. In *IEEE Transactions on Circuit Theory*, volume 18, pp. 411–413. IEEE, 1971.

R.C. Heyser. Determination of loudspeaker signal arrival times: Part 1. In *Journal of the Audio Engineering Society*, volume 19, pp. 734–743, 1971. URL `http://www.aes.org/e-lib/browse.cfm?elib=2136`.

F.S. Hickson and V.E. Newton. Sound localization. In *The Journal of Laryngology & Otology*, volume 95, pp. 29–40. Cambridge Univ Press, 1981.

A.J. Hill and M.O.J. Hawksford. Visualization and analysis tools for low-frequency propagation in a generalized 3D acoustic space. In *Journal of the Audio Engineering Society*, volume 59, pp. 321–337. Audio Engineering Society, 2011.

P.M. Hofman and A.J. Van Opstal. Spectro-temporal factors in two-dimensional human sound localization. In *The Journal of the Acoustical Society of America*, volume 103, pp. 2634–2648. Acoustical Society of America, 1998.

A. Ihlefeld and B.G. Shinn-Cunningham. Effect of source spectrum on sound localization in an everyday reverberant room. In *The Journal of the Acoustical Society of America*, volume 130, p. 324, 2011.

F. Ihlenburg. *Finite element analysis of acoustic scattering*, volume 132. Springer, 1998.

L.A. Jeffress. A place theory of sound localization. In *Journal of comparative and physiological psychology*, volume 41, p. 35. American Psychological Association, 1948.

H. Jeong and Y.W. Lam. Source implementation to eliminate low-frequency artifacts in finite difference time domain room acoustic simulation. In *The Journal of the Acoustical Society of America*, volume 131, pp. 258–268. Acoustical Society of America, 2012.

P.X. Joris, P.H. Smith, and T.C.T. Yin. Coincidence detection minireview in the auditory system: 50 years after Jeffress. In *Neuron*, volume 21, pp. 1235–1238, 1998.

Y. Kahana and P.A. Nelson. Boundary element simulations of the transfer function of human heads and baffled pinnae using accurate geometric models. In *J. Sound Vib.*, volume 300, pp. 552–579. Elsevier, 2007.

M. Karjalainen and C. Erkut. Digital waveguides versus finite difference structures: Equivalence and mixed modeling. In *EURASIP Journal on Applied Signal Processing*, volume 2004, pp. 978–989. Hindawi Publishing Corp., 2004.

A. Kelloniemi. Improved adjustable boundary condition for the 3-d digital waveguide mesh. In *Applications of Signal Processing to Audio and Acoustics, 2005. IEEE Workshop on*, pp. 191–194. IEEE, 2005.

J.L. Kelly and C.C. Lochbaum. Speech synthesis. In *Proc. of 4th International Congress on Acoustics*, pp. 1–4, 1962.

IR Khan and R. Ohba. Explicit formulas for coefficients of maximally flat FIR low/highpass digital filters. In *Electronics Letters*, volume 36, pp. 1918–1919. IET, 2000.

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

K. Kowalczyk and M. van Walstijn. Formulation of locally reacting surfaces in FDTD/K-DWM modelling of acoustic spaces. In *Acta Acustica united with Acustica*, volume 94, pp. 891–906. S. Hirzel Verlag, 2008.

K. Kowalczyk and M. van Walstijn. Wideband and isotropic room acoustics simulation using 2-d interpolated FDTD schemes. In *IEEE Transactions on Audio, Speech, and Language Processing*, volume 18, pp. 78–89. IEEE, 2010.

K. Kowalczyk and M. van Walstijn. Room acoustics simulation using 3-D compact explicit FDTD schemes. In *IEEE Transactions on Audio, Speech, and Language Processing*, volume 19, pp. 34–46. IEEE, 2011.

K. Kowalczyk, M. van Walstijn, and D. Murphy. A phase grating approach to modeling surface diffusion in fdtd room acoustics simulations. In *IEEE Transactions on Audio, Speech, and Language Processing*, volume 19, pp. 528–537. IEEE, 2011.

Asbjørn Krokstad, S Strom, and Svein Sørsdal. Calculating the acoustical room response by the use of a ray tracing technique. In *Journal of Sound and Vibration*, volume 8, pp. 118–125. Elsevier, 1968.

G.F. Kuhn. Model for the interaural time differences in the azimuthal. In *Journal of the Acoustical Society of America*, volume 62, 1977.

A. Kulowski. Algorithmic representation of the ray tracing technique. In *Applied Acoustics*, volume 18, pp. 449–469. Elsevier, 1985.

H. Kuttruff. *Room acoustics.* Taylor & Francis, 2000.

H. Kuttruff. *Acoustics.* Routledge, 2006.

DM Leakey and E.C. Cherry. Influence of noise upon the equivalence of intensity differences and small time delays in two-loudspeaker systems. In *The Journal of the Acoustical Society of America*, volume 29, pp. 284–286. Acoustical Society of America, 1957.

W. Lindemann. Extension of a binaural cross-correlation model by contralateral inhibition. I. simulation of lateralization for stationary signals. In *The Journal of the Acoustical Society of America*, volume 80, p. 1608, 1986a.

W. Lindemann. Extension of a binaural cross-correlation model by contralateral inhibition II. the law of the first wave front. In *The Journal of the Acoustical Society of America*, volume 80, pp. 1623–1630, 1986b.

Ruth Y Litovsky, H Steven Colburn, William A Yost, and Sandra J Guzman. The precedence effect. In *The Journal of the Acoustical Society of America*, volume 106, pp. 1633–1654. Acoustical Society of America, 1999.

R.Y. Litovsky and S.P. Godar. Difference in precedence effect between children and adults signifies development of sound localization abilities in complex listening tasks. In *The Journal of the Acoustical Society of America*, volume 128, p. 1979. Acoustical Society of America, 2010.

Q.H. Liu. The pseudospectral time-domain (pstd) algorithm for acoustic waves in absorptive media. In *Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on*, volume 45, pp. 1044–1055. IEEE, 1998.

T. Lokki, A. Southern, and L. Savioja. Studies on seat-dip effect with 3D FDTD modeling. In *Proc. Forum Acusticum, Aalborg, Denmark, June 27 - July 1*, pp. 1517–1522, 2011.

Tapio Lokki, Alex Southern, Samuel Siltanen, and Lauri Savioja. Acoustics of epidaurus studies with room acoustics modelling methods. In *Acta Acustica united with Acustica*, volume 99, pp. 40–47. S. Hirzel Verlag, 2013.

J. J. Lopez, J. Escolano, and B. Pueo. Simulation of complex and large rooms using a digital waveguide mesh. In *Audio Engineering Society Convention 123*, 10 2007.

J. LoVetri, D. Mardare, and G. Soulodre. Modeling of the seat dip effect using the finite-difference time-domain method. In *The Journal of the Acoustical Society of America*, volume 100, p. 2204, 1996.

J.A MacDonald. A localization algorithm based on head-related transfer functions. In *The Journal of the Acoustical Society of America*, volume 123, p. 4290, 2008.

E.A. Macpherson. A computer model of binaural localization for stereo imaging measurement. In *Journal of the Audio Engineering Society*, volume 39, pp. 604–622. Audio Engineering Society, 1991.

E.A. Macpherson and J.C. Middlebrooks. Listener weighting of cues for lateral angle: the duplex theory of sound localization revisited. In *The Journal of the Acoustical Society of America*, volume 111, p. 2219, 2002.

J.C. Makous and J.C. Middlebrooks. Two-dimensional sound localization by human listeners. In *The Journal of the Acoustical Society of America*, volume 87, pp. 2188–2200, 1990.

Ray Meddis and Enrique A Lopez-Poveda. *Computational Models of the Auditory System*, chapter Auditory Periphery: From Pinna to Auditory Nerve, pp. 7–38. Springer, 2010.

R. Mehra, N. Raghuvanshi, L. Savioja, M.C. Lin, and D. Manocha. An efficient GPU-based time domain solver for the acoustic wave equation. In *Applied Acoustics*, volume 73, pp. 83–94. Elsevier, 2012.

E Meyer and GR Schodder. Uber den einfluss von schallrückwürfen auf richtungslohalisation und lautastärke bei sprache. nach. In *Akad. Wiss. Gottingen, Math. Phys. Klasse IIa*, volume 6, pp. 31–42, 1952.

D. Michéa and D. Komatitsch. Accelerating a three-dimensional finite-difference wave propagation code using gpu graphics cards. In *Geophysical Journal International*, volume 182, pp. 389–402. Wiley Online Library, 2010.

P. Mokhtari, H. Takemoto, R. Nishimura, and H. Kato. Computer simulation of hrtfs for personalization of 3d audio. In *2nd Intl. Symp. on Universal Communication*, pp. 435–440. IEEE, 2008.

B.C.J. Moore. *An introduction to the psychology of hearing*, volume 4. Academic press London, 1982.

B.C.J. Moore and B.R. Glasberg. The role of frequency selectivity in the perception of loudness, pitch and time. In *Frequency selectivity in hearing*, pp. 251–308. Academic Press London, 1986.

P.M.C. Morse. *Vibration and sound*, volume 2. McGraw-Hill New York, 1948.

P.M.C. Morse and K.U. Ingard. *Theoretical acoustics*. Princeton Univ Pr, 1986.

D. Muir and J. Field. Newborn infants orient to sounds. In *Child Development*, pp. 431–436. JSTOR, 1979.

D.T. Murphy and M. Beeson. The KW-boundary hybrid digital waveguide mesh for room acoustics applications. In *IEEE Transactions on Audio, Speech, and Language Processing*, volume 15, pp. 552–564. IEEE, 2007.

D.T. Murphy and M.J. Beeson. Modelling spatial sound occlusion and diffraction effects using the digital waveguide mesh. In *Proc. AES 24th International Conference, Multichannel Audio*, pp. 26–28, 2003.

D.T. Murphy and J. Mullen. Digital waveguide mesh modelling of room acoustics: improved anechoic boundaries. In *Proc. DAFX*, volume 2, pp. 163–168, 2002.

Nvidia. Cuda programming guide. Online, 2008a. URL `http://docs.nvidia.com/cuda/cuda-c-programming-guide/index.html`.

Nvidia. Nvidia visual profiler. Online, 2008b. URL `https://developer.nvidia.com/nvidia-visual-profiler`.

SE Olive and FE Toole. The detection of reflections in typical rooms. In *Journal of the Audio Engineering Society*, volume 37, pp. 539–553. Audio Engineering Society, 1989.

V.E. Ostashev, D.K. Wilson, L. Liu, D.F. Aldridge, N.P Symons, and D. Marlin. Equations for finite-difference, time-domain simulation of sound propagation in moving inhomogeneous media and numerical implementation. In *The Journal of the Acoustical Society of America*, volume 117, p. 503, 2005.

M. Park, P.A. Nelson, and K. Kang. A model of sound localisation applied to the evaluation of systems for stereophony. In *Acta Acustica united with Acustica*, volume 94, pp. 825–839. S. Hirzel Verlag, 2008.

S. Patil and B. Ravi. Voxel-based representation, display and thickness analysis of intricate shapes. In *Computer Aided Design and Computer Graphics, 2005. Ninth International Conference on*, pp. 6–pp. IEEE, 2005.

R.D. Patterson, M.H. Allerhand, and C. Giguère. Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform. In *The Journal of the Acoustical Society of America*, volume 98, pp. 1890–1894, 1995.

J.L. Peña and M. Konishi. From postsynaptic potentials to spikes in the genesis of auditory spatial receptive fields. In *The Journal of neuroscience*, volume 22, pp. 5652–5658. Soc Neuroscience, 2002.

DR Perrott, K. Saberi, et al. Minimum audible angle thresholds for sources varying in both elevation and azimuth. In *The Journal of the Acoustical Society of America*, volume 87, p. 1728, 1990.

A. Pietrzyk and M. Kleiner. The application of the finite-element method to the prediction of soundfields of small rooms at low frequencies. In *Audio Engineering Society Convention 102*, 3 1997.

ME Potter, M. Lamoureux, and MD Nauta. An FDTD scheme on a face-centered-cubic (fcc) grid for the solution of the wave equation. In *Journal of Computational Physics*, volume 230, pp. 6169–6183. Elsevier, 2011.

V. Pulkki, M. Karjalainen, and J. Huopaniemi. Analyzing virtual sound source attributes using a binaural auditory model. In *Journal of the Audio Engineering Society*, volume 47, pp. 203–217. Audio Engineering Society, 1999.

J. Raatgever. *On the binaural processing of stimuli with different interaural phase relations.* PhD thesis, Technische Hogeschool Delft The Netherlands, 1980.

B. Rakerd and WM Hartmann. Localization of sound in rooms, II: The effects of a single reflecting surface. In *The Journal of the Acoustical Society of America*, volume 78, p. 524, 1985.

B. Rakerd and WM Hartmann. Localization of sound in rooms, III: Onset and duration effects. In *The Journal of the Acoustical Society of America*, volume 80, p. 1695, 1986.

B. Rakerd and W.M. Hartmann. Localization of sound in rooms. v. binaural coherence and human sensitivity to interaural time differences in noise. In *The Journal of the Acoustical Society of America*, volume 128, pp. 3052–3063. Acoustical Society of America, 2010.

M. Raspaud, H. Viste, and G. Evangelista. Binaural source localization by joint estimation of ILD and ITD. In *Audio, Speech, and Language Processing, IEEE Transactions on*, volume 18, pp. 68–77. IEEE, 2010.

B.J.W.S. Rayleigh. *The theory of sound*, volume 2. Macmillan, 1896.

J. Redondo, R. Picó, B. Roig, and MR Avis. Time domain simulation of sound diffusers using finite-difference schemes. In *Acta acustica united with acustica*, volume 93, pp. 611–622. S. Hirzel Verlag, 2007.

M.C. Reed and J.J. Blum. A model for the computation and encoding of azimuthal information by the lateral superior olive. In *The Journal of the Acoustical Society of America*, volume 88, p. 1442, 1990.

N. Ricker. Wavelet contraction, wavelet expansion, and the control of seismic resolution. In *Geophysics*, volume 18, pp. 769–792. Society of Exploration Geophysicists, 1953.

S. Sakamoto, H. Nagatomo, A. Ushiyama, and H. Tachibana. Calculation of impulse responses and acoustic parameters in a hall by the finite-difference time-domain method. In *Acoustical science and technology*, volume 29, pp. 256–265. J-STAGE, 2008.

S. Sakamoto, T. Seimiya, and H. Tachibana. Visualization of sound reflection and diffraction using finite difference time domain method. In *Acoustical Science and Technology*, volume 23, pp. 34–39. J-STAGE, 2002.

L. Savioja. Real-time 3D finite-difference time-domain simulation of low-and mid-frequency room acoustics. In *13th Int. Conf on Digital Audio Effects*, volume 1, p. 75, 2010.

L. Savioja and V. Valimaki. Reducing the dispersion error in the digital waveguide mesh using interpolation and frequency-warping techniques. In *IEEE Transactions on Speech and Audio Processing*, volume 8, pp. 184–194. IEEE, 2000.

J.B. Schneider, C.L. Wagner, and S.L. Broschat. Implementation of transparent sources embedded in acoustic finite-difference time-domain grids. In *The Journal of the Acoustical Society of America*, volume 103, p. 136, 1998a.

J.B. Schneider, C.L. Wagner, and O.M. Ramahi. Implementation of transparent sources in FDTD simulations. In *IEEE Transactions on Antennas and Propagation*, volume 46, pp. 1159–1168. IEEE, 1998b.

J. Schnupp, I. Nelken, and A. King. *Auditory neuroscience: Making sense of sound*. MIT Press, 2010.

HP Seraphim. Über die wahrnehmbarkeit mehrerer rückwürfe von sprachschall. In *Acustica*, volume 11, p. 80. S. Hirzel., 1961.

J. Sheaffer and B.M. Fazenda. FDTD/K-DWM simulation of 3D room acoustics on general purpose graphics hardware using compute unified device architecture (CUDA). In *Proceedings of the Institute of Acoustics*, volume 32. Institute of Acoustics, 2010.

J. Sheaffer and B.M. Fazenda. An integrated model of sound localisation in rooms (A). In *Acoustics and Audio Conference 2012*, Ben Gurion University, September 2012. IAA.

J. Sheaffer, B.M. Fazenda, and J.A.S. Angus. Computational modelling techniques for small room acoustics (A). In *Proceedings of the 1st CSE Doctoral School Research Conference*, University of Salford, November 2010. University of Salford.

J. Sheaffer, B.M. Fazenda, D.T. Murphy, and J.A.S. Angus. A simple multiband approach for solving frequency dependent problems in numerical time domain methods. In *Proceedings of Forum Acusticum*, pp. 269–274. S. Hirzel, 2011.

J. Sheaffer, M. van Walstijn, and B.M. Fazenda. A physically-constrained source model for FDTD acoustic simulation. In *Proceedings of the 15th International Conference on Digital Audio Effects (DAFx-12)*, 2012.

J. Sheaffer, C. Webb, and B.M. Fazenda. Modelling binaural receivers in finite difference simulation of room acoustics. In *Accepted to Proceedings of the 21st International Congress on Acoustics (ICA)*, 2013.

B.G. Shinn-Cunningham and K. Kawakyu. Neural representation of source direction in reverberant space. In *Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop on.*, pp. 79–82. IEEE, 2003.

B.G. Shinn-Cunningham, N. Kopco, and T.J. Martin. Localizing nearby sound sources in a classroom: Binaural room impulse responses. In *The Journal of the Acoustical Society of America*, volume 117, p. 3100, 2005.

B.G. Shinn-Cunningham, S. Santarelli, and N. Kopco. Tori of confusion: Binaural localization cues for sources within reach of a listener. In *The Journal of the Acoustical Society of America*, volume 107, p. 1627, 2000.

B.G. Shinn-Cunningham, P.M. Zurek, and N.I. Durlach. Adjustment and discrimination measurements of the precedence effect. In *The Journal of the Acoustical Society of America*, volume 93, p. 2923, 1993.

R.H. Small. Closed-box loudspeaker systems-part 1: Analysis. In *Journal of the Audio Engineering Society*, volume 20, pp. 798–808, 1972.

J. Smith III. *Introduction to Digital Filters: with Audio Applications*, volume 2. Booksurge Llc, 2006.

J.O. Smith III. Physical modeling using digital waveguides. In *Computer Music Journal*, pp. 74–91. JSTOR, 1992.

J.O. Smith III. On the equivalence of the digital waveguide and finite difference time domain schemes. In *arXiv preprint physics/0407032*, 2004.

W.B. Snow. Basic principles of stereophonic sound. In *Journal of the Society of Motion Picture and Television Engineers*, volume 61, pp. 567–589. The Society of Motion Picture and Television Engineers, 1953.

A. Southern and D. Murphy. Low complexity directional sound sources for finite difference time domain room acoustic models. In *Audio Engineering Society Convention 126*, 5 2009.

A. Southern, D. Murphy, G. Campos, and P. Dias. Finite difference room acoustic modelling on a general purpose graphics processing unit. In *Audio Engineering Society Convention 128*, 5 2010.

A. Southern, D. Murphy, and L. Savioja. Spatial encoding of finite difference time domain acoustic models for auralization. In *IEEE Trans. on Audio, Speech, and Lang. Proc.*, volume PP, p. 1, 2012.

A. Southern, D. Murphy, and J. Wells. Rendering walkthrough auralisations using wave-based acoustical models. In *Proc. of the 17th European Signal Processing Conf. EUSIPCO*, volume 9. Citeseer, 2009.

C. Spa, A. Garriga, and J. Escolano. Impedance boundary conditions for pseudo-spectral time-domain methods in room acoustics. In *Applied Acoustics*, volume 71, pp. 402–410. Elsevier, 2010.

G.C. Stecker and J.C. Middlebrooks. Distributed coding of sound locations in the auditory cortex. In *Biological cybernetics*, volume 89, pp. 341–349. Springer, 2003.

R.M. Stern, A.S. Zeiberg, and C. Trahiotis. Lateralization of complex binaural stimuli: A weighted-image model. In *The Journal of the Acoustical Society of America*, volume 84, p. 156, 1988.

J. Strikwerda. *Finite difference schemes and partial differential equations*. Society for Industrial Mathematics, 2004.

T. Su, W. Yu, and R. Mittra. A new look at FDTD excitation sources. In *Microwave and optical technology letters*, volume 45, pp. 203–207. Wiley Online Library, 2005.

P. Sypek, A. Dziekonski, and M. Mrozowski. How to render FDTD computations more effective using a graphics accelerator. In *IEEE Transactions on Magnetics*, volume 45, pp. 1324–1327. IEEE, 2009.

A. Taflove and S.C. Hagness. *Computational electrodynamics*, volume 160. Artech house BostonLondon, 2000.

G. Takanen, M.; Lorho. A binaural auditory model for the evaluation of reproduced stereophonic sound. In *Audio Engineering Society Conference: 45th International Conference: Applications of Time-Frequency Processing in Audio*, 3 2012.

Sakari Tervo, Jukka Pätynen, Antti Kuusinen, and Tapio Lokki. Spatial decomposition method for room impulse responses. In *Journal of the Audio Engineering Society*, volume 61, pp. 17–28, 2013.

J.G. Tolan and J.B. Schneider. Locally conformal method for acoustic finite-difference time-domain modeling of rigid surfaces. In *The Journal of the Acoustical Society of America*, volume 114, p. 2575, 2003.

JM Valin, F. Michaud, and J. Rouat. Robust localization and tracking of simultaneous moving sound sources using beamforming and pinproceedings filtering. In *Robotics and Autonomous Systems*, volume 55, pp. 216–228. Elsevier, 2007.

J. van Dorp Schuitman, D. de Vries, and A. Lindau. Deriving content-specific measures of room acoustic perception using a binaural, nonlinear auditory model. In *The Journal of the Acoustical Society of America*, volume 133, p. 1572, 2013.

S.A. van Duyne and J.O. Smith. Physical modeling with the 2-d digital waveguide mesh. In *Proceedings of the International Computer Music Conference*, pp. 40–40. Intl. Computer Music Association, 1993.

M. van Walstijn and K. Kowalczyk. On the numerical solution of the 2d wave equation with compact FDTD schemes. In *Proc. Int. Conf. Digital Audio Effects (DAFx08)*, pp. 205–212, 2008.

J. Vanderkooy and S.P. Lipshitz. Uses and abuses of the energy-time curve. In *Journal of the Audio Engineering Society*, volume 38, pp. 819–836, 1990. URL `http://www.aes.org/e-lib/browse.cfm?elib=6007`.

M. Vorländer. Simulation of the transient and steady-state sound propagation in rooms using a new combined ray-tracing/image-source algorithm. In *The Journal of the Acoustical Society of America*, volume 86, pp. 172–178. American Institute of Physics, 1989.

M. Vorländer. *Auralization*. Springer, 2008.

H. Wallach. The role of head movements and vestibular and visual cues in sound localization. In *Journal of Experimental Psychology*, volume 27, p. 339. American Psychological Association, 1940.

C.J. Webb and S. Bilbao. Computing room acoustics with CUDA-3D FDTD schemes with boundary losses and viscosity. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 317–320. IEEE, 2011.

K. Wendt. *Das Richtungshören bei der Überlagerung zweier Schallfelder bei Intensitäts-und Laufzeitstereophonie.* PhD thesis, Rheinisch-Westfälische Technische Hochschule Aachen, 1963.

E.M Wenzel, M. Arruda, D.J Kistler, and F.L Wightman. Localization using non-individualized head-related transfer functions. In *The Journal of the Acoustical Society of America*, volume 94, p. 111, 1993.

J.R. Wright. An exact model of acoustic radiation in enclosed spaces. In *Journal of the Audio Engineering Society*, volume 43, pp. 813–820, 1995.

T. Xiao and Q.H. Liu. Finite difference computation of head-related transfer function for human hearing. In *The Journal of the Acoustical Society of America*, volume 113, p. 2434, 2003.

W. Yang and D.W. Grantham. Echo suppression and discrimination suppression aspects of the precedence effect. In *Attention, Perception, & Psychophysics*, volume 59, pp. 1108–1117. Springer, 1997.

K. Yee. Numerical solution of initial boundary value problems involving maxwell's equations in isotropic media. In *IEEE Transactions on Antennas and Propagation*, volume 14, pp. 302–307. IEEE, 1966.

T.C.T. Yin. *Integrative Functions in the Mammalian Auditory Pathway*, chapter Neural mechanisms of encoding binaural localization cues in the auditory brainstem. Springer, 2002.

W.A. Yost and D.R. Soderquist. The precedence effect: Revisited. In *The Journal of the Acoustical Society of America*, volume 76, p. 1377, 1984.

X. Yuan, D. Borup, J. Wiskin, M. Berggren, and SA Johnson. Simulation of acoustic wave propagation in dispersive media with relaxation losses by using FDTD method with PML absorbing boundary condition. In *Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on*, volume 46, pp. 14–23. IEEE, 1999.

X. Yuan, D. Borup, J.W. Wiskin, M. Berggren, R. Eidens, and SA Johnson. Formulation and validation of berenger's pml absorbing boundary for the FDTD simulation of acoustic scattering. In *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, volume 44, pp. 816–822. IEEE, 1997.

P.M. Zurek. The precedence effect and its possible role in the avoidance of interaural ambiguities. In *The Journal of the Acoustical Society of America*, volume 67, p. 952, 1980.

P.M. Zurek. The precedence effect. In *Directional hearing*, pp. 85–105. New York: Springer-Verlag, 1987.