THE PLENACOUSTIC FUNCTION AND ITS APPLICATIONS

THÈSE N^{O} 3653 (2006)

PRÉSENTÉE LE 25 OCTOBRE 2006

À LA FACULTÉ INFORMATIQUE ET COMMUNICATION

Laboratoire de communications audiovisuelles 1

SECTION DES SYSTÈMES DE COMMUNICATION

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Thibaut AJDLER

Ingénieur Civil Electrotechnique, Vrije University Brussel, Belgique
de nationalité belge

acceptée sur proposition du jury:

Prof. E. Telatar, président du jury
Prof. M. Vetterli, Dr L. Sbaiz, directeurs de thèse
Prof. J. R. Mosig, rapporteur
Prof. R. Rabenstein, rapporteur
Prof. V. Martin, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Lausanne, EPFL
2006

Contents

Abstract v

Résumé vii

Acknowledgments ix

Frequently Used Terms, Abbreviations, and Notation xi

1 Introduction 1
- 1.1 Thesis Motivation . . . 1
- 1.2 The Plenacoustic Function . . . 2
- 1.3 Thesis Outline and Contributions . . . 4

2 Background and Related Work 7
- 2.1 Introduction . . . 7
- 2.2 Wave equation . . . 8
  - 2.2.1 Spherical source . . . 8
  - 2.2.2 Circular source . . . 10
- 2.3 Impulse Responses Modeling . . . 11
  - 2.3.1 Room impulse responses . . . 11
  - 2.3.2 Electromagnetic channel impulse responses . . . 16
  - 2.3.3 Head-related transfer functions . . . 17
- 2.4 Wave Field Synthesis and Extrapolation . . . 21
  - 2.4.1 Huygens' principle . . . 21
  - 2.4.2 The Rayleigh I integral . . . 22
  - 2.4.3 Wave field synthesis . . . 24
- 2.5 Conclusion . . . 25

3 Sound Field Analysis and Interpolation Along a Line and Multidimensional Spatial Positions 27
- 3.1 Introduction . . . 27
  - 3.1.1 Related work . . . 28

3.1.2 Contributions 29
3.1.3 Outline 29
3.2 Construction of the Plenacoustic Function 30
3.2.1 Modeling the room 30
3.2.2 Space time representation 31
3.3 Spectrum of the Plenacoustic Function on a Line 31
3.3.1 Far field assumption 31
3.3.2 Spatial decay of the spectrum of the plenacoustic function 32
3.3.3 Temporal frequency decay 38
3.4 Sampling and Reconstruction 39
3.4.1 A Sampling theorem for the plenacoustic function 40
3.4.2 Plenacoustic sampling 44
3.4.3 Reconstruction by interpolation 47
3.5 Finite Length Aperture 48
3.6 Simulations and Measurements 49
3.6.1 Simulation results 49
3.6.2 Experimental results 50
3.7 Plenacoustic Function for Multidimensional Spatial Positions . . . 53
3.7.1 Plenacoustic function on a plane 53
3.7.2 Plenacoustic function in space 56
3.8 Conclusion 57

4 Sound Field Analysis Along a Circle and its Application to HRTF Interpolation 59
4.1 Introduction 59
4.1.1 Related work 60
4.1.2 Contributions 61
4.2 Sound Field Analysis Along a Circle 62
4.2.1 Angular bandwidth of the 2D sound field on a circle 62
4.2.2 Angular bandwidth of the 3D sound field on a circle 69
4.2.3 HRTF sampling and interpolation 75
4.3 HRTF Interpolation in a Temporal Envelope Domain 78
4.3.1 Computing the carrier alignment factors in subbands. . . . 81
4.3.2 Support of the aligned HRTFs 82
4.3.3 Interpolation 85
4.4 Simulations Using HRTFs Models and Measured Data 86
4.4.1 Sinc interpolation 88
4.4.2 Comparison of interpolation methods 91
4.4.3 Discussion 96
4.5 Conclusion 97

5 Dynamic Measurement of Room Impulse Responses 99
5.1 Introduction . . . 99
5.1.1 Related work . . . 100
5.1.2 Contributions . . . 101
5.1.3 Outline . . . 101
5.2 Moving Microphone Signal . . . 102
5.2.1 Doppler effect . . . 102
5.2.2 Microphone moving along a line . . . 103
5.2.3 Microphone moving along a circle . . . 105
5.3 Recorded Signal for a Moving Loudspeaker . . . 106
5.3.1 Loudspeaker moving along a line . . . 106
5.3.2 Loudspeaker moving along a circle . . . 108
5.4 Spatio-Temporal Reconstruction of Room Impulse Responses . . . 110
5.4.1 Reconstruction algorithm for a moving microphone . . . 111
5.4.2 Reconstruction algorithm for a moving loudspeaker . . . 113
5.4.3 Discussion . . . 116
5.4.4 Individualized HRTF measurement . . . 116
5.5 Experiments . . . 117
5.6 Conclusion . . . 119

6 Stochastic Modeling of Spatio-Temporal Channel Impulse Responses Based on the Wave Equation 121
6.1 Introduction . . . 121
6.1.1 Description and problem statement . . . 122
6.1.2 Possible applications . . . 126
6.1.3 Contributions . . . 126
6.1.4 Outline . . . 127
6.2 Model for the Moving Receiver . . . 127
6.2.1 AR-2 process . . . 128
6.2.2 Generalization to AR-N process . . . 130
6.3 Spatio-Temporal Channel Impulse Response . . . 131
6.3.1 Filtering of an input process with the spatio-temporal CIR 131
6.3.2 Power spectral density of the spatio-temporal CIR . . . 132
6.3.3 Trajectory smoothness influence on the butterfly spectrum characteristics . . . 134
6.3.4 Power spectral density of the signal measured by a moving receiver . . . 135
6.4 Multipath Channels . . . 138
6.4.1 Reflection from source . . . 139
6.4.2 Multipath reflections in a rectangular room . . . 140

6.5 Mixing Deterministic and Stochastic Spatio-temporal CIR 142
6.5.1 Deterministic channels 142
6.5.2 Deterministic and stochastic spatio-temporal CIR 143
6.6 Simulations and Experimental Measurements 145
6.6.1 Simulation results 145
6.6.2 Experimental results 146
6.7 Conclusion 148
6.8 Acknowledgments 149

7 Conclusion **151**
7.1 Thesis Summary 151
7.2 Future Research 153
7.2.1 Channel impulse response modeling 153
7.2.2 Impulse response measurements 154
7.2.3 HRTF interpolation 154

A Mathematical Derivations of Multidimensional PAF Spectra **157**
A.1 Derivation of the 2D-FT of the PAF on a line 157
A.2 Derivation of the 3D-FT of the PAF on a plane 158
A.3 Derivation of the 4D-FT of the PAF in space 159

B Spatial Frequency Decay of the PAF for Microphone Lines Non-Parallel to a Wall **161**

C Temporal Frequency Decay of the PAF in a Rectangular Room **167**

D Optimal Alignment of HRTF Carriers **169**

E Filtering of an Input Process by the Stochastic PAF **171**

Bibliography **173**

Abstract

This thesis is a study of the spatial evolution of the sound field.

We first present an analysis of the sound field along different geometries. In the case of the sound field studied along a line in a room, we describe a two-dimensional function characterizing the sound field along space and time. Calculating the Fourier transform of this function leads to a spectrum having a butterfly shape. The spectrum is shown to be almost bandlimited along the spatial frequency dimension, which allows the interpolation of the sound field at any position along the line when a sufficient number of microphones is present. Using this Fourier representation of the sound field, we develop a spatial sampling theorem trading off quality of reconstruction with spatial sampling frequency. The study is generalized for planes of microphones and microphones located in three dimensions. The presented theory is compared to simulations and real measurements of room impulse responses.

We describe a similar theory for circular arrays of microphones or loudspeakers. Application of this theory is presented for the study of the angular sampling of head-related transfer functions (HRTFs). As a result, we show that to reconstruct HRTFs at any possible angle in the horizontal plane, an angular spacing of 5 degrees is necessary for HRTFs sampled at 44.1 kHz. Because recording that many HRTFs is not easy, we develop interpolation techniques to achieve acceptable results for databases containing two or four times fewer HRTFs. The technique is based on the decomposition of the HRTFs in their carrier and complex envelopes.

With the Fourier representation of the sound field, it is then shown how one can correctly obtain all room impulse responses measured along a trajectory when using a moving loudspeaker or microphone. The presented method permits the reconstruction of the room impulse responses at any position along the trajectory, provided that the speed satisfies a given relation. The maximal speed is shown to be dependent on the maximal frequency emitted and the radius of the circle. This method takes into account the Doppler effect present when one element is moving in the scenario. It is then shown that the measurement of HRTFs in the horizontal plane can be achieved in less than one second.

In the last part, we model spatio-temporal channel impulse responses between a fixed source and a moving receiver. The trajectory followed by the moving element is modeled as a continuous autoregressive process. The presented model is simple and versatile. It allows the generation of random trajectories with a controlled smoothness. Application of this study can be found in the modeling of acoustic channels for acoustic echo cancellation or of time-varying multipath electromagnetic channels used in mobile wireless communications.

Keywords: Plenacoustic function, plenoptic function, room impulse responses, spatial sound field sampling.

Résumé

Cette thèse est une étude de l'évolution spatiale du champ sonore. Nous présentons d'abord l'analyse du champs sonore le long de différentes géométries. Dans le cas de l'étude du champ sonore le long d'une ligne dans une chambre, nous décrivons une fonction bidimentionnelle caractérisant le champ sonore dans le temps et l'espace. Calculer la transformée de Fourier de cette fonction mène à un spectre en forme de papillon. Le spectre apparait tre bandlimité le long de la dimension de la fréquence spatiale. Ceci permet l'interpolation du champ sonore à toute position le long de la ligne lorsqu'un nombre suffisant de microphones est présent. Utilisant toujours la représentation de Fourier du champ sonore, nous développons un théoreme d'échantillonnage spatial présentant la qualité de reconstruction en fonction de la fréquence d'échantillonnage spatial. L'étude estgénéralisée pour des microphones disposes suivant des plans ou en trois dimensions. La théorie présentée est comparée à des simulations et mesures réelles de réponses impulsionnelles de chambre.

Nous décrivons une théorie semblable pour des microphones ou haut-parleurs disposés en cercles. Une application de cette théorie est présentée pour l'étude de l'échantillonnage angulaire des "Head-Related Transfer Functions" (HRTFs). A titre de résultat nous montrons qu'afin de reconstruire les HRTFs pour tout angle du planhorizontal, un espace angulaire de 5 degrés est necessaire pour des HRTFs échantillonnés à 44.1 kHz. Parce qu'enregistrer autant de HRTFs n'est pas une tche aisée, nous développons des techniques d'interpolation en vue d'obtenir des résultatsacceptables pour des bases de données contenant deux à quatre fois moins de HRTFs. La technique est basée sur la décomposition des HRTFsdans leurs porteuses et enveloppes complexes.

On montre alors qu'avec la représentation de Fourier du champ sonore, on peut obtenir correctement toutes les réponses impulsionnelles de la chambre mesurées le long d'une trajectoire lorsqu'on utilise un microphone ou haut-parleur mobile. La méthode présentée permet la reconstruction des réponses impulsionnelles de la chambre a n'importe quelle position le long de la trajectoire pourvu que la vitesse satisfasse à une relation donnée. La vitesse maximale admissible dépend de la fréquence maximale du signal émis ainsi que du rayon du cercle considéré. Cette méthode prend en compte l'effet Doppler présent lorsqu'un élément se meut dans le scénario. Il est alors montré que la mesure des HRTFs dans le plan horizontal peut être accompli en moins d'une seconde.

Dans la dernière partie de la thèse, nous modélisons des réponses impulsionnelles spatio-temporelles de canaux entre une source fixe et un receveur mobile. La trajectoire est modélisée comme un processus autorégressif continu. Le modèle est simple et versatile. Il permet de générer une trajectoire aléatoire avec une courbure contrlée. L'application de cette étude peut tre trouvée dans la modélisation de canaux acoustiques pour la suppression d'écho acoustique ou pour des canaux électromagnétiquesvariant dans le temps utilisés dans les systèmes de communications sans fil mobiles.

Mots-clés: Fonction plenacoustique, fonction plenoptique, réponses impulsionnelles de chambre, échantillonnage spatial du champs sonore.

Acknowledgments

I first would like to thank my two thesis supervisors Professor Martin Vetterli and Dr. Luciano Sbaiz. Martin's amazing intuition and "gut feeling" have been leading my thesis from the beginning. Our meetings were always a source of inspiration and great motivation. Luciano's amazing talent to solve any possible problem was a wonderful help during this thesis work. Thank you so much for the "Mövenpick afternoons". After our office hours, I hope to keep on with our great "sushi evenings" and philosophical discussions.

I would like to thank all my colleagues at LCAV for all the great times spent together. A specific thank goes to my office mates Christof Faller for the fruitful discussions and collaborations and to Andrea Ridolfi for your latex and stochastic lessons. Thanks to Guillermo Barrenetxea and Henri Dubois-Ferriere, my office neighbours, for the great times spent together as well as to Olivier Roy and Bob Konsbruck for our interesting technical discussions.

Special thanks go to Jocelyne Plantefol for her fast and efficient responses to all organization questions I encountered.

I would also like to thank Christoph Steiner, Florian Troesch and Professor Armin Wittneben from ETHZ for their help during the measurement of UWB channel impulse responses performed in their laboratory. These measurements have then further been possible at EPFL thanks to Professor Edoardo Charbon.

Many thanks to the jury members of my thesis, Professors Vincent Martin, Rudolf Rabenstein and Juan Mosig for their very interesting comments and remarks on my thesis.

My gratitude also goes to Dr. Igor Kozintsev and Dr. Rainer Lienhart for allowing me to spend a few unforgettable months at Intel Corporation.

Last but not least, I am thankful to my family for its support. A special thanks goes to my father for reminding me some geometry he learned forty years ago. Finally, many thanks to Evelyne for her love and patience.

Frequently Used Terms, Abbreviations, and Notation

Terms and abbreviations

2D: two-dimensional;

2D-FT: two-dimensional Fourier transform;

3D: three-dimensional;

AR-N process: autoregressive process of order N;

CIR: channel impulse response;

FT: Fourier transform;

FFT: fast implementation of the Fourier transform, denoted fast Fourier transform (FFT).

Free-field: An open space with no physical objects from which sound is reflected.

HRIR: head-related impulse response, modeling transduction of sound from a source to left and right ear entrances in free-field;

HRTF: head-related transfer function, Fourier transform of the HRIR;

ICC: inter-channel coherence;

ICLD: inter-channel level difference;

ICTD: inter-channel time difference;

MSE: mean squared error;

PAF: plenacoustic function;

PSD: power spectral density;

RIR: room impulse response;

SNR: signal-to-noise ratio;

SPAF: stochastic plenacoustic function;

WFS: wave field synthesis;

Notation and variables

To study decay of functions, the following notations are used:

- $f(x) \sim g(x)$ means that $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$.
- $f(x) \in O(g(x))$ means that there exist positive constants c and k, such that $|f(x)| \leq cg(x)$, $\forall x \geq k$.

$\|\cdot\|$	Euclidean norm;
b	size of the subband when subband decomposition is used;
c	speed of wave propagation. For acoustics, $c \approx 340$ [m/s] and in electromagnetism, $c \approx 3 \cdot 10^8$ [m/s];
f	temporal frequency [Hz];
$h(x,t)$	room impulse response along space (x-axis) and time;
$h(\theta,t)$	room impulse response or HRTF along angular (θ-axis) and time;
$h(\theta)$	time of arrival of the sound from the source to the receiver;
$p(x,t)$	sound field along space (x-axis) and time;
$q(x,t)$	windowed sound field along space (x-axis) and time;
t	time [s];
v	speed [m/s];
$w(x,t)$	window function;
Δx	distance between two consecutive samples (here in the x direction);
$\delta(t)$	Dirac delta function;
$\delta_k(t)$	Kronecker symbol;
ϕ	spatial frequency [rad/m];
ϕ_S	spatial sampling frequency [rad/m];
Ω	elevation angle [degrees];
ω	temporal frequency [rad/s];
θ	azimuthal angle [degrees];

l_θ	angular frequency or wave number in the θ direction;
l_{θ_S}	angular sampling frequency or wave number in the θ direction;
τ	time related to the spatial trajectory [s];
γ	temporal frequency related to the spatial trajectory [rad/s];
Γ	random trajectory;
ν	relative speed;
σ	standard deviation of a process;
E_i	Exponential integral function;
$\mathcal{F}_x$	Fourier transform along the x direction;
$H_i^{(1,2)}$	i^{th} order Hankel function of the first or second kind;
$I(t)$	input stochastic process; function of the time;
J_i	Bessel function of order i;
K_i	modified Bessel function of the second kind of order i;
$O(\tau, t)$	Output stochastic process; function of the τ and t;
$R_O(\Delta\tau, \Delta t)$	autocorrelation of the process O;
$S_O(\gamma, \omega)$	power spectral density of the process O;

Fourier Transforms

Fourier transform (FT) The (continuous time) Fourier transform is defined for continuous signals $f(t)$ as

$$\hat{p}(\omega) = \int_{-\infty}^{\infty} p(t) e^{-j\omega t} \mathrm{d}t.$$

Its inverse is called the inverse Fourier transform (IFT), and can be written as

$$p(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{p}(\omega) e^{j\omega t} \mathrm{d}\omega.$$

Fourier series (FS) The (continuous time) Fourier series is defined for periodic, continuous signals $p(t)$ with period T as

$$\hat{p}(k) = \frac{1}{T} \int_{0}^{T} p(t) e^{-j2\pi kt/T}, \quad k = -\infty, ..., \infty.$$

Its inverse is called the inverse Fourier series (IFS):

$$p(t) = \sum_{k=-\infty}^{\infty} \hat{p}(k) e^{j2\pi kt/T}.$$

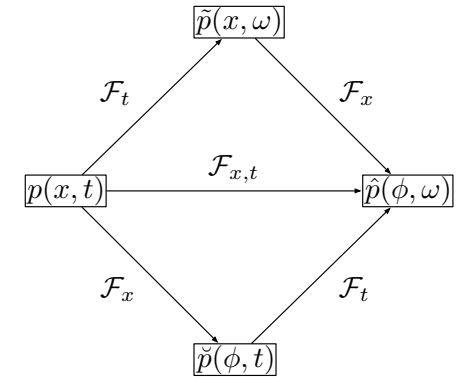
In this thesis, Fourier transforms are taken along different dimensions. It is possible to define three different Fourier transforms of $p(x, t)$. A scheme is

shown to present the different possible Fourier transforms in Fig. 1(a).

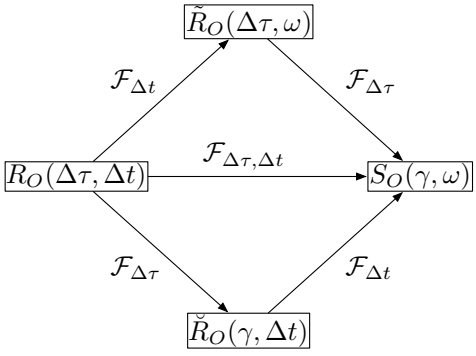
$$\begin{aligned}\tilde{p}(x,\omega) &= \mathcal{F}_t\{p(x,t)\}\\ \breve{p}(\phi,t) &= \mathcal{F}_x\{p(x,t)\}\\ \hat{p}(\phi,\omega) &= \mathcal{F}_x\{\tilde{p}(x,\omega)\} = \mathcal{F}_t\{\breve{p}(\phi,t)\} = \mathcal{F}_{x,t}\{p(x,t)\}.\end{aligned}$$

Similarly, different Fourier transforms for the autocorrelation function $R_O(\Delta\tau,\Delta t)$ can be defined as shown in Fig. 1(b).

$$\begin{aligned}\tilde{R}_O(\Delta\tau,\omega) &= \mathcal{F}_{\Delta t}\{R_O(\Delta\tau,\Delta t)\}\\ \breve{R}_O(\gamma,\Delta t) &= \mathcal{F}_{\Delta\tau}\{R_O(\Delta\tau,\Delta t)\}\\ S_O(\gamma,\omega) &= \mathcal{F}_{\Delta\tau}\{\tilde{R}_O(\Delta\tau,\omega)\} = \mathcal{F}_{\Delta t}\{\breve{R}_O(\gamma,\Delta t)\} = \mathcal{F}_{\Delta\tau,\Delta t}\{R_O(\Delta\tau,\Delta t)\}.\end{aligned}$$



(a)



(b)

Figure 1: Different Fourier transforms for (a) the two-dimensional signal $p(x,t)$; (b) the two-dimensional autocorrelation of the stochastic process $O(\tau,t)$.

Chapter 1

Introduction

1.1 Thesis Motivation

With the emergence of increasingly powerful computers and high-capacity channels, the processing of multiple audio channels has become more and more accessible. The simultaneous recording of hundreds of channels is becoming nowadays possible using a simple laptop and a multichannel sound card. These technological advances have created a large amount of new possible techniques in such fields as communications, speech processing, source separation and localization.

Edison and his tinfoil phonograph were the first to record voice in 1877, while stereo recording was patented in 1931 by Blumlein. Since then many new techniques involving stereo have been developed for the recording or playback of audio, but the emergence of multichannel recording and reproduction techniques is quite recent. Dolby introduced its 5.1 system in 1992 and since then new techniques involving more and more channels have been explored. In 2004, the largest microphone array was built at MIT using 1020 microphones. This increase in the number of channels in new audio systems represents a great news for signal processors but also new challenges due to the quantity of data to process and the demand for ever increasing performance systems.

Two of the most basic questions that one can ask when using a large number of microphones in a room are the following: "Assume you are in a concert hall, and you want to faithfully describe the sound pressure field at any location in the hall. If you record the acoustic event with an array of microphones, how many do you need to be able to reproduce the sound pressure field at any point?" or "Conversely, assume a virtual acoustical environment, where sources are moving, while the sound pressure is measured in a particular spot. How finely do you need to simulate the acoustic impulse responses to be able

to place the source at any location?"

These questions are at the basis of the work undertaken in this thesis and their answers, as well as related ones, are developed in the course of the different chapters of this work. All the elements necessary to answer these questions lie in the spatio-temporal acoustic pressure field and its properties. We have decided to call this field the *plenacoustic function* in reference to the plenoptic function introduced by Adelson and Bergen [4]. In the rest of this introduction chapter, we present in more detail the plenacoustic function (PAF) and compare it with the plenoptic function. Further, an outline of the thesis is given along with the contributions of each chapter.

1.2 The Plenacoustic Function

The name "plenacoustic function" has been mentioned for the first time in [71]. The first analysis of the function has been presented in [9,11,10]. This name has been given by reference to the plenoptic function. Remark that the choice of the names "plenacoustic" and "plenoptic" functions present mixing Greek and Latin roots. The Latin expressions would be the "plenaudio" and "plenvideo" functions while the Greek expressions would be the "panacoustic" and panoptic" functions. Since the name of the "plenoptic" function was already chosen by Adelson and Bergen [4], we followed with a similar etymological barbarism.

As a first step to define the PAF, let us introduce the plenoptic function. It is a seven dimensional function, $f(x, y, z, \theta, \alpha, \lambda, t)$ which describes the intensity of the light field seen at location (x, y, z) when looking in direction (θ, α), at wavelength λ and time t as described in Fig. 1.1(a). Remark that in this case, one can define a very precise direction (θ, α) of arrival of a light ray. This is due to the very small wavelength of light (on the order of 100 nm). Therefore, it is possible to build devices measuring a good approximation of the plenoptic function. In the acoustical case, the wavelengths are 10^6 time larger. This makes it unfeasible in practice to construct a device that would measure the sound pressure at any arbitrary position and the component of its gradient due to a source originating from a very precise direction.

Both the plenoptic function and the PAF are related to the wave equation. In the case of light, a camera is sensitive to the intensity of the radiation while the phase information is lost. On the other hand, a camera can be very directional since a lens is focalizing the image on the sensor. This explains the choice of the definition of the plenoptic function. Remark that the plenoptic function does not correspond to the complete solution of the wave equation but simply to what can be sensed by the eye or a camera. In the case of acoustic waves, the solution of the wave equation is the pressure in every point of the

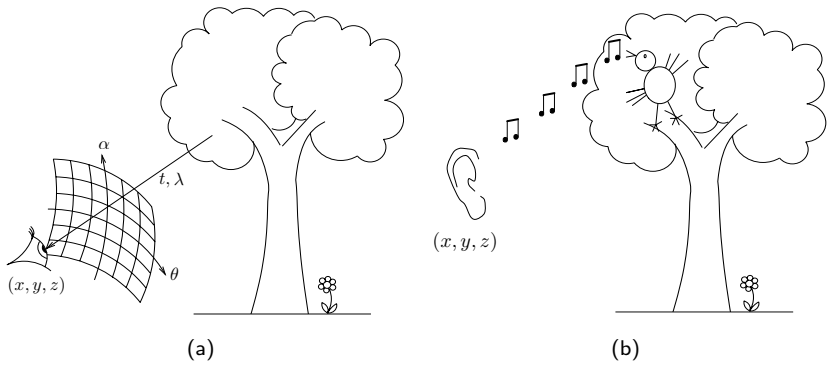


Figure 1.1: Schemes for the (a) plenoptic function; (b) plenacoustic function.

space. Since the human ear, and microphones are sensitive to this quantity, we present the acoustic analogy of the plenoptic function, i.e. the PAF, as the sound pressure field.

Taking into account these physical differences existing between the optical and the acoustical case, we define the PAF as follows.

Definition 1.2.1. The plenacoustic function denoted as $p(x, y, z, t)$ is the sound pressure recorded at location (x, y, z) and time t given one or multiple acoustic events.

This definition is illustrated in Fig. 1.1(b). Note that from the knowledge of the sound pressure in every point of the three dimensions, one can calculate variations of pressure in all directions and obtain the particle velocity vector for every point using Euler's equation [85]

$$\rho_0 \frac{\partial v_n(x, y, z, t)}{\partial t} = -\frac{\partial p(x, y, z, t)}{\partial \mathbf{n}}, \tag{1.1}$$

where v_n stands for the particle velocity in the direction $\mathbf{n}$ and ρ_0 for the medium density. Therefore, from this information, one can transform the omnidirectional PAF into a directional PAF. This is achievable since the PAF represents the complete solution of the wave equation. Remark that the velocity particle can only be derived along all directions when the PAF is known in the three spatial dimensions. In a practical case, one would measure the PAF along a line. The knowledge of the particle velocity as well as the directionality would then only be partial.

As defined previously, the PAF is the solution of the scalar acoustic wave

equation given by:

$$\nabla^2 p(x,y,z,t) - \frac{1}{c^2}\frac{\partial^2 p(x,y,z,t)}{\partial t^2} = s(x,y,z,t), \tag{1.2}$$

where c represents the speed of sound propagation and $s(x,y,z,t)$ represents a distribution of sources located in space. The wave equation and its solution will be discussed in detail in Section 2.2. Often, one is concerned with the case of a single point source. Namely, for a given source s, we denote the room impulse response (RIR) at location (x,y,z) by $h(x,y,z,t)$, and if the source generates a signal $s(t)$, the PAF is

$$p(x,y,z,t) = \int_{-\infty}^{\infty} s(\tau)h(x,y,z,t-\tau)\mathrm{d}\tau. \tag{1.3}$$

When the emitted sound is a Dirac pulse, the PAF becomes simply the spatio-temporal RIR. The PAF is then called the Green's function [85] as will be described in Section 2.2. Assuming that our system is linear and time invariant, we consider the presence of multiple sources as the superposition of single sources. The sound pressure field can then be regarded as the sum of all point sources convolved with their spatio-temporal RIR.

From the point of view of the physicist, the PAF is simply the solution of the wave equation with appropriate boundary conditions, and a given driving function. From the point of view of the numerical analyst, the system would be very complex for any reasonable room, even for very simple cases. For the signal processor, acoustic RIRs have been studied, measured and simulated for many scenarios, and it is thus natural to study the PAF globally. A natural question for a signal processor is of course the sampling question: is there a discrete set of points in time and space from which the full PAF can be reconstructed? The equivalent question for the plenoptic function was posed and solved by Chai et al. [30] with further results from Zhang et al. [115, 116]. For the time dimension, bandlimited sources are assumed to allow sampling in time. The more interesting question is of course sampling in space, which directly relates to the number of microphones necessary to acquire the PAF. This question, as well as related ones, are discussed in the different chapters of this thesis.

1.3 Thesis Outline and Contributions

In this thesis, we present the plenacoustic function as a means for studying and developing different aspects related to the sampling of the sound field. Effort was made to always keep a balance between the theoretical parts of the work and the more experimental side of the research. As a result, next to the

development of sampling theorems and the derivation of difficult integrals, we also present experimental measurements to validate each theoretical result.

Chapter 2 describes some works related to the general topic of this thesis. Therefore, as the plenacoustic function is nothing else than the solution of the wave equation, this equation is presented and discussed. Its solution is studied and the Green's function for the wave equation in two and three dimensions is explored. As the solution of the wave equation is only achievable analytically in very few cases, models are presented for its solution in different contexts. First, room impulse responses are discussed and modeled according to different methods. Further, head-related transfer functions are described in detail. These are filters describing the effect of the shape of the head, torso and pinnae on the sound arriving at the entrance of the ear. They are widely used in headphone playback systems to achieve more realism. Then, electromagnetic channel impulse responses, satisfying the same wave equation (only with a different propagation speed), are discussed and different models are presented. Finally, as an application of array processing, wave field synthesis is presented. This technique aims at reproducing the sound field as accurately as possible using loudspeaker arrays. Even if this method is not directly used further in this thesis, different concepts described in its study are mentioned in the different chapters of the thesis.

Chapter 3 studies the plenacoustic function along different geometries such as a line, a plane and the three dimensional space filled with microphones. These plenacoustic functions are represented using their multi-dimensional Fourier spectrum. This representation is very suitable to observe and describe aliasing components appearing when sampling the sound field. A contribution of this chapter is to show that the support of the spatio-temporal spectrum of the plenacoustic function along a line has a butterfly shape and that the energy decays exponentially outside of this essential support. Using this result, we present a spatial sampling theorem describing the quality of reconstruction of the sound field as a function of the spatial sampling frequency. This is obtained without making any far field assumptions as usually done in the literature. Interpolation of the sound field is then discussed and simulation results are presented together with experimental measurements of room impulse responses in the acoustic and electromagnetic case.

In *Chapter 4*, the plenacoustic function and its sampling are studied for circular arrays of microphones. A first contribution is to present an angular sampling theorem describing the quality of reconstruction of the sound field in function of the angular sampling frequency. Further, the dual problem considering one microphone and different loudspeakers located on a circle is explored. This arrangement corresponds to the typical head-related transfer

function (HRTF) measurement setup. As application of the presented angular sampling theorem, sampling and interpolation of HRTFs is discussed. As the number of HRTFs required by the angular sampling theorem is often large, another interpolation technique is developed to achieve good results when the angular spacing is undersampled by a factor of two or four. The contribution of this second part of Chapter 4 is to present a new technique of HRTFs interpolation achieving better results than previous existing methods in a mean squared sense.

An interesting application of Chapters 3 and 4 is described in *Chapter 5*. There, using a continuously moving receiver or source, a novel technique is described that reconstructs the room impulse responses at any position along the followed trajectory. The method has been implemented on a real system using a motorized microphone holder and the results confirm the presented theory. This technique appears to be very useful for the measurement of HRTFs. All HRTFs in the horizontal plane can be measured in less than one second. This fast method makes the use of individualized HRTFs for headphone playback easier to achieve since measurement of our own HRTFs becomes less time consuming.

The plenacoustic function is described in *Chapter 6* for smooth stochastic trajectories. A contribution of the chapter is to develop a versatile model allowing the creation of trajectories with controlled smoothness. This is achieved by using continuous autoregressive processes. The control on the smoothness allows the generation of a large variety of realistic trajectories such as moving persons or driving cars. The stochastic plenacoustic function describes the sound field at each of the positions along the generated trajectories. The autocorrelation function and the power spectral density of the stochastic plenacoustic function are studied and closed form solutions are given for specific cases. As the main contribution of the chapter, we show that the power spectral density of the stochastic plenacoustic function has a butterfly shape and that the opening of the butterfly shape is modified by varying the smoothness of the followed trajectory. Experimental measurements are used to confirm the theoretical results.

Finally, *Chapter 7* concludes this thesis and discusses some ideas for future work.

Chapter 2

Background and Related Work

2.1 Introduction

This background chapter aims at introducing different concepts discussed in the course of this thesis. Some tools that will be frequently used in this work are presented in detail in this background chapter.

The main topic of this thesis is the study of the sound propagation in space. The equation governing the evolution of the sound waves over space and time is known as the wave equation. The two and three dimensional (2D and 3D) wave equations are presented in Section 2.2. For each case, the solution is explored in detail.

This background chapter presents some basic tools that will be further used in the next chapters. One of the main contribution of this thesis is the description of the interpolation of room impulse responses, electromagnetic channel impulse responses and head-related transfer functions. Therefore, these different functions are introduced and models to describe them are presented in Section 2.3.

Across the different chapters of the thesis, the reader will be presented different schemes for interpolation of the sound field along different geometries such as linear arrays and circular arrays. A very interesting application of this kind of array processing can be found in wave field synthesis (WFS). This technique aims at reproducing a sound field by using loudspeaker arrays. Although this technique is not directly the topic of this thesis, a short introduction of the method is presented in Section 2.4 since related work of interest for the present thesis has been discussed in the context of WFS.

2.2 Wave equation

In this section, the wave equation is described in two and three dimensions. For each case, the Green's function is defined and derived. The Green's function is a basic block of the work of the next chapters, therefore its derivation is described in detail.

2.2.1 Spherical source

Considering a spherical source emitting a Dirac pulse at position $\mathbf{s} = (x_s, y_s, z_s)$ at instant t_s, the sound field generated by this source is given by the following equation:

$$\nabla^2 p(x,y,z,t) - \frac{1}{c^2}\frac{\partial^2 p(x,y,z,t)}{\partial t^2} = -\delta(x-x_s)\delta(y-y_s)\delta(z-z_s)\delta(t-t_s), \quad (2.1)$$

with c the speed of the wave propagation. The previous expression (2.1) can be rewritten as

$$\nabla^2 p(\mathbf{r},t) - \frac{1}{c^2}\frac{\partial^2 p(\mathbf{r},t)}{\partial t^2} = -\delta(\mathbf{r})\delta(t), \quad (2.2)$$

where for simplicity the source emits its pulse at $t_s = 0$ from position $\mathbf{s} = (0,0,0)$. Note that $\mathbf{r} = (x,y,z)$. Taking the Fourier transform with respect to time in (2.2) leads to

$$\nabla^2 \tilde{p}(\mathbf{r},\omega) + \left(\frac{\omega}{c}\right)^2 \tilde{p}(\mathbf{r},\omega) = -\delta(\mathbf{r}). \quad (2.3)$$

Since free space is assumed, $\tilde{p}(\mathbf{r},\omega)$ only depends on $r = \|\mathbf{r}\|$ due to symmetry. Using the spherical coordinates, (2.3) can be rewritten as

$$\frac{1}{r^2}\frac{\mathrm{d}}{\mathrm{d}r}\left(r^2\frac{\mathrm{d}\tilde{p}(r,\omega)}{\mathrm{d}r}\right) + \left(\frac{\omega}{c}\right)^2 \tilde{p}(r,\omega) = -\frac{\delta(r)}{4\pi r^2}. \quad (2.4)$$

Since the right-hand side in (2.4) is zero except for the origin, multiplying both sides by r and considering $r > 0$ leads to

$$\frac{\mathrm{d}^2}{\mathrm{d}r^2}\left[r\tilde{p}(r,\omega)\right] + \left(\frac{\omega}{c}\right)^2 \left[r\tilde{p}(r,\omega)\right] = 0. \quad (2.5)$$

Note that (2.5) corresponds to the one dimensional Helmholtz equation whose solution is known as the d'Alembert solution [85]

$$\tilde{p}(r,\omega) = a\frac{e^{-j\frac{\omega}{c}r}}{r} + b\frac{e^{j\frac{\omega}{c}r}}{r}, \quad (2.6)$$

with a and b being arbitrary constants. In our case, due to the presence of a source at the origin, only the wave traveling towards $+r$ is assumed, therefore the second term in (2.6) is canceled. To obtain the value of a, we substitute (2.6) into (2.2), and then integrate (2.2) within a small sphere V of radius r including the origin. The first term in (2.2) is then processed as follows using Green's theorem [85]

$$\int_V \nabla^2 \tilde{p}(\mathbf{r},\omega)\mathrm{d}V = \oint_S \nabla \tilde{p}(\mathbf{r},\omega)\cdot \mathrm{d}S = 4\pi r^2 \nabla \tilde{p}(\mathbf{r},\omega)\cdot \bar{r}, \tag{2.7}$$

where $\bar{r}$ stands for the unit vector pointing in the outward radial direction, normal to the surface of the sphere S. Therefore,

$$\nabla \tilde{p}(\mathbf{r},\omega)\cdot \bar{r} = a\frac{\partial}{\partial r}\left(\frac{e^{-j\frac{\omega}{c}r}}{r}\right). \tag{2.8}$$

Replacing (2.8) in (2.7) leads to

$$\int_V \nabla^2 \tilde{p}(\mathbf{r},\omega)\mathrm{d}V = -4\pi r^2 a\left(\frac{e^{-j\frac{\omega}{c}r}}{r^2} + j\frac{\omega}{c}\frac{e^{-j\frac{\omega}{c}r}}{r}\right). \tag{2.9}$$

The integral of the second term in (2.2) leads to

$$\begin{aligned}
\int_V \left(\frac{\omega}{c}\right)^2 \tilde{p}(\mathbf{r},\omega)\mathrm{d}V &= \left(\frac{\omega}{c}\right)^2 \int_{r=0}^{r} \tilde{p}(r,\omega)4\pi r^2 \mathrm{d}r && (2.10)\\
&= 4\pi a\left(\frac{\omega}{c}\right)^2 \int_{r=0}^{r} re^{-j\frac{\omega}{c}r}\mathrm{d}r && (2.11)\\
&= 4\pi a\left(\frac{\omega}{c}\right)^2 \left(-\frac{re^{-j\frac{\omega}{c}r}}{j\frac{\omega}{c}} + \left(\frac{c}{\omega}\right)^2\left(e^{-j\frac{\omega}{c}r}-1\right)\right), && (2.12)
\end{aligned}$$

where the last equality is obtained after integration by parts. The integral of the third term is simply unitary since the volume contains the Dirac source. Taking the limit for $r \to 0$, we obtain that

$$a = \frac{1}{4\pi}. \tag{2.13}$$

Therefore, (2.6) is now

$$\tilde{p}(r,\omega) = \frac{e^{-j\frac{\omega}{c}r}}{4\pi r}. \tag{2.14}$$

Replacing now r by the distance between the source and the receiver denoted as $\|\mathbf{r}\|$, the solution of (2.2) is

$$\tilde{p}(\mathbf{r},\omega) = \frac{e^{-j\frac{\omega}{c}\|\mathbf{r}\|}}{4\pi\|\mathbf{r}\|}. \tag{2.15}$$

If the source is not located at the origin but at position $\mathbf{s}$, consider the change of variable $r = \|\mathbf{r} - \mathbf{s}\|$ in (2.14), which leads to

$$\tilde{p}(\mathbf{r}, \omega) = \frac{e^{-j\frac{\omega}{c}\|\mathbf{r}-\mathbf{s}\|}}{4\pi\|\mathbf{r} - \mathbf{s}\|}. \tag{2.16}$$

Calculating the inverse Fourier transform of (2.16) gives

$$p(x, y, z, t) = \frac{\delta\left(t - \frac{\sqrt{(x-x_s)^2+(y-y_s)^2+(z-z_s)^2}}{c}\right)}{4\pi\sqrt{(x - x_s)^2 + (y - y_s)^2 + (z - z_s)^2}}. \tag{2.17}$$

2.2.2 Circular source

To obtain the Green's function for the 2D wave equation, two derivations are possible. The first version considers the wave equation in 2D and solves it for an excitation in the plane [51]. The other derivation followed by most of the textbooks [85, 112] makes use of the 3D Green's function to derive the 2D version as a special case. For simplicity, we choose to present the second version.

Consider the circular source to be in the plane formed by the axes x and y. The direction z is perpendicular to the plane of interest. We consider a 3D setup where the excitation is not a unique spherical source but an infinite line of spherical sources along the z direction. The wave field due to this infinite line of sources is described as

$$\tilde{p}_{2D}(x, y, \omega) = \int_{-\infty}^{\infty} \frac{e^{-j\frac{\omega}{c}\sqrt{(x-x_s)^2+(y-y_s)^2+z^2}}}{4\pi\sqrt{(x - x_s)^2 + (y - y_s)^2 + z^2}} \mathrm{d}z. \tag{2.18}$$

To solve this integral, we introduce the integral definition of the Hankel function [53]:

$$H_0^{(2)}(\frac{\omega}{c}\xi) = \frac{j}{\pi}\int_{-\infty}^{\infty} \frac{e^{-j\frac{\omega}{c}\sqrt{\xi^2+z^2}}}{\sqrt{\xi^2 + z^2}} \mathrm{d}z, \tag{2.19}$$

where $H_0^{(2)}$ stands for the order zero Hankel function of the second kind. Therefore, (2.18) is rewritten as

$$\tilde{p}_{2D}(x, y, \omega) = \frac{-j}{4} H_0^{(2)}\left(\frac{\omega}{c}\sqrt{(x - x_s)^2 + (y - y_s)^2}\right). \tag{2.20}$$

This last equation corresponds to the Green's function for the 2D wave equation.

2.3 Impulse Responses Modeling

This section describes different techniques used for the modeling of impulse responses. In Section 2.3.1, RIRs models are presented. A lot of attention is devoted to the *image source model* [13] widely used in Chapter 3. A short introduction of different models for EM channel impulse responses is given in Section 2.3.2 and further used in Chapter 6. Interpolation of head-related transfer function is described in Chapter 4. Therefore, a introduction to these functions is presented in Section 2.3.3.

2.3.1 Room impulse responses

In this section, we present a brief overview of different techniques existing for the modeling and simulation of RIRs. The sound propagation is described by means of the wave equation as presented in Section 2.2. An impulse response from a source to a listener can be obtained by solving this equation. Unfortunately, an analytical solution can only be found for very simple geometries. In most of the cases, the solution needs to be approximated and different approaches exist for the computational modeling of room acoustics. Many techniques have been presented; we have chosen three of them that we describe in more detail in the following sections:

- Image source model
- Ray-tracing methods
- Wave-based methods.

Image source model

The image source is described in this section for the simple case of a rectangular box-shaped room. The idea of the technique is to model the effect of the walls as new virtual sources located on the far side of these walls. To model the reverberation in a room, one considers a large number of free field virtual sources (theoretically an infinite number) and by adding their effects, the RIR is modeled.

Consider the dimensions of the room to be (L_x, L_y, L_z). The source s_1 is located at $(x_{s_1}, y_{s_1}, z_{s_1})$. The receiver r has coordinates (x_r, y_r, z_r). The norm of the vector $\mathbf{m}_1$ corresponds to the distance between the source and the receiver with

$$\mathbf{m}_1 = (x_{s_1} - x_r, y_{s_1} - y_r, z_{s_1} - z_r)\,. \tag{2.21}$$

As discussed in Section 2.2, the impulse response in free field due to the source s_1 is given by

$$p(s_1, r, t) = \frac{\delta\left(t - \frac{\|\mathbf{m}_1\|}{c}\right)}{4\pi\|\mathbf{m}_1\|}. \tag{2.22}$$

In the presence of one wall along the x direction as described in Fig. 2.1(a), a reflection will arrive at the receiver after the free field sound. This reflection can be modeled as one virtual source on the far side of the wall. This new virtual source is denoted as s_2 and its distance to the receiver is $\|\mathbf{m}_2\|$. The RIR is described as

$$p(s_1, r, t) = \frac{\delta\left(t - \frac{\|\mathbf{m}_1\|}{c}\right)}{4\pi\|\mathbf{m}_1\|} + \frac{\delta\left(t - \frac{\|\mathbf{m}_2\|}{c}\right)}{4\pi\|\mathbf{m}_2\|}. \tag{2.23}$$

Considering two perpendicular walls in the directions x and y, three additional virtual sources are created next to the original source as shown in Fig. 2.1(b). In the case of three perpendicular walls along the three directions x, y and z shown in Fig. 2.1(c), seven virtual sources are present next to the original source in the RIR whose expression becomes

$$p(s_1, r, t) = \sum_{i=1}^{8} \frac{\delta\left(t - \frac{\|\mathbf{m}_i\|}{c}\right)}{4\pi\|\mathbf{m}_i\|}, \tag{2.24}$$

with $\|\mathbf{m}_i\|$ the distance between source s_i and r. Note that for the eight considered sources, $\mathbf{m}_i$ corresponds to

$$\mathbf{m}_i = \left(x_{s_1} \pm x_r, y_{s_1} \pm y_r, z_{s_1} \pm z_r\right). \tag{2.25}$$

In a rectangular room, parallel to each wall, a second wall is present at some finite distance. Similarly to the effect of a person looking at himself between two mirrors, an infinite number of images are created. Each of the first eight sources are repeated with a period of $2L_x$, $2L_y$ and $2L_z$ along the x, y and z directions respectively. The general formula for RIRs in a rectangular room with perfectly reflecting walls is given by

$$p(s_1, r, t) = \sum_{i=1}^{8} \sum_{v=-\infty}^{\infty} \frac{\delta\left(t - \frac{\|\mathbf{m}_i + \mathbf{m}_v\|}{c}\right)}{4\pi\|\mathbf{m}_i + \mathbf{m}_v\|}, \tag{2.26}$$

with $\mathbf{m}_v = (2lL_x, 2nL_y, 2oL_z)$ and (l, n, o) an integer vector triplet. Note that the first sum shows that in a 3D case, seven virtual sources are created in addition to the original source due to the first reflections on three walls.

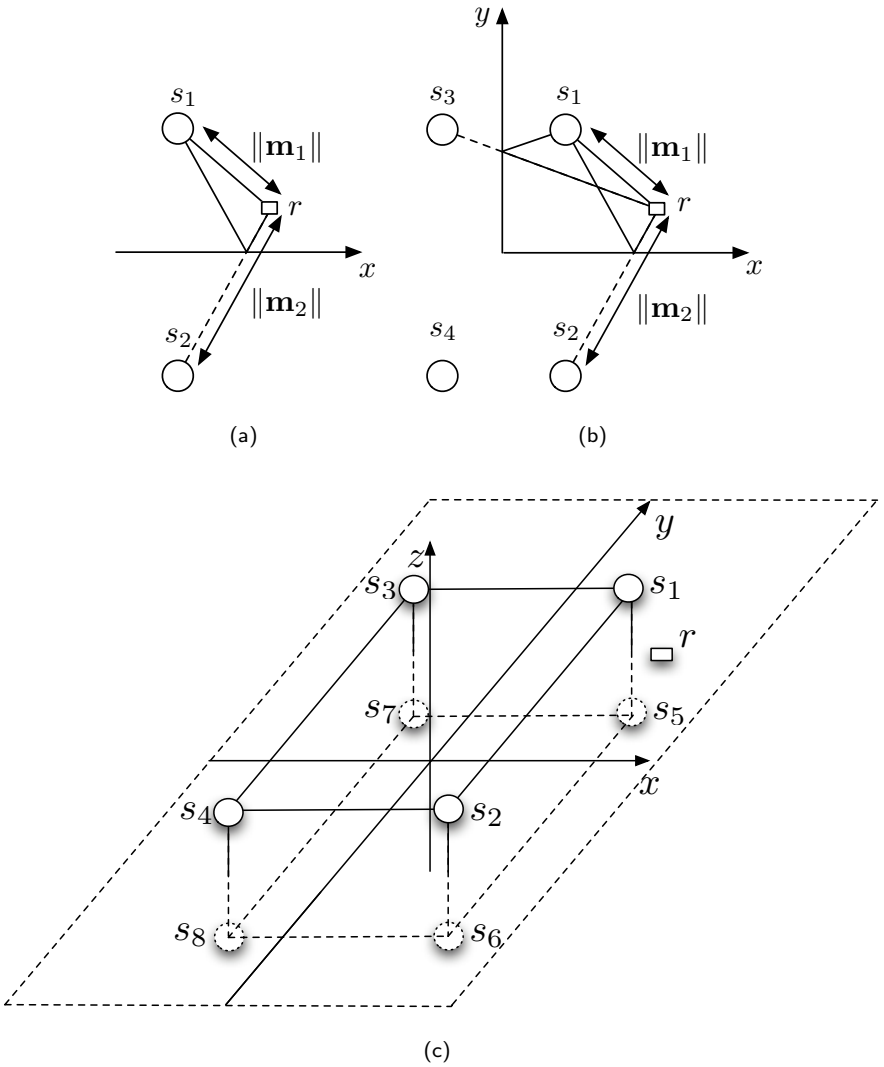


Figure 2.1: Schemes for the image source model considering (a) one wall; (b) two walls; (c) three walls.

The second sum shows that sound between two parallel rigid walls is infinitely reverberated. This construction is shown in the 2D case in Fig. 2.2. In case the walls are not totally reflective, a modified version of (2.26) is given in [13], where the reflection factors of each wall are considered.

A generalization of this technique can be found for more complex geometries [22]. The basic idea is very similar, but an additional step needs to be performed to check the visibility between the virtual source and the receiver. Some virtual sources need indeed not be considered due to occlusions. Note that the image

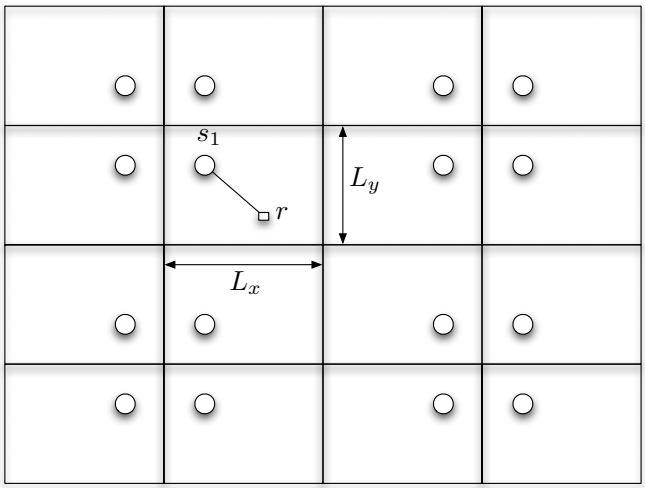


Figure 2.2: Image source model for a 2D case.

source model is computationally very intensive. The number of virtual sources increases exponentially with the duration of the impulse response. Therefore, this technique is mostly used for the simulation of early reflections while the late reverberation is modeled using other less computationally intensive techniques described in the sequel.

Ray-based methods

Ray-based methods are based on geometrical room acoustics where sound is considered to act like rays [75]. This assumption is valid when the wavelength of sound is small compared to the surfaces in the room. Phenomena taking into account the wave nature of the sound field propagation such as diffraction and interferences [85] are therefore ignored in this method.

The most commonly used ray-based method is the ray-tracing method [74]. In the basic algorithm, the sound source emits sound rays which are reflected at the different surfaces in the room according to specular reflections which assumes that the incident angle of an incoming ray is the same as the incident angle of the outgoing ray. More advanced rules can include more complex reflections or diffusions phenomena [88]. Each emitted ray is followed until some termination criteria are met (such as minimum energy of the ray or maximum propagation time).

Although this method is general and relatively easy to implement, some drawbacks exist. When considering rays emitted by the source, an angular sampling needs to be performed in order to limit the number of possible rays

and the computational time of the algorithm. This sampling can lead to aliasing artifacts on the RIR and the loss of important paths. Another problem is occurring because of the modeling of the receiver and the sources as volumes of space. This is necessary to admit intersection with infinitely thin rays but can lead to false hits and paths counted multiple times.

The advantage of this method over the image source model is that it can easily take into account obstacles and listeners. These are usually modeled as volumetric objects in the room.

More advanced techniques such as beam-tracing [48] are used nowadays for increased performance. Instead of individual rays, they consider beams of rays through 3D polyhedral environments. The beams of rays have the advantage of covering every possible angle and therefore finding all possible paths up to the termination criteria. This technique can also integrate models of diffraction or scattering and are used nowadays for their real-time capabilities. They furthermore scale to support larger architectural environments.

Wave-based methods

Numerical wave-based methods are introduced when the wave equation cannot be solved analytically. These methods include the finite element method (FEM), the boundary element method (BEM) and the finite-difference time-domain method (FDTD) [24]. These methods find an approximate solution to the wave equation by solving the system of equations resulting from the discretization of the considered environment. In the FEM, the complete space is discretized while in BEM only the boundaries are discretized. The major drawback with these methods is that the discretization must be fine enough to capture irregularities and discontinuities of the sound field which becomes very time consuming. Therefore, these methods are suitable only for the low temporal frequency reproduction of the RIRs [70].

Methods such as TDFD also consider the wave equation but simplify the derivatives in finite differences. This simplifies greatly the problem and techniques using this principle such as digital waveguide meshes are widely used. The benefit of FDTD techniques is that one can create a denser mesh structure where required (e.g. near corners or challenging locations). The digital waveguide technique has received a lot of attention in the literature due to its simplicity and versatility [99, 98, 100, 118]. This technique is general and applications can be found in other domains such as musical instruments modeling [102].

2.3.2 Electromagnetic channel impulse responses

While most of the work in this thesis focuses on the acoustic wave propagation, similar results can also be derived in the study of the electromagnetic (EM) wave propagation. The wave equation is indeed the same, only the speed of propagation differs. The speed of wave propagation in the EM case is indeed the speed of light propagation and is of about $3 \cdot 10^8$ m/s.

Most of the research on modeling in the acoustic case focuses on room impulse responses. In the EM case, where the propagation happens over much larger distances than in the case of sound, the setups to be modeled are very different. In most of the EM models, receiving and emitting antennas are considered to be located outside of buildings. These buildings are creating multipath reflections. While rooms are most easily modeled with simple geometries, urban designs such as cities or neighborhoods are far more complex geometries [114]. This difficulty to model precisely the environment leads often to the use of statistical models. Furthermore, this environment is usually varying over time, which leads to the introduction of models for time-varying channels.

Modeling of channels has a long history. A lot of early activity in the 1960's is still relevant today. The characterization of time-varying channel for mobile communications has been studied by Bello [16], Clarke [35], Gans [49] and Kennedy [67]. More recent advances in the topic can be found in [52] and [107].

A lot of models have been proposed to describe spatio-temporal Channel impulse responses (CIRs) [46]. The goal is to model the channel in the most accurate way to be able to predict it. With the emergence of antenna arrays, it is possible to reduce the fading [52] observed by receivers using spatial diversity techniques [62]. A few existing models for the description of those channels are now presented. Among those, one directly can separate between deterministic and statistical models. For the deterministic models, the knowledge of the environment must be known in order to estimate the trajectories followed by the electromagnetic wave. For this purpose image source models [101] can be used as well as ray-tracing based algorithms. The second method is more versatile since it can take into account different effects such as reflection, diffraction and scattering as described in [109]. A large literature exists on the study of statistical models for CIRs. Early models consider angle of arrivals for the scatterers present in the channel as uniformly distributed. In Lee's model [80], scatterers are evenly spaced around the mobile. Under this assumption, Jakes derived theoretical results on the correlation between two receivers [62]. Based on experimental measurements, other models [97] have been presented taking

into account clustering effects observed in real data. In the present work, the purpose is not to present a new model of multipath propagation but to use simple existing models for the study of the spatial evolution of the CIRs. We are observing how the channels are varying when the source or the receiver are moving from one position to the next. In this context we are studying how fast the channel is modified depending on its temporal frequency content. This study is described in Chapter 6.

2.3.3 Head-related transfer functions

The direction-dependent transfer characteristics from a sound source to the ear is called *head-related impulse response* (HRIR) in the time-domain representation and *head-related transfer function* (HRTF) in the frequency-domain representation. They describe the acoustic filtering of a sound wave propagating in free field due to the presence of a listener. The HRTF is defined by [84]

Definition 2.3.1.

$$\hat{h}_{l,r}(\omega) = \frac{\text{FT of sound pressure at entrance of ear canal (l,r)}}{\text{FT of sound pressure in the middle of the head with listener absent}},$$

where FT stands for the Fourier transform.

HRTFs are therefore filters quantifying the effect of the shape of the head, body and pinnae on the sound arriving at the entrance of the ear canal. These filters are very important in 3D audio playback over headphones. A typical situation is described in Fig. 2.3. The source in free field emits the sound $s(t)$ that is recorded at the two ears. The HRIRs describing the filters between the source and the left ear or the right ear are denoted as $h_l(t)$ and $h_r(t)$ respectively. The sound recorded at the left ear, $e_l(t)$, and the right ear, $e_r(t)$, are therefore given by the following expressions

$$\begin{aligned} e_l(t) &= \quad s(t) * h_l(t) \\ e_r(t) &= \quad s(t) * h_r(t). \end{aligned} \tag{2.27}$$

Spatial hearing

Theoretically, by providing the signals $e_l(t)$ and $e_r(t)$ at the two ears of a subject using headphones[1], the person should perceive the exact same impression as if the original sound had been played by the source. This phenomenon is known

[1]Consider here that the headphone has a perfectly flat response.

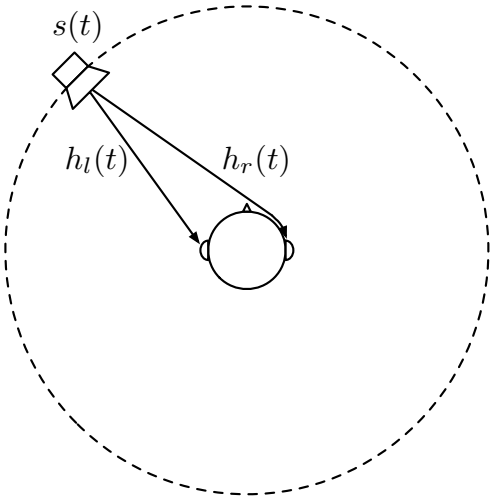


Figure 2.3: One source emitting sound $s(t)$ with the HRTFs corresponding to the left and right ears.

as *externalization*. We are able to deliver to the subject the impression that the sound is coming from outside of the head [44]. As is known from the literature of stereophony and stereophonic audio playback [20], without use of HRTFs the subject only can get the impression that the sound is coming from inside the head. It is then possible to modify different characteristics of the two-channels system to vary the origin and the width of the sound image. These parameters are Inter-channel time difference (ICTDs), inter-channel level differences (ICLD) and inter-channel coherence (ICC) [47].

To recreate as accurately as possible an impression of externalization, one should use individualized HRTFs. It means that each and everyone should possess his own HRTF dataset. Playback would be using the individualized HRTF set for optimal compatibility. For practical reasons, one often uses standard HRTFs measured on artificial heads [50]. When using non-individualized HRTFs, two main problems occur. Externalization is slightly less well perceived and front/back confusions are more often present. These confusions are related to the fact that some source positions that should appear in front are perceived in the back. In [27, 44], it is mentioned that better externalization can be achieved by adding some room models to the playback system as can be found in Section 2.3.1. There have been attempts to create model of human heads, such as the Valdemar model, which should correspond to humans in a generic way, so that HRTF recordings from these artificial heads would allow good localization results [34].

HRTFs are measured in anechoic chambers for every possible azimuthal and

elevation angle. The effect of the distance from the source is often neglected. The reason for this is that in real 3D audio systems the distance cue is not provided by HRTFs but by the room model used. Hartmann has discussed the role of the first room reflections on the perception of space and distance [55,56]. Nevertheless some publications can be found about the influence of the distance of the source on HRTFs [43,42]. In most of the HRTFs databases, the distance is of approximatively 1 to 1.5 m as is the case in three databases accessible online [50,1,12].

HRTF measurement

A few techniques exist to measure HRTFs. The simplest and most common technique consists in putting a microphone at the entrance of the ear canal and recording the sound emitted by a source located at a certain azimuth and elevation angle. By comparing the recorded signal and the emitted one, it is possible to extract the influence of the listener. This procedure needs to be repeated for every desired direction as shown in Fig. 2.4. Different stimuli can be used for the purpose of measuring HRTFs. White Gaussian noises, maximum lengths sequences (MLS) [23] or Golay codes are often used. Comparison of these signals can be found in [33].

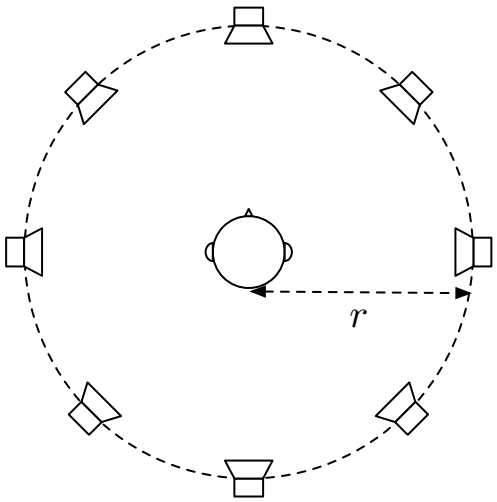


Figure 2.4: Setup for the recording of HRTFs.

A very interesting method has been presented by Zotkin et al. [120]. By putting a loudspeaker emitting sound near the ear entrance, the sound field is recorded with receivers located on a circular or on a spherical array of microphones. Like this, all HRTFs can be measured simultaneously.

A third technique is described in this thesis in Chapter 5 where either a setup with a moving source and a fixed person or with a fixed source and a moving person is considered. As a result of the proposed algorithm, it can be shown that the measurement of all possible HRTFs along the horizontal plane can be achieved in less than one second [8].

HRTF modeling and interpolation

When an HRTF dataset is measured, all possible positions for the source cannot be considered. An angular sampling needs to be performed. A major question arises then about what should this angular sampling be and what should be the method to interpolate the HRTFs between the considered sampling positions. In this thesis we answer these two questions in Chapter 4. In the present chapter, we present some existing techniques for HRTF modeling and interpolation.

One of the most simple and straightforward methods applies linear interpolation using the nearest neighbor HRTFs for obtaining HRTFs at any angle in between those [15].

A model for HRTFs has been proposed by Kulkarny et al. [73] which decomposes HRTFs as a minimum phase and an all-pass function. This model is based on previous studies [69] showing that HRTFs can be approximated by a minimum phase function up to 10 kHz. Minaar et al. [83] interpolated the minimum-phase components of HRTFs linearly in the time domain and claim to be able to predict the needed resolution of measurements without causing audible artifacts. Since the minimum phase components have a minimum phase lag, phase delay, and group delay for a given magnitude, they are optimally aligned in time.

This idea of alignment in time has been used by other authors to improve the quality of interpolation. It has indeed been shown that the performance of interpolation in the time or frequency domain can be improved by compensating the HRTFs prior to interpolation according to the time of arrival of sound [81]. That is, the HRTFs are time aligned and interpolation is carried out on the time-aligned HRTFs. Additionally, the time of arrival is interpolated separately. In [81], linear, spline and sinc interpolation are considered and the best results are obtained using linear interpolation and time alignment of the HRTFs. Hartung et al. [57] have compared other methods such as the inverse distance weighting method and the spherical splines method [95, 28]. As a result of psychoacoustics tests, they found out that the second method performed slightly better.

Other methods used for HRTF modeling are their representation using a

limited number of basis functions with corresponding weighting factors depending on azimuthal and elevation angles. The basis functions can for example be obtained by principal components analysis [69], independent component analysis [77] or spatial feature extraction and regularization [32]. Obviously, the more basis functions are used the better the modeling results are. Using measured HRTFs at known positions, the corresponding weighting factors are calculated. Interpolation of these weighting factors leads to the newly interpolated HRTFs.

Infinite impulse responses have also been considered by [21, 72] to model HRTFs. Two main models have been considered, the all-pole model and the pole-zero model. In [96], linear interpolation of the poles and zeros using a gradient search algorithm has been studied.

A very original and novel technique was described by [43]. There HRTFs are obtained as solution of a scattering problem [85]. The obtained solution is expanded using multipole expansion [54]. Using this representation, regularized fitting of measured HRTFs is applied and from there the parameters in the expansion can be found to describe the precise set of HRTFs. Then interpolation and range extrapolation can be achieved.

2.4 Wave Field Synthesis and Extrapolation

The typical audio reproduction systems such as stereo or 5.1 [47] are limited to reconstructing the sound field in one point, the sweet spot. These systems can only accurately render the sound field in one point. Therefore, only one listener located in the sweet spot listens to the sound as it was intended to be. At other positions, the sound field is not directly controlled. Often, however, It is desirable to reproduce the sound field such that it is reproduced accurately over a larger area.

The wave field synthesis technique achieves this by approximating the sound field not only in a single point, as has been done previously, but by aiming at reconstructing the sound field over a relatively large area. The price to pay to achieve this better spatial reconstruction of the sound field is the use of a much larger number of loudspeakers.

2.4.1 Huygens' principle

In 1690, Huygens proposed that a primary source can be replaced by an infinite number of secondary sources located along its wave front. This is illustrated in Figure 2.5. More generally, the secondary sources can be placed on a fixed surface S, e.g. on a plane, while still reproducing the sound field of the primary source. This is achieved by weighting and delaying the source signals as will

be described in the sequel.

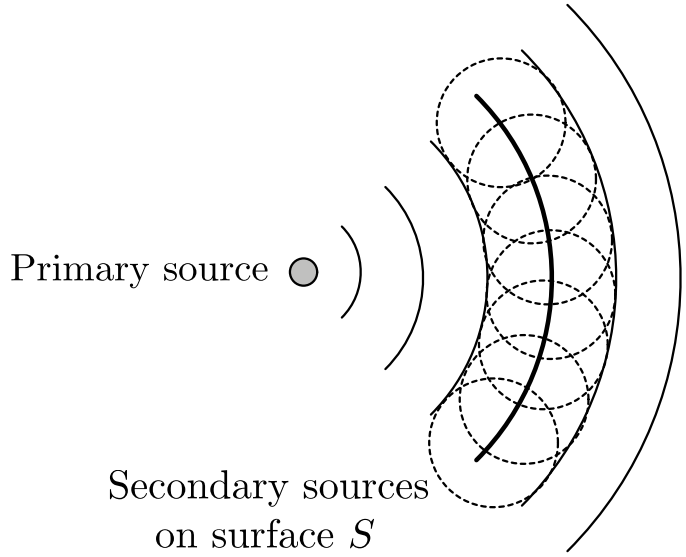


Figure 2.5: Huygens' principle: Any wave front can be modeled by an infinite number of point sources placed on the wave front.

The quantitative formulation of Huygens' principle is given by the Kirchhoff-Helmholtz integral,

$$\tilde{p}(\mathbf{r},\omega) = \frac{1}{4\pi} \oint_S \left[\tilde{p}(\mathbf{r}_S,\omega) \frac{\partial}{\partial \mathbf{n}} \left(\frac{e^{-j\frac{\omega}{c}\|\mathbf{r}\text{-}\mathbf{r}_S\|}}{\|\mathbf{r}\text{-}\mathbf{r}_S\|} \right) - \frac{\partial \tilde{p}(\mathbf{r}_S,\omega)}{\partial \mathbf{n}} \frac{e^{-j\frac{\omega}{c}\|\mathbf{r}\text{-}\mathbf{r}_S\|}}{\|\mathbf{r}\text{-}\mathbf{r}_S\|} \right] \mathrm{d}S, \quad (2.28)$$

which can be derived from the wave equation (2.2) using Green's integral theorem [18,17,110]. Figure 2.6 [2] illustrates the different variables used in (2.28). S represents the surface of a source free volume V, $\|\mathbf{r}\text{-}\mathbf{r}_S\|$ is the distance from the surface point $\mathbf{r}_S$ to a point $\mathbf{r}$ within the volume V, $\mathbf{n}$ is the vector normal to the surface and $\tilde{p}(\mathbf{r}_S,\omega)$ is the Fourier transform of the sound pressure field on S.

Consider a source free volume V delimited by a surface S. From (2.28), it can be observed that the sound pressure $\tilde{p}(\mathbf{r},\omega)$ inside of the volume V, can be obtained by measuring the pressure on the surface $\tilde{p}(\mathbf{r}_S,\omega)$ as well as its particle velocity being proportional to $\frac{\partial \tilde{p}(\mathbf{r}_S,\omega)}{\partial \mathbf{n}}$.

2.4.2 The Rayleigh I integral

Simplification of the Kirchhoff-Helmholtz integral can be obtained when considering all primary sources in one half space. The surface S is degenerated

[2] This figure is a slight modification of Fig. 2 in [105].

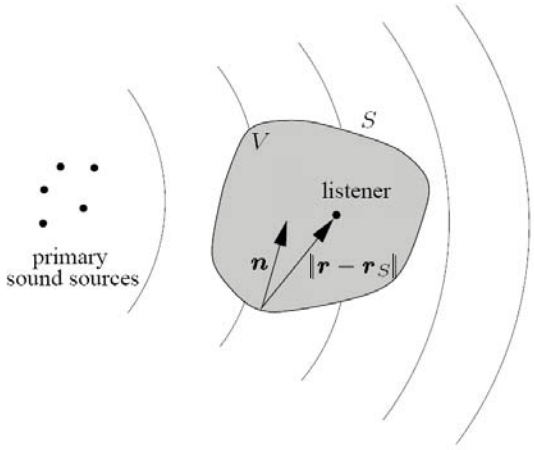


Figure 2.6: Definition of the parameters used for the Kirchhoff-Helmholtz integral.

to a plane between the primary sources and the listening area as illustrated in Figure 2.7(a). The Rayleigh I integral [18, 17, 110] corresponds to a special case of the Kirchhoff-Helmholtz where only the particle velocity needs to be measured on the surface S. The details of its derivation can be found in [110] where the first term in (2.28) is canceled by a careful choice of the pressure field. The Rayleigh I Integral is given by

$$\tilde{p}(\mathbf{r},\omega) = \rho_0 c \frac{j\omega}{2\pi c} \int \left[\tilde{v}_n(\mathbf{r}_S,\omega) \frac{e^{-j\frac{\omega}{c}\|\mathbf{r}\text{-}\mathbf{r}_S\|}}{\|\mathbf{r}\text{-}\mathbf{r}_S\|} \right] \mathrm{d}S\,, \tag{2.29}$$

where ρ_0 is the density of air, c is the speed of sound and $\tilde{v}_n(\mathbf{r}_S,\omega)$ denotes the particle velocity perpendicular to the surface S. From (2.29) and more generally from (2.28), two interesting applications can be considered, the wave field extrapolation and WFS. The first one considers the measurement of the particle velocity along the surface S and by processing it, the sound field can be extrapolated inside the volume using the Rayleigh integral [60]. A second approach is to measure the field on the surface S and to reproduce the same field inside of the volume by means of drivable sources located on the surface. At each position $\mathbf{r}_S$, a monopole is driven with a signal proportional to the particle velocity perpendicular to the surface S,

$$\tilde{d}(\mathbf{r}_S,\omega) = \rho_0 c \frac{j\omega}{2\pi c} \tilde{v}_n(\mathbf{r}_S,\omega)\,, \tag{2.30}$$

to reconstruct a sound field equivalent to the original sound field as illustrated in Figure 2.7(b).

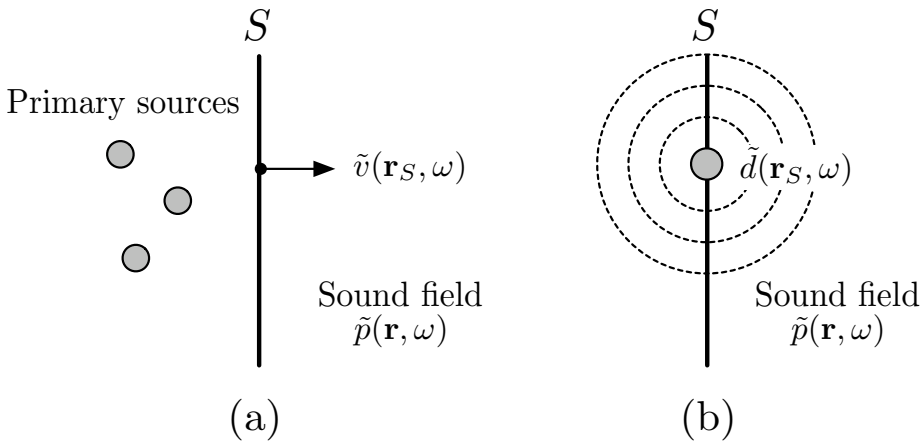


Figure 2.7: (a) The sound field in the source free area to the right of S is completely specified by the particle velocity $\tilde{v}_n(\mathbf{r}_S, \omega)$ orthogonal to S. (b) The sound field is reconstructed using monopole sources on S driven by the signals $\tilde{d}(\mathbf{r}_S, \omega)$.

2.4.3 Wave field synthesis

The presented theory has shown that it is theoretically possible to reproduce the sound field generated by sound sources located in one half space, into the other half space. The condition is to place an infinite number of sources in the infinite plane separating the two half spaces and to drive these sources with the correct signal as described in Section 2.4.2. From this theoretical scheme, many simplifications have been applied to develop a more realistic setup. We present here a very brief overview of the different steps performed to allow a realizable reproduction system using WFS.

The main drawback of the presented technique is the necessity to use an infinite number of loudspeakers along two dimensions. An important work done by Berkhout et al. [17] was to study an equivalent form of the Rayleigh I integral for the case where only one line of loudspeakers was used. The idea is to place the loudspeaker line at the same height as the ears of listeners and to reproduce the sound field correctly in the largest possible area. It was unfortunately shown that the sound field is then correctly reproduced only along one line parallel to the loudspeaker array and at the same height [110]. By varying some parameters in the driving signals, the distance of the reproduction line can be modified. By an appropriate choice of this distance of reproduction, a large listening area will have a sound field close to the original field.

A second important step in the simplification of the problem is to discretize and limit the aperture of the loudspeaker line. These two factors are fundamental in the design of the WFS system. A too large distance between consecutive loudspeakers limits the correct reproduction for high frequencies. This topic of the spatial sampling of the loudspeakers in WFS systems has been described in [18] and more recently by [104]. The length of the array also plays an important role in the correct reproduction of the sound field over a large area [17].

WFS systems suffer from the fact that reverberation cannot be totally reproduced. Reverberation as modeled by the image sources, shows that some new sources are created on the half space that should be source free. These sources can therefore not be reproduced in the listening area. To counteract this problem, one usually uses artificial reverberation to mimick their effect [110,61] but this remains a drawback of this technique.

2.5 Conclusion

This chapter has presented general concepts that are recurrent in this thesis. The plenacoustic function describes the sound field at any position in space. Therefore, the wave equation has been given and its solution explored in different cases. The wave equation can be solved in simple cases but as soon as the geometry of the problem become slightly complex, models have to be introduced to study the sound field in rooms. Therefore, impulse response models have been described for rooms, electromagnetic channels and head-related transfer functions. As an application of spatio-temporal processing, wave field synthesis has been discussed. This technique reproduces the sound field generated by primary sources using loudspeaker arrays driven by the appropriate signals.

Chapter 3

Sound Field Analysis and Interpolation Along a Line and Multidimensional Spatial Positions

3.1 Introduction

This chapter analyzes the sound field along a line and multidimensional spatial positions such as planes and spaces filled with microphones. The main discussion presented in this work relates to the following question: "How many microphones are needed in space in order to recreate as accurately as possible the sound field at all positions?" Beyond the fundamental interest of characterizing precisely the sound field and its sampling, the results are useful in spatial audio applications. For example, it indicates to what extent a microphone array can be used to interpolate any spatial location. Or conversely, how many spatial positions of a source are needed to synthesize arbitrary positions for a virtual source.

For the time dimension, bandlimited sources are assumed to allow sampling in time. The more interesting question is of course sampling in space, which directly relates to the number of microphones necessary to acquire the PAF. When considering far field assumptions, it can be shown [63] that the support of the spatio-temporal PAF is bandlimited to a spatial frequency ϕ which is

related to the temporal frequency ω in a linear manner[1]

$$\phi = \frac{\omega}{c}. \tag{3.1}$$

Thus, if the time domain signal is bandlimited to ω_0, the spatial frequency is limited to $\frac{\omega_0}{c}$, and the far field PAF can be sampled with any spatial distance Δx satisfying

$$\Delta x > \frac{c\pi}{\omega_0}. \tag{3.2}$$

Without the far field assumption, the PAF is not perfectly bandlimited anymore and contains energy in the region $\phi \geq \frac{\omega_0}{c}$. In this chapter, the PAF is studied without the far field assumption and its spectral characteristics are derived both in the case of free field and in reverberating rooms. From this analysis, we derive a precise quantitative sampling theorem which trades off spatial sampling density for signal reconstruction quality. The developed theory is compared with simulations and experimental results.

3.1.1 Related work

As mentioned earlier, previous literature exists on the bandlimitedness of the spatial sound pressure field. In [87, 63, 117] using microphone arrays, the sound pressure field is studied both along the temporal and the spatial dimensions. The spectrum of the spatio-temporal wave equation is studied under the far field assumption which simplifies the obtained sampling results [87, 63, 117]. Recently, and in parallel to our work, Coleman [37, 36] has investigated the wide-band electromagnetic impulse response in far field, deriving sampling results under this assumption for linear arrays and planar arrays of sensors.

In [18], Berkhout studies in detail the extrapolation of waves in the field of seismic wave theory. Results on the sampling and the extrapolation of wave fields are obtained using the representation of the wave field by its spatio-temporal spectrum. Also practical aspects such as the aperture size of the array are discussed. Recent techniques have shown interesting extrapolation results using a one-dimensional (1D) microphone arrays (mostly circular arrays) but limitations occur when trying to extrapolate real three-dimensional (3D) RIRs [60]. Similarly to the extrapolation of the sound field, a technique called wave field synthesis (WFS) [39] has been described in detail in Chapter 2. It is explained there that from the knowledge of the sound field in a region of

[1] In acoustics as well as in some array signal processing books the spatial and temporal frequencies are expressed as wave numbers. Then, $\frac{\omega}{c}$ is usually denoted as k and the spatial frequency ϕ as k_x.

space, the WFS reproduces the sound field in other regions of space. The WFS is based on the Huygens' principle stating that the propagation of a wave through a medium can be described by adding the contributions of all secondary sources positioned along a wave front. Measuring the sound field on an infinite plane of microphones would allow its accurate reproduction at any point of the source-free half space. Dual to the WFS is wave field analysis [19] where the ray parameter versus intercept time domain representation is used. This representation is shown to be equivalent to taking the linear Radon transform of the spatio-temporal RIRs. The method gives an improved insight in the structure of complex sound pressure fields by being able to separate the different contributions of the sound pressure field (direct path and each reflection on the walls).

3.1.2 Contributions

This chapters presents the following contributions:

- A detailed study of the decay of the PAF spectrum along the spatial and temporal frequency axes, both in free field and in rooms with reverberation, for the case of the PAF studied along a line.
- A detailed study of the spatial decay of the PAF spectrum for the PAF studied along a plane and in three dimensions.
- A sampling theorem describing the trade-off existing between spatial sampling frequency and quality of reconstruction of the sound field when sampled along aline.
- Simulations and experimental results for the acoustic and electromagnetic cases verifying the presented theory.

3.1.3 Outline

The outline of the chapter is the following. In Section 3.2, the PAF and its construction is presented. Section 3.2.1 reviews RIRs and how they can be simulated, while Section 3.2.2 describes the space-time representation. Section 3.3 studies the spectrum of the PAF on a line in a room. Its spatial and temporal frequency decay are described in Section 3.3.2 and 3.3.3 respectively. Section 3.4 studies the sampling of the PAF. The sampling theorem of the PAF is presented in Section 3.4.1 followed by a discussion on different sampling patterns in Section 3.4.2. Based on the sampled function, one would like to reconstruct the field in every possible position which is shown in Section 3.4.3. Limitations due to the finite length of the array are taken in

account in Section 3.5. The theory presented in this chapter is then verified using simulations in Section 3.6.1 and measurements done in real environments in Section 3.6.2. Section 3.7 is devoted to the generalization of the PAF to multidimensional spatial positions. The conclusions are drawn in Section 3.8.

3.2 Construction of the Plenacoustic Function

To study the sound pressure field along a line of microphones in a room, one needs to study the sound pressure field from every possible source position in the room to any possible microphone position on the line. For simplicity the technique is presented for a single source but it will be shown later that the technique works as well for multiple sources. Consider a source s emitting a signal $s(t)$. The microphones located on the line will not record exactly $s(t)$. The sound at microphone m_1 is $s(t)$ convolved with the RIR corresponding to the direct path between s and m_1, followed by a possibly infinite number of reflections on the walls (each microphone will receive a sum of delayed and attenuated versions of $s(t)$). At another microphone position m_2, the recorded signal will be different since the RIR from s to m_2 is different than the RIR from s to m_1. The only parameter changing between the different spatial positions is the RIR. Therefore the rest of the analysis of this chapter will be focused on the spatial evolution of the RIRs. Also for that reason, the name of the PAF will be used for the rest of the chapter as the spatio-temporal RIRs.

3.2.1 Modeling the room

In order to construct the PAF in a room, the RIRs at any point in the room need to be known. The image method discussed in [13] and in Section 2.3.1 for the simulations of RIRs is used. The method is based on the creation of virtual sources in order to simulate the effect of the reflections on the walls. In the case of a rectangular rigid-walls room of size (L_x, L_y, L_z), the RIRs are given by (2.26). Special attention has to be given to the problem of the quantization rounding in the computation of the RIR. In [13], the delay corresponding to each virtual source was rounded to the closest sample in time in order to reduce the complexity of the simulations. This leads to aliasing in time and space. In our simulations, each dirac has been replaced by a sinc function of appropriate bandwidth delayed with the exact non integer delay. This removes the aliasing effect. However, as the sinc functions have a very slow decay in time, one has to consider long enough RIRs to allow the sincs to sufficiently vanish.

3.2.2 Space time representation

With the RIRs as defined in (2.26), the PAF is constructed for a line in the room. In that case, one can construct a two-dimensional (2D) function by gathering all the RIRs at any position on the line, leading to a 2D continuous function of space and time. Space represents the position, time being the duration of the RIR. This representation is shown in Fig. 3.1(a) when a pulse is recorded on a line of microphones in free field and in Fig. 3.1(b) for the case of a room.

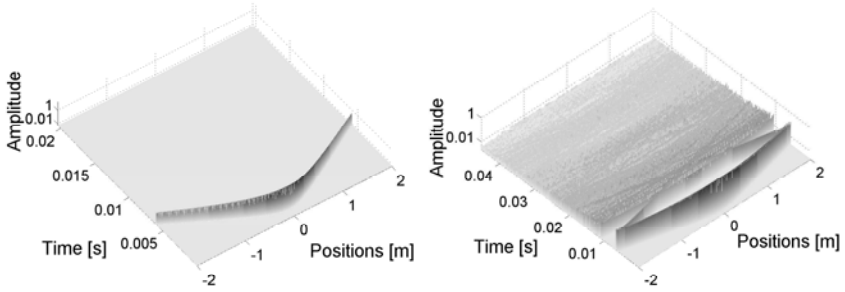


Figure 3.1: PAF in time and space. (a) In free field. (b) Inside a room.

3.3 Spectrum of the Plenacoustic Function on a Line

In this section, the PAF on a line and its associated spectrum are studied. First, a description of the 2D spectrum of the PAF under the far field assumption is given in Section 3.3.1. Further, the PAF on a line and its associated spectrum are studied without any far field assumption. An analytical expression of the two-dimensional Fourier transform (2D-FT) of the PAF is given in Section 3.3.2 together with an analysis of the spatial frequency decay of the spectrum of the PAF. This analysis is performed for a single source and further in the case of a rectangular room. The analysis of the temporal frequency decay in the case of a single source and for a rectangular room is described in Section 3.3.3.

3.3.1 Far field assumption

Often, in the literature [87,63], the far field assumption is considered to simplify matters. In the context of the sound field studied along a line, the far field

assumption considers the source far enough away from the microphones such that the wave front arriving on the line of microphones appears as a plane wave. The plane wave is defined by its angle of arrival α as shown in Fig. 3.2(a). The PAF in space and time can then be expressed as:

$$p(x,t) = \delta\left(t - t_s - \frac{x\cos\alpha}{c}\right), \tag{3.3}$$

where c stands for the speed of sound propagation and t_s for the time taken by the sound to travel from the source to the center of the microphone line. For simplicity, we consider $t_s = 0$ meaning that at time $t = 0$, the plane wave passes by the origin of the x-axis. The 2D-FT of (3.3) is shown in Fig. 3.2(b) and given by

$$\hat{p}(\phi,\omega) = 2\pi\delta\left(\phi + \omega\frac{\cos\alpha}{c}\right), \tag{3.4}$$

with ϕ and ω being respectively the spatial and temporal frequencies. Considering now that plane waves arrive from every possible angles, the support of the spectrum of the impulse responses gathered by the microphones lies exactly on a butterfly shaped spectrum as is presented in Fig. 3.2(c). Outside of the spectrum, no energy is present [37, 36]. This result allows us to sample the sound field and perfectly reconstruct it, since the spectrum is perfectly bandlimited. Therefore, as mentioned already in the introduction of this chapter, choosing the spatial spacing between consecutive microphones as in (3.2) leads to perfect reconstruction of the field. This result is only possible due to the far field assumption and is not realistic. In the next section, the 2D-FT of the sound field is described when no far field assumption is applied. It is then shown that there is still energy outside of the butterfly shaped spectrum which prevents the perfect reconstruction.

3.3.2 Spatial decay of the spectrum of the plenacoustic function

An analytical expression for the 2D-FT of the PAF is given for the free field case, followed by a general formula in the case of a rectangular room.

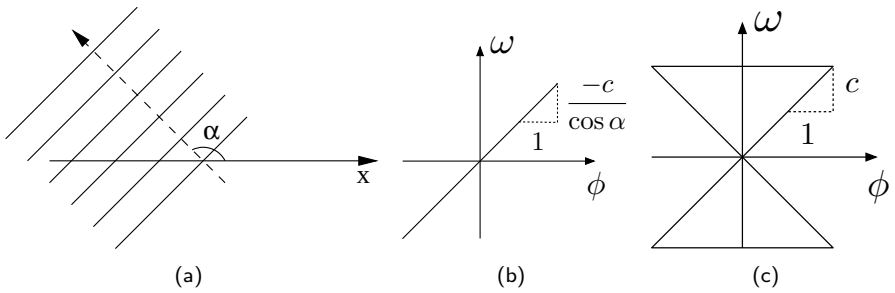


Figure 3.2: Far field assumption for the spatio-temporal study of the wave field. (a) Scheme of a plane wave arriving on a line of microphone with angle α. (b) two-dimensional spectrum of the spatio-temporal impulse response. (c) Considering all possible angles of arrival, the support of the wave field is described by a butterfly shaped spectrum.

Free field case

The evolution of the RIR along the x-axis is studied. As described in Chapter 2, the PAF in space and time domain is given by the following formula [85]:

$$p(x,t) = \frac{\delta\left(t - \frac{\sqrt{(x-x_s)^2+(y_m-y_s)^2+(z_m-z_s)^2}}{c}\right)}{4\pi\sqrt{(x-x_s)^2+(y_m-y_s)^2+(z_m-z_s)^2}}. \tag{3.5}$$

The x component of the microphone is only varied. For simplicity, the subscript in the variable x_m was removed, denoting it by x. The variables y_m, z_m, x_s, y_s and z_s are constant. Calling $d^2 = (y_m - y_s)^2 + (z_m - z_s)^2$, (3.5) is rewritten as

$$p(x,t) = \frac{\delta\left(t - \frac{\sqrt{(x-x_s)^2+d^2}}{c}\right)}{4\pi\sqrt{(x-x_s)^2+d^2}}. \tag{3.6}$$

The 2D-FT of this function is calculated in Appendix A.1. The obtained result for $\omega \in \mathbb{R}^+$ and $\phi \in \mathbb{R}$ is[2]:

$$\hat{p}(\phi,\omega) \quad = \quad -\frac{j}{4}e^{-j\phi x_s}H_0^{(1)*}\left(d\sqrt{\left(\frac{\omega}{c}\right)^2 - \phi^2}\right), \tag{3.7}$$

with ϕ and ω being respectively the spatial and temporal frequencies. The magnitude of (3.7) is plotted for $d = 1$ in Fig. 3.3 for positive frequencies. $H_0^{(1)*}$ represents the complex conjugate of the zero order Hankel function of the first

[2]Since $p(x,t)$ is a real function, one has that $\hat{p}(-\phi,-\omega) = \hat{p}^*(\phi,\omega)$, with $\hat{p}^*$ the complex conjugate of $\hat{p}$. To simplify the notation, all further derivations are done for $\omega \in \mathbb{R}^+$.

kind. This function is infinite at zero. Therefore when either $d = 0$ or $|\phi| = \frac{\omega}{c}$ the plenacoustic spectrum becomes infinite. When $d = 0$, it corresponds to the situation where the source is located on the line of the microphones. For the case of $|\phi| = \frac{\omega}{c}$, a more intuitive explanation is given. Consider a sinusoid of temporal frequency ω rad/s emitted from a certain position. The signal acquired by the microphones located at positions tending to infinity is at one instant an attenuated sinusoid of spatial frequency $\frac{\omega}{c}$ rad/m. For these microphone positions the source appears as being on the line. This explains the infinite value of the spectrum for this spatial frequency.

The values where $|\phi| > \frac{\omega}{c}$ correspond to the evanescent mode of the waves. The waves lose their propagating character to become exponentially fast decaying waves [112, 18]. Therefore, considering $\omega, \phi \in \mathbb{R}$, most of the energy is contained in the part of the spectrum satisfying

$$|\phi| \leq \frac{|\omega|}{c}. \tag{3.8}$$

This result will be used later in the sampling of the PAF. As the spectrum is decaying very fast along the spatial frequency axis, one will be able to derive a sampling theorem to sample and reconstruct the PAF along the spatial axis (see Section 3.4).

An intuitive explanation is now given for the butterfly shape of the spectrum. At low temporal frequencies, the sound wavelength is very large and therefore spatially, the wave is varying slowly which explains the small spatial support. For higher temporal frequencies, the sound wavelength is smaller what makes the spatial variation larger. The spatial support for higher temporal frequencies is therefore larger.

Note also that for $|\phi| \geq \frac{\omega}{c}$, the argument of the Hankel function in (3.7) becomes imaginary and (3.7) can be rewritten as:

$$\hat{p}(\phi, \omega) \quad = \quad \frac{1}{2\pi} e^{-j\phi x_s} K_0 \left(d\sqrt{\phi^2 - \left(\frac{\omega}{c}\right)^2} \right), \tag{3.9}$$

where K_0 is a modified Bessel function of the second kind and order zero. The modified Bessel function of the second kind has the following asymptotical behavior (see [112]):

$$K_0(x) \sim \sqrt{\frac{\pi}{2x}} e^{-x}. \tag{3.10}$$

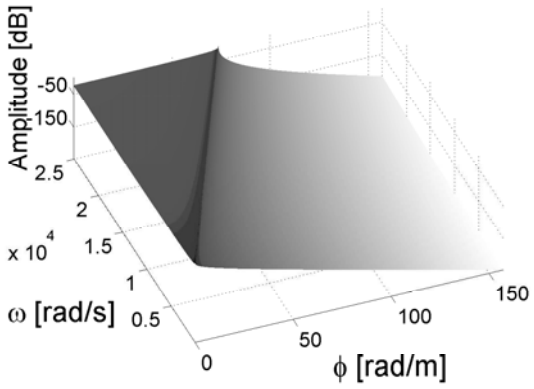


Figure 3.3: Theoretical 2D spectrum of the PAF according to (3.7) for $d = 1$.

Note that further numerical computations show that

$$K_0(x) \leq \sqrt{\frac{\pi}{2x}} e^{-x} \quad \text{for} \quad x > 0. \tag{3.11}$$

For large ϕ, (3.9) can be rewritten using (3.10) as:

$$\hat{p}(\phi, \omega) \sim \frac{1}{2\pi} e^{-j\phi x_s} \sqrt{\frac{\pi}{2d\sqrt{\phi^2 - (\frac{\omega}{c})^2}}} e^{-d\sqrt{\phi^2 - (\frac{\omega}{c})^2}}. \tag{3.12}$$

For a fixed $\omega = \omega_0$, (3.12) asymptotically behaves as:

$$\hat{p}(\phi, \omega_0) \sim \frac{e^{-j\phi x_s}}{2\sqrt{\pi}} \frac{e^{-d\phi}}{\sqrt{d\phi}}. \tag{3.13}$$

The decay along the spatial frequency axis is faster than exponential.

Rectangular room

In the case of a rectangular room of size (L_x, L_y, L_z) with perfectly reflecting walls, all the reflections are considered as virtual sources as explained in Section 3.2.1 and the superposition principle is applied. The expression for the PAF is then given by (2.26). Each virtual source leads to a spectrum that follows (3.7). The total spectrum of the PAF is the sum of the spectra of each virtual source taken separately, leading to an infinite sum. One would like to know how this sum is decaying for large spatial frequencies.

Results on the decay of the spatial frequency are presented in the easier case of all the virtual sources located in the plane. Similar results can be obtained in the general case of sources located in space.

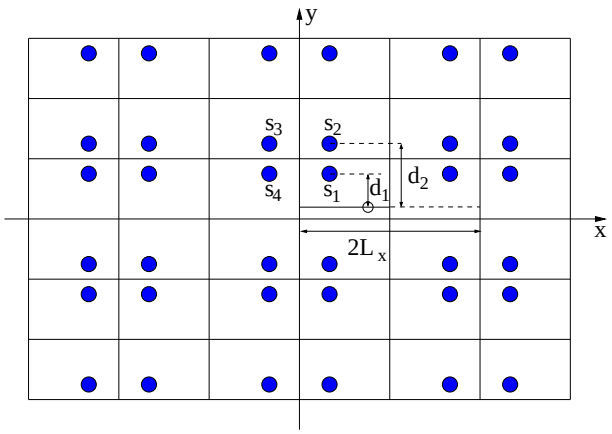


Figure 3.4: Image source model with the original source s_1 and all the other virtual sources.

The image model given in Fig. 3.4 was used. Our original source is s_1 (with coordinates (x_{s_1}, y_{s_1})) and in its immediate neighborhood, one can see three other virtual sources (s_2, s_3 and s_4). These four mother sources will create an infinite number of repetitions to form all the virtual sources in the plane. These four mother sources will be repeated in the x and y directions with a periodicity of $2L_x$ and $2L_y$, respectively.

As the room has finite size, an infinite line of microphones cannot be considered. The microphone line is considered to cover the whole length of the room. In Fig. 3.4, the line is parallel to the x axis.

Define the distances from sources s_1 and s_2 to the line of microphones as d_1 and d_2 respectively (with $d_1 \leq d_2$). The other distances for the sources s_4 and s_3 to the line are in this case also d_1 and d_2.

An interesting aspect of this construction is that by the $2L_x$ periodicity of the source positions along the x axis, the sound recorded on an infinite line is also periodic with period $2L_x$. Further, using the symmetry of the construction, one realizes that the sound heard at position a is the same as the one heard at position $2L_x - a$ with $a \in [0, L_x]$.

By sampling the spectrum of the PAF on the ϕ axis at multiples of $\phi_0 = \frac{\pi}{L_x}$, one exactly obtains the Fourier series of the sound pressure field recorded on a line from 0 to $2L_x$ when this sound pressure field is $2L_x$ periodic.

Considering the four mother sources (s_1 to s_4) with their periodic repetitions along the x axis, the discrete spectrum of the PAF can be expressed as

follows (for large $n\phi_0$):

$$\hat{p}(n\phi_0,\omega_0) \sim \left(\frac{e^{-jn\phi_0 x_{s_1}}}{2\sqrt{\pi}} + \frac{e^{-jn\phi_0 x_{s_2}}}{2\sqrt{\pi}}\right) \left(\frac{e^{-d_1 n\phi_0}}{\sqrt{d_1 n\phi_0}} + \frac{e^{-d_2 n\phi_0}}{\sqrt{d_2 n\phi_0}}\right). \tag{3.14}$$

Therefore,

$$\hat{p}(n\phi_0,\omega_0) \sim c_1(n)\left(\frac{e^{-d_1 n\phi_0}}{\sqrt{d_1 n\phi_0}} + \frac{e^{-d_2 n\phi_0}}{\sqrt{d_2 n\phi_0}}\right), \tag{3.15}$$

with $c_1(n)$ a bounded function in n.

Consider now the $2L_y$ periodic repetitions of the sources along the y axis. Call the sources $s_{1,i}$ the sources with coordinates $(x_{s_1}, y_{s_1} + i2L_y)$ and similarly $s_{2,i}$ the sources with coordinates $(x_{s_2}, y_{s_2} + i2L_y)$. Call $D_{1,i}$ the distances between the line of microphones and the sources $s_{1,i}$, and $D_{2,i}$ the distances between the line of microphones and the sources $s_{2,i}$.

$$D_{1,i} = |d_1 + i2L_y|, \tag{3.16}$$

$$D_{2,i} = |d_2 + i2L_y|. \tag{3.17}$$

When considering all the source repetitions in the x and y directions, the spectrum becomes:

$$\hat{p}(n\phi_0,\omega_0) \sim c_1(n)\sum_{i=-\infty}^{\infty}\left(\frac{e^{-D_{1,i} n\phi_0}}{\sqrt{D_{1,i} n\phi_0}} + \frac{e^{-D_{2,i} n\phi_0}}{\sqrt{D_{2,i} n\phi_0}}\right). \tag{3.18}$$

The right side of (3.18) can be upperbounded by:

$$c_1(n)\sum_{i=0}^{\infty}\left(\frac{e^{-(d_1+i2L_y)n\phi_0}}{\sqrt{d_1 n\phi_0}} + \frac{e^{-(d_2+i2L_y)n\phi_0}}{\sqrt{d_2 n\phi_0}} + \frac{e^{-(d_1'+i2L_y)n\phi_0}}{\sqrt{d_1' n\phi_0}} + \frac{e^{-(d_2'+i2L_y)n\phi_0}}{\sqrt{d_2' n\phi_0}}\right), \tag{3.19}$$

with $d_1' = 2L_y - d_1$ and $d_2' = 2L_y - d_2$. Since $d_1 \leq d_2 \leq L_y$, (3.19) can be upperbounded by

$$\frac{c_1(n)\left(e^{-d_1 n\phi_0} + e^{-d_2 n\phi_0} + e^{-d_1' n\phi_0} + e^{-d_2' n\phi_0}\right)}{\left(1 - e^{-2L_y n\phi_0}\right)\sqrt{n d_1 \phi_0}}. \tag{3.20}$$

Asymptotically for large n, the above expression is of the following order:

$$\hat{p}(n\phi_0, \omega_0) \in O\left(\frac{e^{-d_1 n\phi_0}}{\sqrt{n}}\right). \tag{3.21}$$

This shows that for a reverberant room, the decay is faster than exponential when the line of microphones is parallel to a wall[3].

3.3.3 Temporal frequency decay

The study of the temporal frequency decay is of interest to fully characterize the plenacoustic function. Nevertheless, in most cases sounds are bandlimited along the temporal frequency due to the bandwidth of the emitters and receivers. Therefore, the results of this section will be briefly presented since they only are interesting from a theoretical point of view. The more detailed analysis of the presented results can be found in [5]. Similarly to Section 3.3.2, results on the temporal frequency decay for the free field case are first presented before generalizing them for a rectangular room.

Free field

The spectrum of the PAF is given by expression (3.7). The asymptotic behavior of the Hankel function is given by [112]:

$$H_0^{(1)}(x) \sim \sqrt{\frac{2}{\pi x}} e^{j(x-\pi/4)}. \tag{3.22}$$

For large ω, (3.7) can be rewritten using (3.22) as:

$$\hat{p}(\phi, \omega) \sim -\frac{j e^{-j(\phi x_s - \pi/4)}}{2\sqrt{2\pi}} \frac{e^{-jd\sqrt{(\frac{\omega}{c})^2 - \phi^2}}}{\sqrt{d\sqrt{(\frac{\omega}{c})^2 - \phi^2}}}. \tag{3.23}$$

Considering a fixed $\phi = \phi_0$, (3.7) asymptotically behaves as:

$$\hat{p}(\phi_0, \omega) \sim -\frac{j\sqrt{c}}{2\sqrt{2\pi}} e^{-j(\phi_0 x_s - \pi/4)} \frac{e^{-jd\frac{\omega}{c}}}{\sqrt{d\omega}}. \tag{3.24}$$

[3]The case where the line of microphones is not covering the whole length of the room can be seen as a windowing and is discussed in Section 3.5. The case where the line is not parallel to the wall is studied in Appendix B. There, the line is extended along the periodic repetitions of the room and it is shown that the measured sound field has a spectrum decaying slightly faster than an exponential. The restriction of this infinite line inside the room can then again be seen as a windowing as explained in Section 3.5.

Therefore,

$$\hat{p}(\phi_0, \omega) \sim \frac{c_2(\omega)}{\sqrt{\omega}}, \tag{3.25}$$

with $c_2(\omega)$ a bounded function of ω. This last relation shows that the decay of the PAF spectrum along the temporal frequency is, up to a constant, asymptotic as $\frac{1}{\sqrt{\omega}}$.

Rectangular room

In the case of a rectangular room, one follows the same construction as in Section 3.3.2. Considering the four mother sources (s_1 to s_4) in Fig. 3.4 with their periodic repetitions along the x axis, the discrete spectrum of the PAF can be expressed as follows (for large ω and a finite $n = n_0$):

$$\begin{aligned} \hat{p}(n_0\phi_0, \omega) &\sim \frac{-j\sqrt{c}e^{j\pi/4}}{2\sqrt{2\pi}} \left(e^{-jn_0\phi_0 x_{s_1}} + e^{-jn_0\phi_0 x_{s_2}}\right) \\ &\qquad \left(\frac{e^{-jd_1\frac{\omega}{c}}}{\sqrt{d_1\omega}} + \frac{e^{-jd_2\frac{\omega}{c}}}{\sqrt{d_2\omega}}\right) \qquad (3.26) \\ &\sim c_3(n_0)\left(\frac{e^{-jd_1\frac{\omega}{c}}}{\sqrt{d_1\omega}} + \frac{e^{-jd_2\frac{\omega}{c}}}{\sqrt{d_2\omega}}\right), \qquad (3.27) \end{aligned}$$

with $c_3(n_0)$ being independent of ω.

Considering the $2L_y$ periodic repetitions of the sources along the y axis, one obtains:

$$\hat{p}(n_0\phi_0, \omega) \sim c_3(n_0) \sum_{i=-\infty}^{\infty} \left(\frac{e^{-jD_{1,i}\frac{\omega}{c}}}{\sqrt{D_{1,i}\omega}} + \frac{e^{-jD_{2,i}\frac{\omega}{c}}}{\sqrt{D_{2,i}\omega}}\right). \tag{3.28}$$

In Appendix C, this sum is shown to converge and can be written as

$$\hat{p}(n_0\phi_0, \omega) \sim \frac{c_4(\omega)}{\sqrt{\omega}}, \tag{3.29}$$

with $c_4(\omega)$ a bounded function of ω.

3.4 Sampling and Reconstruction

In the previous sections, the decay of the 2D spectrum of the PAF has been studied both along the temporal and the spatial frequency axes. One has observed that the spectrum of the PAF lies on a support that is almost bandlimited. This result is valid for a single source, but also for a finite number of sources as well as for reverberation. In the scope of this section, we are

mostly interested in the sampling of the sound pressure field along the spatial axis. We consider that the sound pressure field positions are recorded with omnidirectional point microphones[4]. Since no spatial anti-aliasing filter can be applied in the spatial direction, the speed of the spatial frequency decay of the 2D spectrum of the PAF is the key factor for the quality of the reconstruction. Along the temporal direction, the signal can be filtered in order to avoid aliasing.

In this section, the sampling theorem of the PAF is presented. The quality of the reconstruction is studied when sampling the sound pressure field in space. Further, interpolation techniques are discussed in order to reconstruct the signal from the available samples.

3.4.1 A Sampling theorem for the plenacoustic function

In order to uniformly sample the PAF along the spatial direction, a uniformly spaced infinite number of impulse responses are considered. Call ϕ_S the spatial sampling frequency defined as $\frac{2\pi}{\Delta x}$ where Δx is the sampling interval between two consecutive positions of the measured impulse responses. Next to the spatial sampling, the RIRs also need to be sampled at a certain temporal sampling rate depending on the desired audio bandwidth. Call ω_S the temporal sampling frequency, or $\omega_S = \frac{2\pi}{\Delta t}$ with Δt the sampling period of the impulse responses.

Consider now the spectrum of the PAF given by (3.7) at a particular temporal frequency ω_0. It has approximatively the shape given in Fig. 3.5(a). When the PAF is sampled, repetitions of the spectrum occur as shown in Fig. 3.5(b) and the obtained spectrum is denoted as $\hat{p}_S(\phi, \omega)$. As the spectrum is not perfectly bandlimited, the repetitions will affect the reconstruction. A theorem quantifying the signal to noise ratio (SNR) of the reconstruction of the PAF is presented for one source emitting in free field. Call $\mathrm{SNR}(\phi_S, \omega_0)$ the SNR of the reconstruction for a sinusoid emitted at frequency $\omega = \omega_0$ with the microphones positioned with a spatial sampling frequency ϕ_S. In the present case, the SNR is defined as follows:

$$\mathrm{SNR}(\phi_S, \omega_0) = \frac{\int_{\phi=-\infty}^{\infty} |\hat{p}(\phi, \omega_0)|^2 \mathrm{d}\phi}{4 \int_{\phi=\frac{\phi_S}{2}}^{\infty} |\hat{p}(\phi, \omega_0)|^2 \mathrm{d}\phi}. \tag{3.30}$$

The numerator in (3.30) corresponds to the energy of the spectrum of the PAF at temporal frequency ω_0. The denominator in (3.30) corresponds to the energy contained in the spectral repetitions that will contaminate the reconstruction

[4] If the membrane of the microphone has a finite dimension, a spatial filtering is applied on the recorded sound.

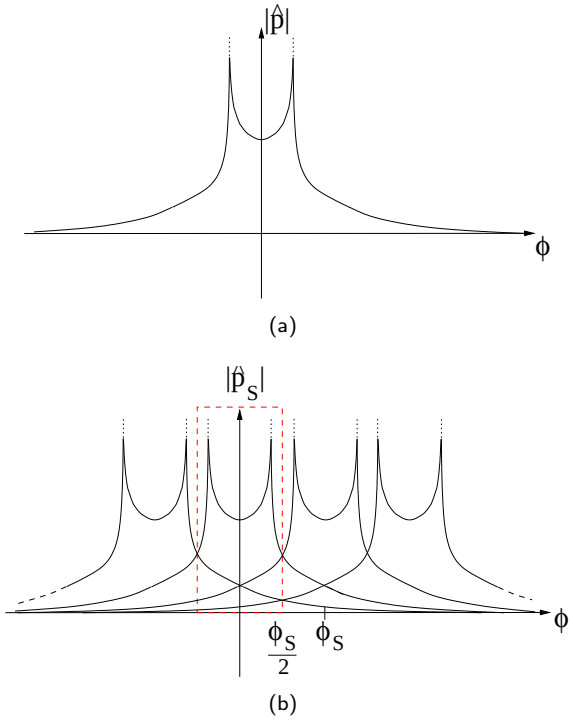


Figure 3.5: Magnitude of the spectrum of the PAF. (a) Cut of the spectrum of the PAF for a particular temporal frequency. (b) For a particular temporal frequency, spectrum of the sampled PAF (denoted as $\hat{p}_S$) consisting of spectral repetitions of the original PAF spectrum.

in the spectral domain of interest. Two different kinds of energy are present in this denominator: the "in band" and the "out of band" energy. The "in band" energy corresponds to the energy of all the spectral repetitions in the domain of interest, namely $[-\frac{\phi_S}{2}, \frac{\phi_S}{2}]$. The "out of band energy" is the energy present in the spectrum that is outside of the domain of interest. It can be shown that the "in band" and the "out of band" energies are equal in the case of an infinite line of microphones. This explains the factor 4 in the denominator of (3.30).

Theorem 3.4.1. Assume one single source emitting in free field at a frequency $\omega = \omega_0$. When sampling the spatio-temporal sound field with a spatial sampling frequency of ϕ_S, and reconstructing it using an ideal interpolator, the SNR of the reconstructed signal in the band $[-\frac{\phi_S}{2}, \frac{\phi_S}{2}]$ can be expressed as

$$\mathrm{SNR}(\phi_S, \omega_0) = \frac{1}{2d \int_{\phi=\frac{\phi_S}{2}}^{\infty} |H_0^{(1)}(d\sqrt{(\frac{\omega_0}{c})^2 - \phi^2})|^2 \mathrm{d}\phi}, \tag{3.31}$$

where $H_0^{(1)}$ is the zero order Hankel function of the first kind.

When considering $\phi_S > \frac{2\omega_0}{c}$ and using the exponential integral function defined as $E_i(x) = \int_x^\infty \frac{e^{-t}}{t}\mathrm{d}t$, the SNR can be lowerbounded:

$$\mathrm{SNR}(\phi_S, \omega_0) \geq \frac{\pi}{4E_i\left(2d\sqrt{\left(\frac{\phi_S}{2}\right)^2 - \left(\frac{\omega_0}{c}\right)^2}\right)}. \tag{3.32}$$

Proof: The numerator in (3.30) can be rewritten using the Parseval's relation as follows:

$$2\pi \int_{x=-\infty}^{\infty} |\tilde{p}(x, \omega_0)|^2 \mathrm{d}x, \tag{3.33}$$

with $\tilde{p}(x, \omega_0)$ the inverse Fourier transform of $\hat{p}(\phi, \omega_0)$ along the spatial axis. We have that $\tilde{p}(x, \omega_0) = \frac{e^{-j\frac{\omega_0\sqrt{x^2+d^2}}{c}}}{4\pi\sqrt{x^2+d^2}}$, and therefore the numerator in (3.30) is

$$\frac{2\pi}{16\pi^2} \int_{x=-\infty}^{\infty} \frac{1}{x^2 + d^2}\mathrm{d}x = \frac{1}{8d}. \tag{3.34}$$

Using (3.7) and (3.34) in (3.30) leads to (3.31).

When considering $\phi_S > \frac{2\omega_0}{c}$, the denominator in (3.30) is

$$\frac{1}{\pi^2} \int_{\phi=\frac{\phi_S}{2}}^{\infty} \left| K_0\left(d\sqrt{\phi^2 - \left(\frac{\omega_0}{c}\right)^2}\right)\right|^2 \mathrm{d}\phi. \tag{3.35}$$

Using (3.11), (3.35) can be upperbounded by:

$$\frac{1}{2\pi} \int_{\phi=\frac{\phi_S}{2}}^{\infty} \frac{e^{-2d\sqrt{\phi^2-(\frac{\omega_0}{c})^2}}}{d\sqrt{\phi^2 - (\frac{\omega_0}{c})^2}}\mathrm{d}\phi. \tag{3.36}$$

Using the change of variable $z = \sqrt{\phi^2 - (\frac{\omega_0}{c})^2}$, (3.36) can be rewritten as:

$$\frac{1}{2\pi d} \int_{z=\sqrt{\left(\frac{\phi_S}{2}\right)^2-\left(\frac{\omega_0}{c}\right)^2}}^{\infty} \frac{e^{-2dz}}{\sqrt{z^2 + (\frac{\omega_0}{c})^2}}\mathrm{d}z. \tag{3.37}$$

Then, (3.37) can be upperbounded by:

$$\frac{1}{2\pi d} \int_{\sqrt{\left(\frac{\phi_S}{2}\right)^2-\left(\frac{\omega_0}{c}\right)^2}}^{\infty} \frac{e^{-2dz}}{z}\mathrm{d}z = \frac{1}{2\pi d} E_i\left(2d\sqrt{\left(\frac{\phi_S}{2}\right)^2 - \left(\frac{\omega_0}{c}\right)^2}\right), \tag{3.38}$$

where E_i represents the exponential integral function. Using (3.38) and (3.34), (3.32) is obtained. ■

As a numerical check of the tightness of the bound, the SNR has been computed

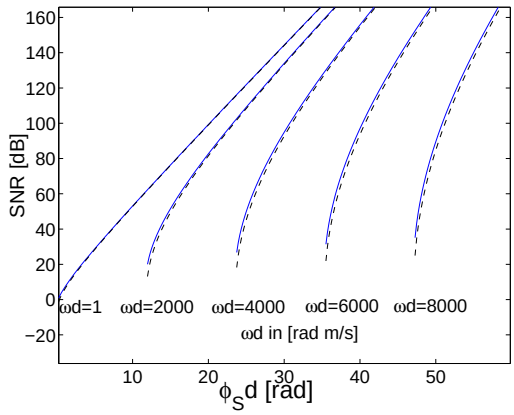


Figure 3.6: In full lines, SNR on the reconstruction signal for different $\phi_S d$ for normalized temporal frequencies ranging from 1 rad m/s to 8000 rad m/s. In dotted line, the corresponding SNR lower bounds are given.

as a function of the spatial sampling frequency and the temporal frequency. To take into account the distance d, it was observed that (3.31) and (3.32) can be expressed as a function of normalized frequencies $\phi_S d$ and ωd. This allows us to obtain a numerical evaluation for these two equations independently of d. The results are shown in Fig. 3.6 in full lines for (3.31) and in dotted lines for (3.32). In order to avoid numerical instability due to the infinite value of the spectrum at the position $\phi_S = \frac{2\omega}{c}$, the simulations start for each ωd at a value of $\phi_S d$ larger than $\phi_S d = \frac{2\omega d}{c}$. At one normalized temporal frequency, it can be observed that the SNR increases for larger normalized spatial sampling frequencies. Note that the lowerbound follows tightly the SNR obtained numerically. A numerical example is given.

Example 3.4.1. As an example, it can be observed from Fig. 3.6 that to reconstruct the sound field up to 8000 rad/s (or 1.3 kHz), a spatial distance of 12.35 cm is necessary to achieve a reconstruction quality of 100 dB when considering a unitary distance between the source and the line of microphones. Remark that when making the far field assumption, using (3.2) would lead to a spacing of 13.35 cm. □

Instead of considering the SNR at a specific temporal frequency, the previous result for the SNR on the reconstruction can be generalized when the signal is in a frequency band $[\omega_1, \omega_2]$ with power spectral density $|\beta(\omega)|^2$. The

SNR on the reconstruction is then given by the following formula:

$$\mathrm{SNR}(\phi_S, [\omega_1, \omega_2]) = \frac{\int_{\omega_1}^{\omega_2} |\beta(\omega)|^2 \mathrm{d}\omega}{2d \int_{\omega_1}^{\omega_2} |\beta(\omega)|^2 \left(\int_{\frac{\phi_S}{2}}^{\infty} |H_0(d\sqrt{(\frac{\omega}{c})^2 - \phi^2})|^2 d\phi \right) \mathrm{d}\omega}. \quad (3.39)$$

When considering $\phi_S > \frac{2\max(\omega_1,\omega_2)}{c}$, (3.39) can be lowerbounded as follows:

$$\mathrm{SNR}(\phi_S, [\omega_1, \omega_2]) \geq \frac{\pi \int_{\omega_1}^{\omega_2} |\beta(\omega)|^2 \mathrm{d}\omega}{4 \int_{\omega_1}^{\omega_2} |\beta(\omega)|^2 E_i \left(2d\sqrt{\left(\frac{\phi_S}{2}\right)^2 - \left(\frac{\omega}{c}\right)^2} \right) \mathrm{d}\omega}. \quad (3.40)$$

Generalization of the sampling theorem for the cases of multiple sources in free field or inside a room is matter of current research.

3.4.2 Plenacoustic sampling

In this section the effect of the sampling in time and space of the PAF is represented in the spectral domain. This representation takes into account the results of the previous section. Sampling in space is done considering a margin taking into account the evanescent decay of the PAF spectrum. The sampling of the PAF will first be shown in details for the rectangular sampling pattern. Further, it will be shown that quincunx sampling can also be used to sample the PAF.

Rectangular sampling

The schematic top view of the spectrum of the PAF is shown in Fig. 3.7(a). When sampling the PAF along the spatial dimension with a spatial sampling frequency of ϕ_S, repetitions of the spectrum occur as shown in Fig. 3.7(b). When one considers the far field approximation of the sound pressure field, the spatial sampling frequency needed to reconstruct the PAF up to a temporal frequency ω_0 is given by

$$\phi_S > \frac{2\omega_0}{c}. \quad (3.41)$$

When one does not consider any far field approximation, Fig. 3.6 has to be considered. In that figure, the spatial sampling frequency is given for each possible temporal frequency for a wanted reconstruction SNR. When one wants to ensure that the highest frequency of the signal is still reconstructed with a given SNR, the spatial sampling frequency needs to be slightly increased

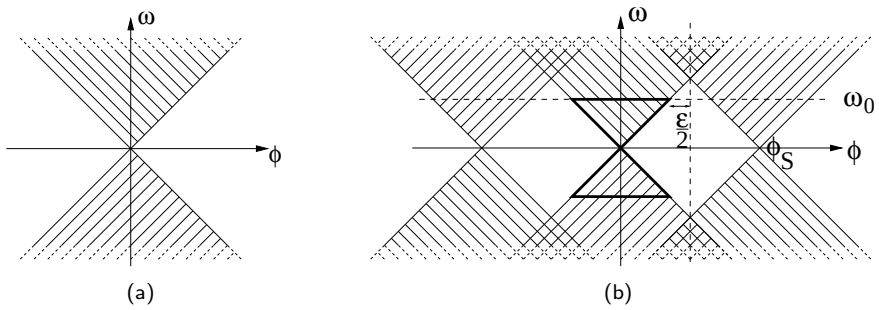


Figure 3.7: PAF spectrum. (a) Top view of the PAF spectrum. (b) Top view of the PAF spectrum with its repetitions due to the spatial sampling.

compared to (3.41). To reconstruct the sound pressure field up to ω_0 with a given SNR_0 at that frequency, the spatial sampling frequency is given by

$$\phi_S > \frac{2\omega_0}{c} + \varepsilon(\mathrm{SNR}_0, \omega_0), \tag{3.42}$$

where $\varepsilon(\mathrm{SNR}, \omega)$ is obtained from Fig. 3.6 when considering a specific temporal frequency and a given SNR. It represents the difference between the value read on the graph and the far field approximation (3.41). Inside the region in bold in Fig. 3.7(b), it can be observed that for a particular ϕ_S chosen according to (3.42):

$$\forall |\omega| \leq \omega_0 \quad \mathrm{SNR}(\phi_S, \omega) \geq \mathrm{SNR}_0. \tag{3.43}$$

When $|\omega| \geq \omega_0$, due to the periodic repetitions along the spatial axis, the $\mathrm{SNR}(\phi_S, \omega)$ decreases.

The temporal sampling of the PAF is now considered. The signal is first bandlimited to ω_0 satisfying (3.42) and sampled with a temporal sampling frequency of $\omega_S = 2\omega_0$. Repetitions of the spectra occur now also along the temporal frequency. The obtained spectrum for the PAF sampled in space and time is shown in Fig. 3.8(a).

Conversely, considering temporal sampling first, one can say that if the maximal temporal frequency present in the signal is ω_0, then sampling the signal at a temporal sampling frequency $\omega_S = 2\omega_0$, one obtains the signal whose spectrum is shown in Fig. 3.8(b). When sampling this signal along the spatial dimension, it is necessary to choose ϕ_S as in (3.42) in order to avoid aliasing as shown in Fig. 3.8(a). Using the rectangular sampling, the sampling of the PAF happens in space and time domain as shown in Fig. 3.9(a). The corresponding spectrum is shown in Fig. 3.9(b). The final expression for our

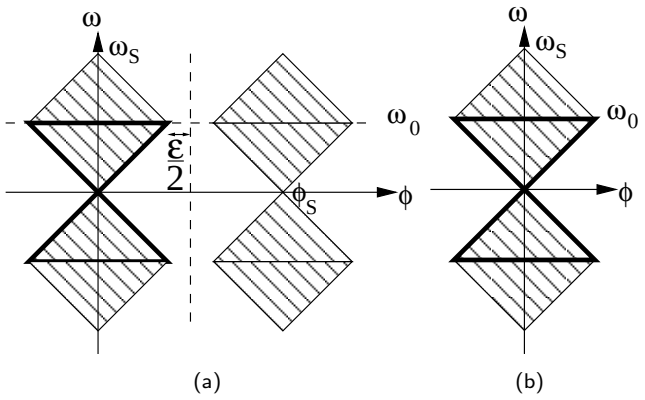


Figure 3.8: PAF spectrum. (a) PAF spectrum with its spectral repetitions along the temporal and the spatial frequencies. (b) PAF spectrum with its spectral repetitions along the temporal frequencies. In both figures, the region in bold corresponds to the original spectrum bandlimited along the temporal frequency without spectral repetitions.

sampled PAF 2D spectrum (denoted as $\hat{p}_S$) becomes:

$$\hat{p}_S(\phi,\omega) = \frac{1}{\Delta x \Delta t} \sum_{k_1,k_2=-\infty}^{\infty} \hat{p}\left(\phi - \frac{2\pi k_1}{\Delta x}, \omega - \frac{2\pi k_2}{\Delta t}\right). \tag{3.44}$$

Quincunx sampling

A tighter packing of the spectrum can be achieved by using quincunx sampling. In time domain, the grid to be used is shown in Fig. 3.9(c). In the corresponding spectrum, the spectral repetitions are placed such that they fill better the whole frequency space as shown in Fig. 3.9(d). In the quincunx sampling the spatial sampling frequency is now only $\frac{\phi_S}{2}$. This corresponds to a distance between two samples on the space axis of $2\Delta x = \frac{4\pi}{\phi_S}$. This shows that using quincunx sampling one only needs to sample the even microphones at even times while the odd microphones are sampled at odd times. This leads to a gain of factor two in the processing. However it does not reduce the number of necessary microphones. Similar approaches have been used in [36] in the study of the far field electromagnetic field. The precise value of ϕ_S to be used in the case of quincunx sampling to ensure a SNR of reconstruction at a certain temporal frequency is not given in this work but could be derived similarly to the expression (3.31).

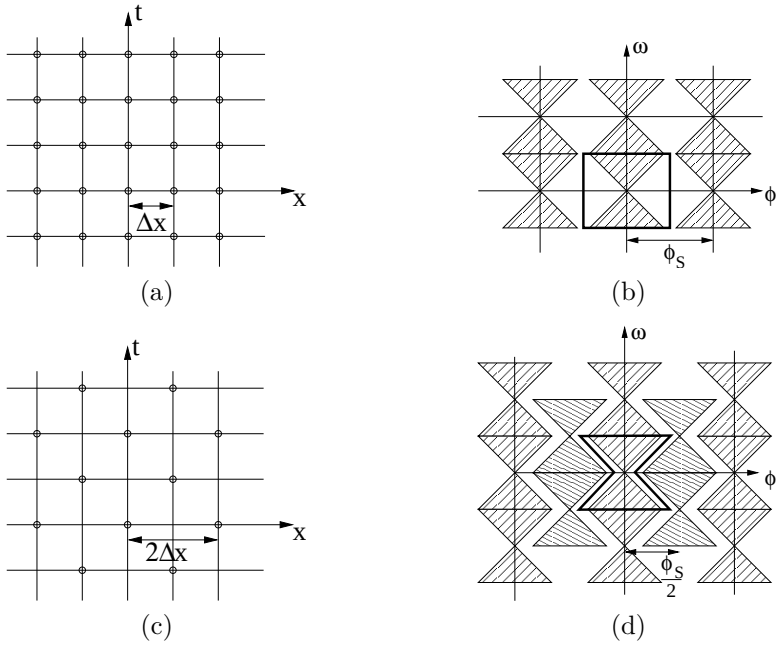


Figure 3.9: Sampling and interpolation of the PAF. (a) Rectangular sampling grid. (b) Plenacoustic spectrum with its repetitions for a rectangular sampling grid with the interpolation filter in bold. (c) Quincunx sampling grid. (d) Plenacoustic spectrum with its repetitions for a quincunx sampling grid with the interpolation filter in bold.

3.4.3 Reconstruction by interpolation

Knowing the sound field at every point of the sampling grid, the usual interpolation techniques [111, 108] are applied in order to reconstruct the sound pressure field at any location. The interpolation filter to be used is dependent on the sampling grid, and may be separable in time and space. When the samples have been obtained by rectangular sampling, the interpolation filter is a low-pass filter with support $[-\frac{\phi_S}{2}, \frac{\phi_S}{2}]$. The support of the filter is shown in bold in Fig. 3.9(b). When the samples have been obtained by quincunx sampling, the filter used for interpolation is a "fat fan filter" [108]. The support of the filter is shown in bold in Fig. 3.9(d). One can compute the ideal filters in closed forms, however these ideal interpolation filters are not realizable in practice. Therefore, one needs to slightly increase the spatial sampling frequency to take into account the decay of the real filter. For their design, standard techniques can be used [108].

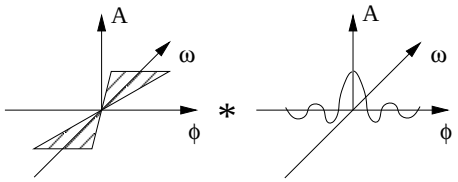


Figure 3.10: Effect of the windowing due to the finite length of the array.

3.5 Finite Length Aperture

In this section, the field is not measured along an infinite line but on a finite interval inside the room. This can be seen as a windowing of the PAF in the spatial domain. Consider a rectangular window $w(x)$. In the present case, the window is simply a function of the spatial position. Calling the windowed PAF $q(x,t)$. It satisfies $q(x,t) = p(x,t)w(x)$. In frequency domain this is written as

$$\hat{q}(\phi,\omega) = \hat{p}(\phi,\omega) * \hat{w}(\phi,\omega) = \hat{p}(\phi,\omega) * (\hat{w}(\phi)\delta(\omega)) \,. \tag{3.45}$$

The situation is schematically shown in Fig. 3.10.

The size of the aperture has an influence on the decay of the windowed PAF. Taking measurements from positions between $-\frac{L}{2}$ and $\frac{L}{2}$ leads to a convolution of the PAF spectrum with the following sinc function:

$$\hat{w}(\phi) = \int_{x=-\frac{L}{2}}^{\frac{L}{2}} e^{-j\phi x} \mathrm{d}x = L\mathrm{sinc}\left(\frac{\phi L}{2\pi}\right). \tag{3.46}$$

At a given ϕ, the larger the value of L, the faster the decay will be. This fact can be observed in Fig. 3.11. A section of the 2D spectrum of the PAF at a particular temporal frequency (2000π rad/s) is presented. One can observe that for larger aperture sizes the spectrum decays more rapidly as given in (3.46).

The decay of the spectrum of the PAF along the spatial frequency will now be slower than the one in (3.9). For a particular temporal frequency $\omega = \omega_0$, the decay is:

$$\hat{q}(\phi,\omega_0) = \frac{e^{-j\phi x_s}}{2\pi} K_0\left(d\sqrt{\phi^2 - \left(\frac{\omega_0}{c}\right)^2}\right) * \hat{w}(\phi). \tag{3.47}$$

In the case of a rectangular window, this decay is the convolution of a sinc with a modified Bessel function K_0.

Combining the finite aperture effect with the sampling of the PAF, the following expression for the 2D-FT of the sampled windowed PAF (denoted as

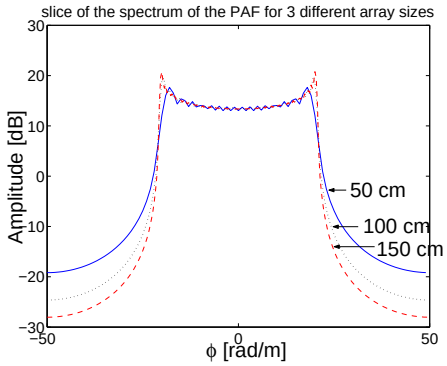


Figure 3.11: A section of the PAF at a particular temporal frequency (2000π rad/s). The curves represent data acquired on intervals of different lengths: 50 cm (full line), 100 cm (dotted line) and 150 cm (dashed line). A larger interval leads to a faster decay. For this graph, we used $d = 1$ in (3.7).

$\hat{q}_S$) is obtained:

$$\hat{q}_S(\phi,\omega) = \frac{1}{\Delta x \Delta t} \sum_{k_1,k_2=-\infty}^{\infty} \hat{q}\left(\phi - \frac{2\pi k_1}{\Delta x}, \omega - \frac{2\pi k_2}{\Delta t}\right). \tag{3.48}$$

3.6 Simulations and Measurements

In this section, simulation results are presented for the interpolation of RIRs. These results are then compared with real measurements.

3.6.1 Simulation results

RIRs have been simulated on a line in a room using the image source model. For simulation purposes, one derives a dense set of impulse responses, keeps a subset, and interpolates the missing ones. In the simulations, the case of rectangular sampling of the sound field was considered. To compare the interpolated with the simulated RIRs, the normalized mean square error (MSE) criterion was used:

$$\text{MSE} = 10\log_{10} \frac{\sum_{i=1}^{T} \left(r[i] - r_e[i]\right)^2}{\sum_{i=1}^{T} r^2[i]}, \tag{3.49}$$

with T the length in samples of the simulated RIRs, r the simulated RIR and r_e the interpolated RIR.

In the presented simulations, $2n+1$ RIRs were simulated every cm along a line in the room. From these simulations, we kept $n+1$ measurements spaced with 2 cm to interpolate the n "in-between" positions. With the spacing of

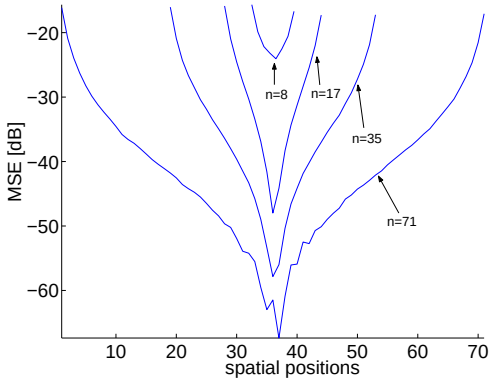


Figure 3.12: Interpolation error for different array sizes. We use the same spacing between the microphones, namely 2 cm.

2 cm, the RIRs were reconstructed up to a temporal frequency of 16000π rad/s. In this setup, due to the limited length of the array, the decay of the spectrum is mostly determined by the decay of the windowing function as discussed in Section 3.5. We chose ε large enough in (3.42) to ensure a sufficient decay in the spectrum of the finite length aperture PAF. Note that this ε can be substantially larger than the one obtained in Fig. 3.6. The MSE for those n positions is shown in Fig. 3.12. Different lengths of the array centered around the same spatial position were considered, namely $n = 8$, 17, 35 and 71. Remark that the MSE is only given for the interpolated positions since the positions used to apply the interpolation are perfectly reconstructed. One can observe that using the array corresponding to $n = 71$ leads to a very small error (less than -60 dB) for the interpolation in the middle of the array. When using the same spacing between the microphones but reducing the number of RIRs, the interpolation error increases due to the border effects introduced by the finite length of the array. Remark also that the curves on Fig. 3.12 are not symmetric. This is due to the fact that the microphone array was not symmetrically located inside the room.

3.6.2 Experimental results

Experimental results were carried out in a partially sound insulated room with RIRs measured at different spatial positions. One loudspeaker (Genelec 1029A) was used together with a microphone array (composed of 8 Panasonic $WM61A$). A frequency logarithmic sweep [86] was used to measure the RIRs. 71 RIRs were measured with a microphone spacing of 2 cm along a line in the room. The spectrum of the measured PAF is shown in Fig. 3.13(a). Similarly to the technique presented in the simulations, we kept 36 measurements spaced

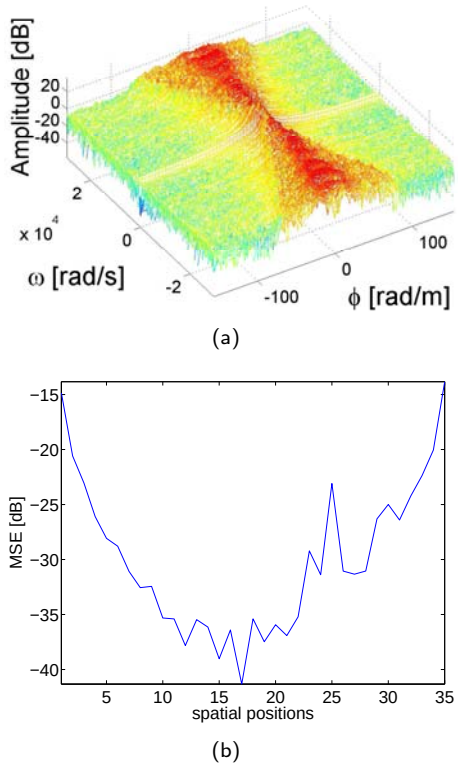


Figure 3.13: Experimental PAF. (a) 2D-FT of the measured plenacoustic function. (b) Interpolation error on measured RIRs.

with 4 cm to interpolate the 35 "in-between" positions. The RIRs were low-pass filtered to 8000π rad/s. Considering (3.42), ε was chosen large enough to ensure a sufficient decay in the spectrum of the windowed PAF. The MSE on the 35 interpolated RIRs measurements is shown in Fig. 3.13(b). Note that the MSE shown in Fig. 3.13(b) is obtained when using only the first 100 ms of the RIRs. When considering the full RIRs (1 s) poorer results are obtained (on the order of -25 dB). This results leads to the thinking that only the beginning of the RIRs is well interpolated. To justify this conjecture, successive measurements were performed at the same spatial positions with a 30 s interval. We studied the MSE between two successive measurements using a sliding window of 25 ms. The results were averaged over 100 measurements. It can be observed in Fig. 3.14 that the MSE is very low at the beginning of the RIR and increases with time. After 100 ms, the MSE between consecutive measurements is already of about -38 dB. This fact can probably be understood by a variation of the speed of sound propagation over time. This effect is the most severe for

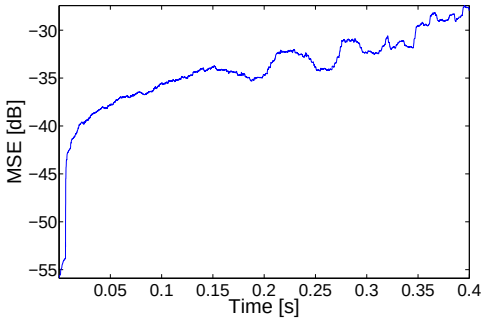


Figure 3.14: MSE on successive RIRs along a sliding window of 25 ms. The results were averaged over 100 pairs of successive measurements.

waves traveling over a wider area. Therefore, the reverberant part of the RIR undergoes the largest relative timing changes. The speed of sound fluctuation also has a larger impact for larger microphone to loudspeaker distances. This speed of sound variation can be the consequence of variation of different parameters such as temperature, humidity and pressure in the room. In [45], similar observations have been done and attributed to variation of temperature. They showed that a variation of 0.1° can create a misalignment between RIRs of more that 25 dB.

Remark finally that due to the limit of our eight inputs microphone array, we had to move the array to the next positions in order to measure the 71 RIRs (our intrusion in the room probably modified the temperature and other factors between two sets of measurements). Better results would be obtained if all the measurements could be captured simultaneously, which was not possible due to hardware limitations.

For completeness, a set of electromagnetic (EM) channel impulse responses (CIRs) is also presented in this chapter. Measurements of EM CIRs have been done in a room along a line at 78 positions spaced by 2 cm. The excitation signal was produced by an pseudonoise generator (Centellax $TG1P1A$) containing energy in a range between 2 GHz to 5 GHz. The impulse responses were measured by antennas using an oscilloscope Lecroy $SDA - 6000$. The 2D spectrum of the CIRs is shown in Fig. 3.15. This spectrum also exhibits a butterfly shape. Nevertheless, as the excitation signal contained small energy for frequencies below 2 GHz and the used antennas only picked up signal of higher frequencies, the butterfly shape is only visible above that frequency, what explains the trapezoidal shape of the spectrum. Note that in the electromagnetic case, the support of the butterfly is now dependent on the speed of light propagation.

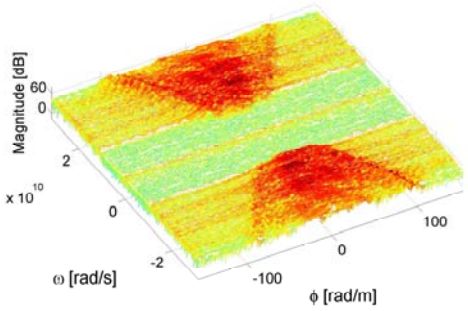


Figure 3.15: 2D spectrum of measured electromagnetic channel impulse responses.

3.7 Plenacoustic Function for Multidimensional Spatial Positions

The previous sections were devoted to the detailed study of the sampling and interpolation of the sound pressure field on the line. This study can obviously be generalized to other spatial positions of microphones (or loudspeakers). Therefore, one wants to study the shape of spectra associated with different microphones setups. Section 3.7.1 studies the spectrum of the PAF associated to a plane of microphones. The optimal sampling pattern for positioning the microphones is studied. Further, the three-dimensional space filled of microphones is presented in Section 3.7.2.

3.7.1 Plenacoustic function on a plane

Study of the spectrum

In Section 3.3, the shape and the properties of the PAF on a line in the room were studied. In this section, one considers a more general case where the RIRs are studied on a plane. Consider a plane in the space filled with receivers in the x and the y directions. Further, a source is located at position (x_s, y_s, z_s). At any receiver position (x_m, y_m, z_m) the direct path coming from the source is

$$p(x_m, y_m, t) = \frac{\delta(t - \frac{a}{c})}{4\pi a}, \tag{3.50}$$

with $a = \sqrt{(x_m - x_s)^2 + (y_m - y_s)^2 + (z_m - z_s)^2}$. The derivation of the 3D-FT of (3.50) is done in Appendix A.2 and follows the same approach as the

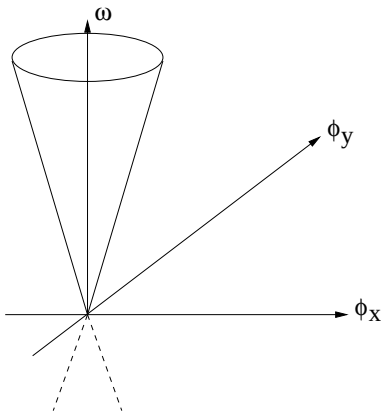


Figure 3.16: Schematic view of the three-dimensional spectrum of the PAF.

one presented in [18]. The result is:

$$\hat{p}(\phi_x, \phi_y, \omega) = \begin{cases} \frac{-j\Phi}{2} \frac{e^{-j|z|\sqrt{\left(\frac{\omega}{c}\right)^2 - \phi_q^2}}}{\sqrt{\left(\frac{\omega}{c}\right)^2 - \phi_q^2}} & \text{for} \quad |\phi_q| \leq \frac{\omega}{c} \\ \frac{\Phi}{2} \frac{e^{-|z|\sqrt{\phi_q^2 - \left(\frac{\omega}{c}\right)^2}}}{\sqrt{\phi_q^2 - \left(\frac{\omega}{c}\right)^2}} & \text{for} \quad \frac{\omega}{c} \leq |\phi_q|. \end{cases} \tag{3.51}$$

with $\phi_q^2 = \phi_x^2 + \phi_y^2$ and $\Phi = e^{-j(\phi_x x_s + \phi_y y_s)}$. Note that ϕ_x and ϕ_y represent the spatial frequencies for the microphones in the x and y directions respectively. The obtained spectrum has a conical shape as shown in Fig. 3.16. The surface of the cone is given by:

$$\frac{\omega}{c} = \sqrt{\phi_x^2 + \phi_y^2}. \tag{3.52}$$

Similarly to the results obtained with the line of microphones, one sees that the decay of the spectrum is also faster than an exponential outside of the conical shape. Remark that in the specific case of the source located on the plane of the microphones, the decay becomes slower and is, up to a constant, asymptotic as $\frac{1}{\phi_q}$.

Optimal sampling pattern

Similarly to the analysis presented in Section 3.4.2, the optimal sampling pattern for the positioning of the microphones on the plane is studied. The first approach is to use the rectangular sampling as shown in Fig. 3.17(a). A spacing of Δx_1 and Δy_1 was used for the spacing between the microphones in the x and y directions. Fig. 3.17(b) shows the corresponding packing of the circles in the Fourier spectrum for one temporal frequency (typically the highest frequency

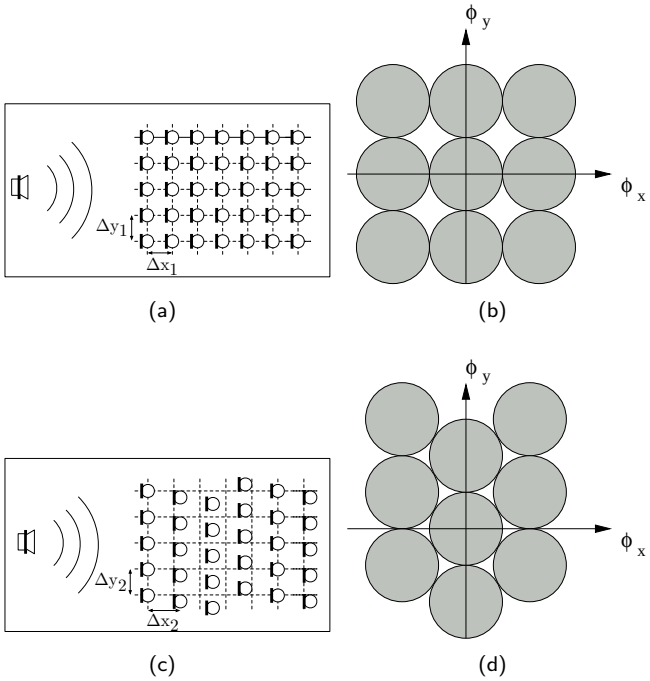


Figure 3.17: Sampling of the PAF on a plane. (a) Placement of the microphones on the plane on a rectangular sampling grid. (b) Plenacoustic spectrum with its repetitions for a rectangular sampling grid. (c) Hexagonal sampling grid. (d) Plenacoustic spectrum with its repetitions for a hexagonal sampling grid.

present in the emitted signal).

The conical shape of the spectrum allows us to obtain a tighter packing of the circles. The use of an hexagonal sampling pattern leads to a reduction of about 15% in the number of necessary microphones. Fig. 3.17(c) shows the new positions of the microphones on the plane. In this case,

$$\Delta x_2 = \frac{2}{\sqrt{3}}\Delta x_1\,,\ \Delta y_2 = \Delta y_1. \tag{3.53}$$

Fig. 3.17(d) shows the corresponding spectrum with its spectral repetitions. Other packings of the cones can be realized to lower the temporal sampling frequency of the A/D converters but do not reduce further the number of microphones needed to sample the sound pressure field on a plane [37].

Simulation results

We simulated RIRs on a plane in a room using the image source model. The 3D Fourier transform was applied on the gathered data. By looking at sections of

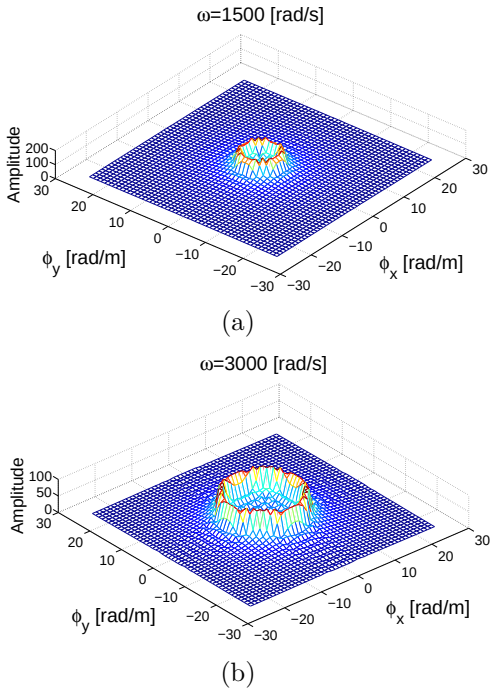


Figure 3.18: Spectrum of the PAF obtained by simulations at different temporal frequencies. (a) $\omega = 1500$ rad/s. (b) $\omega = 3000$ rad/s.

this spectrum for $\omega = 1500$ rad/s and $\omega = 3000$ rad/s, one obtains Fig. 3.18(a) and (b), respectively. One can see that with growing temporal frequencies, the support of the PAF spectrum also increases as given by (3.52).

3.7.2 Plenacoustic function in space

In this section microphones located in the 3D space are considered. Similarly to the setup presented in Section 3.7.1, a source is located at position (x_s, y_s, z_s) and microphones at positions (x_m, y_m, z_m). The PAF is also given by (3.50). Note that in the present setup also z_m is a variable. Call ϕ_z the spatial frequency of the microphone positions in the z direction. Calculating the 4D-FT of (3.50) is done in Appendix A.3. The result is:

$$\hat{p}(\phi_x, \phi_y, \phi_z, \omega) = \frac{e^{-j(\phi_x x_s + \phi_y y_s + \phi_z z_s)}}{\phi_x^2 + \phi_y^2 + \phi_z^2 - (\frac{\omega}{c})^2}, \tag{3.54}$$

which represents a cone in four dimensions. For a particular temporal frequency, the section of this cone is a sphere. The size of the sphere obeys the

following rule:

$$\frac{\omega}{c} = \sqrt{\phi_x^2 + \phi_y^2 + \phi_z^2}. \tag{3.55}$$

Remark that the decay outside of the sphere is not exponential as it was for the plane and line of microphones. This is due to the presence of the source at one of the microphone positions. At a particular temporal frequency, the optimal packing of the spheres is given by face-centered cubic lattice packing [106]. It reduces the number of samples by a factor of about 29.3% when compared to a normal rectangular sampling pattern.

3.8 Conclusion

In this chapter, we have introduced and studied the plenacoustic function. It characterizes the sound pressure field at any point in space. This function has been studied and its spectrum for the linear, the planar and the three-dimensional case has been calculated without making any far field assumption. The decay of the spectrum has been studied along both the temporal and spatial frequency axis. Based on the support of the spectrum, the number and the spacing between the microphones needed to reconstruct the sound pressure field up to a certain temporal frequency has been determined. A quantitative sampling theorem, trading off sampling rate for SNR, has been presented. As an example, it has been shown that to reconstruct the sound field up to 1.3 kHz, a spatial distance of 12.35 cm is necessary to achieve a reconstruction quality of 100 dB for a source located at unitary distance from the line of microphones. Using the far field assumption, this spacing would be of 13.35 cm. The optimal sampling pattern for the microphone positions has also been given for the linear, the planar and the three-dimensional case. Finally, simulations and experimental results were presented and compared with the theoretical results.

Chapter 4

Sound Field Analysis Along a Circle and its Application to HRTF Interpolation

4.1 Introduction

In this chapter, we explore the spatial evolution of the sound field along a circle. We study how many sensors need to be placed on a circular array to be able to recover the field at any position along the circle. This leads to an angular sampling theorem of the sound field along a circle. This theorem describes quantitatively the trade-off existing between the quality of reconstruction of the sound field and the angular spacing between the sensors. Further, as application of the presented theory, sampling and interpolation of head-related transfer functions (HRTFs) is described. HRTF sampling is being looked at in terms of sound field sampling. Considering the dual problem of a source at the ear entrance it is shown that the angular bandwidth on circles around a listener is limited. Consequently, angular interpolation can be applied to obtain highly precise interpolated HRTFs. It is shown that the necessary spacing between the HRTFs is quite small. Thus, this technique is only applicable if finely spaced HRTF measurements are available. For example, at a sampling rate of 44.1 kHz, HRTFs need to be available every 4.9° to avoid spatial aliasing.

While the sound field based HRTF interpolation method enables very interpolation in a least mean square sense (much more precise than other methods), its drawback is that it looses its advantage when applied to more coarsely spaced HRTFs than would be dictated by the angular Nyquist theorem. The theorem indeed often dictates a very dense angular sampling which makes the

method difficult to apply in some practical situations.

To counteract this limitation of this first interpolation scheme, we are proposing an HRTF interpolation algorithm which attempts to take advantage of spatial bandwidth considerations while improving its performance for the case when less HRTF measurements are present than required by the angular Nyquist theorem. In the low frequency band, where very little spatial aliasing occurs, the proposed algorithm applies an angular interpolation method similar to the previously proposed method. At higher frequencies, the interpolation is carried out in subbands. Angular aliasing is avoided by applying angular interpolation to the complex temporal envelope in each subband. Given the interpolated complex envelopes the subband signals are re-generated by modulating them with a separately interpolated carrier signal. The newly presented technique is then compared with different existing techniques. It is shown that the new method still achieves good interpolation results in case of angular undersampling by a factor of two to four.

4.1.1 Related work

Circular microphone or loudspeaker arrays, discussed in detail in this chapter, are used in different techniques existing in the literature such as beamforming [14, 63, 61], wave field analysis and synthesis [90, 92] and HRTF interpolation [20]. Some related work is now described.

Related works exist on the topic of sound field sampling but have mostly considered the plane wave assumption [65, 64, 103]. In [92], Poletti has quantified the aliasing, noise and transducer variability of circular microphone arrays under the far field assumption. Note that there exists a major difference between the plane wave studied along a line or a circle. The linear array is the natural way of analyzing the plane wave since one plane wave at one temporal frequency corresponds exactly to one point in the spectrum [7, 91]. In other words, one plane wave corresponds to one basis function of the Fourier decomposition. As was previously described in the introduction of Chapter 3, when considering a plane wave along a line, the sound field can be shown to have a compact support in the frequency domain and if the Nyquist sampling condition is fulfilled, perfect reconstruction can be achieved. When the same plane wave is studied along a circle, perfect reconstruction is not possible anymore. It can be shown that the decomposition of a plane wave in cylindrical coordinates excites all possible modes [65]. Therefore, theoretically, sensors should be placed infinitely close to each other to be able to perfectly reconstruct the sound field. In practice, by neglecting the modes containing little energy, a very good reconstruction can still be achieved. Jones et al. [64] have derived bounds

on the error made by reconstructing the field when only a limited number of those modes are considered.

Substantial work has also been done in the field of wave field synthesis (WFS) using circular arrays and many existing systems make use of circular loudspeaker arrays [90]. Recently Spors et al. have investigated the aliasing artifacts present when reproducing the sound field with WFS using linear and circular arrays of loudspeakers [104]. Also, wave field extrapolation algorithms can be adapted to describe the case of circular microphone arrays [59,60,61]. In [60], the auralization of the sound field is described. The field is measured along a circular array and further extrapolated at other positions. Different possible geometries are described and compared for this purpose and it is concluded that a circular array of microphones leads to the best extrapolation results. The description of HRTFs as well as a different aspects relative to their measurement and interpolation are described in Section 2.3.3.

4.1.2 Contributions

The contributions presented in this chapter are the following:

- A thorough analysis of the sampling of the two-dimensional sound field is described together with an angular Nyquist theorem for circular arrays.
- Approximation formula are given for the essential support of spectrum of the three-dimensional sound field gathered on a circle.
- Best interpolation results for HRTFs in the horizontal plane are obtained when sampled with an angular spacing of 5° for reconstruction up to 44.1 kHz.
- A novel technique is presented for undersampled HRTF datasets where the HRTFs are decomposed into their envelope and carrier signals. This method achieves acceptable results in mean square error sense for HRTFs measured with an angular undersampling factor of two or four.

Outline

The chapter is organized as follows. Section 4.2 studies the bandwidth limit of the sound field on a circle. It first considers a two-dimensional setup in Section 4.2.1 where an angular sampling theorem is developed and results are generalized for the three-dimensional case in Section 4.2.2. Given this result, sampling and interpolation of HRTFs are considered in Section 4.2.3. A new technique, presented in Section 4.3, is described in order to achieve good results in the case HRTFs are sampled too coarsely along the angular dimension. It is

based on the decomposition of HRTFs in their complex envelopes and carrier signals. Numerical simulations and experiments are presented in Section 4.4 together with a comparison with other existing methods. Conclusions are given in Section 4.5.

4.2 Sound Field Analysis Along a Circle

In this section, the angular bandwidth of the sound field along a circle is analyzed. In a first step, the study is performed in two dimensions in Section 4.2.1. A sampling theorem is presented to describe quantitatively the quality of reconstruction as a function of the angular sampling frequency. The three-dimensional (3D) case is then further discussed in Section 4.2.2. Based on these results, sampling of HRTFs is discussed in Section 4.2.3 together with a first technique for their interpolation.

4.2.1 Angular bandwidth of the 2D sound field on a circle

In this section, two setups are explored. The first setup considers a plane wave arriving on the circle of microphones. The second setup considers a circular wave source. For the second setup, a sampling theorem is presented.

Plane wave assumption

The first setup considered in this section consists of a circular microphone array of radius r. The coordinates of the different microphones are $\mathbf{r} = (r_x, r_y)$, with $r_x = r\cos\theta$ and $r_y = r\sin\theta$. A plane wave emitted by a plane source is incoming on the circle of microphone with a given angle θ_s as is shown in Fig. 4.1(a).

The plenacoustic function (PAF) along the circle is now defined as $p(\theta, t)$, where the sound field recorded at each angle θ is described as function of the time t. The sound field along the circle for a far field source emitting a Dirac pulse is given by

$$p(\theta, t) = \delta\left(t - t_s - \frac{r\cos(\theta - \theta_s)}{c}\right), \tag{4.1}$$

with c the speed of sound propagation and t_s the time taken by the wave to travel from the source position to the center of the microphone array. For the simplicity of the calculations, we consider $t_s = 0$, meaning that for $t = 0$, the wave should pass through the center of the circular array. The PAF along a circle is shown in Fig. 4.1(b). Taking the Fourier transform of (4.1) along the

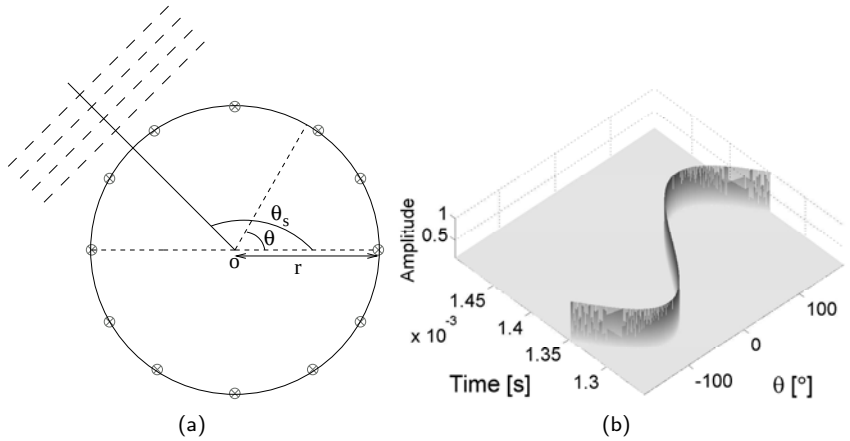


Figure 4.1: Sound field analysis along a circle. (a) A plane wave is incoming on a circle of radius r with an angle θ_s. (b) Plenacoustic function studied along this circle.

temporal axis leads to

$$\tilde{p}(\theta, \omega) = e^{-j\frac{\omega}{c} r \cos(\theta - \theta_s)}. \tag{4.2}$$

In order to study the spatial bandwidth of the sound field recorded on a circle, the Fourier transform of (4.2) along the angular direction needs to be calculated. Denote the angular frequency as l_θ. Remark that $l_\theta \in \mathbb{Z}$ due to 2π periodicity of $\tilde{p}(\theta, \omega)$ in θ. We therefore have that

$$\hat{p}(l_\theta, \omega) = \frac{1}{2\pi} \int_0^{2\pi} \tilde{p}(\theta, \omega) e^{-jl_\theta \theta} \mathrm{d}\theta, \tag{4.3}$$

and

$$\tilde{p}(\theta, \omega) = \sum_{l_\theta = -\infty}^{\infty} \hat{p}(l_\theta, \omega) e^{jl_\theta \theta}. \tag{4.4}$$

To calculate $\hat{p}(l_\theta, \omega)$ in the present case, we make use of the Anger-Jacobi expansion [38, 64, 103]. It describes a plane wave as a function of a sum of Bessel functions:

$$e^{-j\frac{\omega}{c} r \cos(\theta - \theta_s)} = \sum_{l_\theta = -\infty}^{\infty} j^{l_\theta} J_{l_\theta}\left(\frac{\omega}{c} r\right) e^{jl_\theta(\theta - \theta_s)}. \tag{4.5}$$

Therefore,

$$\hat{p}(l_\theta, \omega) = j^{l_\theta} J_{l_\theta}\left(\frac{\omega}{c} r\right) e^{-j l_\theta \theta_s}. \tag{4.6}$$

The magnitude spectrum $|\hat{p}(l_\theta, \omega)|$ is plotted in Fig. 4.2 for a circle of radius 0.6 m. Note that most of the energy is present in a butterfly shaped region. Outside of this region, the energy present is decaying very fast. The intuition behind the shape of the spectrum is similar to the one presented in the case of the infinite line of microphones in Chapter 3. For low temporal frequency, due to the large wavelength, the angular frequency support is small. For increasing temporal frequency the butterfly is widening due to larger angular frequency support. This support increases due to the smaller wavelength indicating faster changes along the angular dimension. Similarly to the results of Chapter 3, the support of (4.6) can be approximated to the region satisfying

$$|l_\theta| \leq \frac{|\omega|}{c} r. \tag{4.7}$$

For a given temporal frequency, the spectrum behaves as (for large l_θ)

$$|\hat{p}(l_\theta, \omega)| \sim \frac{1}{\Gamma(l_\theta + 1)} \left(\frac{\omega}{2c} r\right)^{l_\theta}, \tag{4.8}$$

where $\Gamma(l_\theta) = (l_\theta - 1)!$. Using Stirling's approximation which represents an asymptotic formula for the Γ function, it can be shown that for large l_θ and a finite ω, the magnitude spectrum decays as

$$|\hat{p}(l_\theta, \omega)| \sim \frac{1}{\sqrt{2\pi l_\theta}} \left(\frac{re\omega}{2cl_\theta}\right)^{l_\theta} \tag{4.9}$$

$$\sim \frac{c_1}{\sqrt{l_\theta}} \left(\frac{c_2}{l_\theta}\right)^{l_\theta}, \tag{4.10}$$

with c_1 and c_2 two constants. This very fast decay can be observed in Fig. 4.2.

Circular source

The second setup in this section explores the more realistic case, where the sound source is not considered as infinitely far away but at a finite distance. The sound source is located at a distance s from the center of the array and its coordinates are $\mathbf{s} = (s_x, s_y)$ with $s_x = s\cos\theta_s$ and $s_y = s\sin\theta_s$. The setup is shown in Fig. 4.3. The same analysis is valid for distances s larger or smaller than the radius of the circle as will be shown in the sequel. Consider the source to emit sound in free field. The Green's function for the case of a circular sound

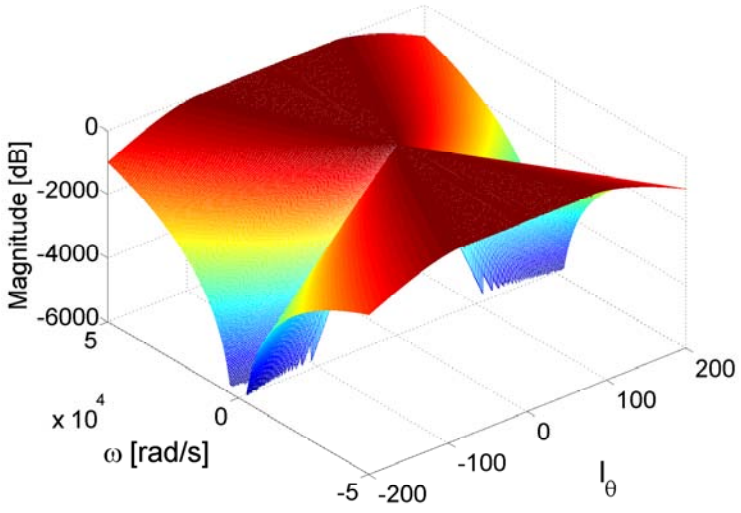


Figure 4.2: Butterfly spectrum of the PAF along a circle in far field.

source in two dimensions is given by [85, 112] and was derived in Chapter 2

$$\tilde{p}(x, y, \omega) = \frac{-j}{4} H_0^{(2)} \left(\frac{\omega}{c} \sqrt{(x - s_x)^2 + (y - s_y)^2} \right), \tag{4.11}$$

with $H_0^{(2)}$ the order zero Hankel function of the second kind. In our case, we are interested in the field studied along a circular array. The time taken by sound to travel between the source and the different receivers located on the circle is

$$h(\theta) = \frac{\sqrt{(s_x - r\cos\theta)^2 + (s_y - r\sin\theta)^2}}{c}. \tag{4.12}$$

With (4.12), the Green's function can be rewritten as

$$\tilde{p}(\theta, \omega) = \frac{-j}{4} H_0^{(2)} \left(\omega h(\theta) \right) = \frac{-j}{4} H_0^{(2)} \left(\frac{\omega}{c} \sqrt{r^2 + s^2 - 2rs\cos(\theta - \theta_s)} \right). \tag{4.13}$$

Note that for large argument of the Hankel function of order zero, we have that [112]

$$H_0^{(2)} \left(\frac{\omega}{c} r \right) \sim \sqrt{\frac{2c}{\pi\omega r}} e^{-i\left(\frac{\omega}{c} r - \frac{\pi}{4}\right)}. \tag{4.14}$$

Similarly to the Anger-Jacobi expansion described in (4.5), it is also possible to express a circular wave as a sum of Bessel functions. This is known as the

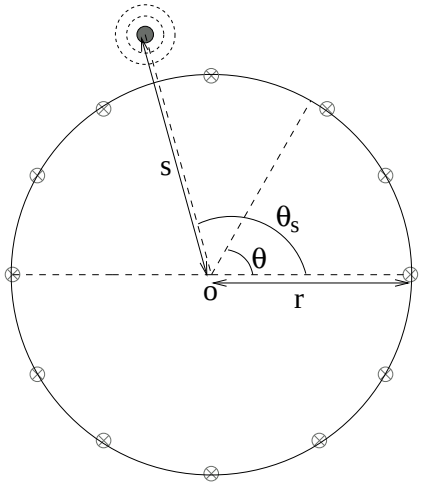


Figure 4.3: A monopole source emits sound in free field. The sound field is recorded along a circle of radius r with equally spaced microphones. The source is located at a finite distance s from the circle.

addition theorem [112, 38]:

$$H_0^{(2)}\left(\frac{\omega}{c}\|\mathbf{s}-\mathbf{r}\|\right) = \sum_{n=-\infty}^{\infty} J_n\left(\frac{\omega}{c}r_<\right) H_n^{(2)}\left(\frac{\omega}{c}r_>\right) e^{jn(\theta-\theta_s)}, \tag{4.15}$$

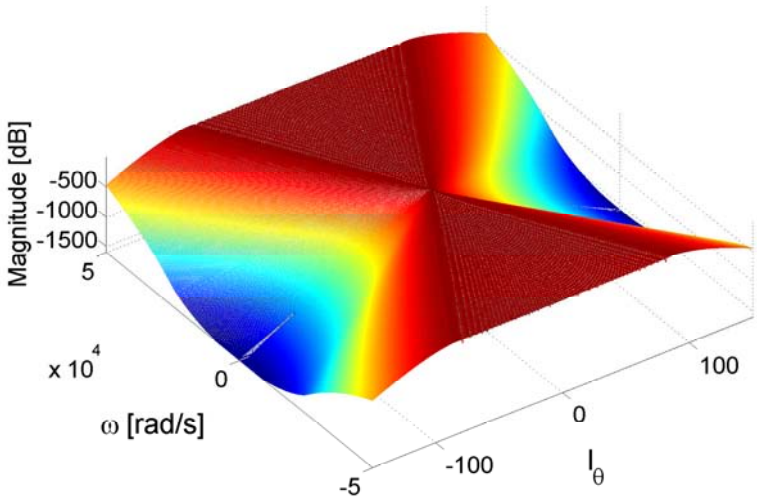
with $\|\mathbf{s}-\mathbf{r}\|$ describing the distance between the source and the receivers, $r_<$ the lesser of r and s and $r_>$ the greater. In this section, we consider $s \geq r$. The exact same results are obtained for the case where $s \leq r$, only s and r need to be swapped in the equations. It will be shown to be the case in the study of the angular bandwidth of HRTFs in Section 4.2.3.

Comparing the Fourier series in (4.15) and in (4.4) using (4.13) leads to:

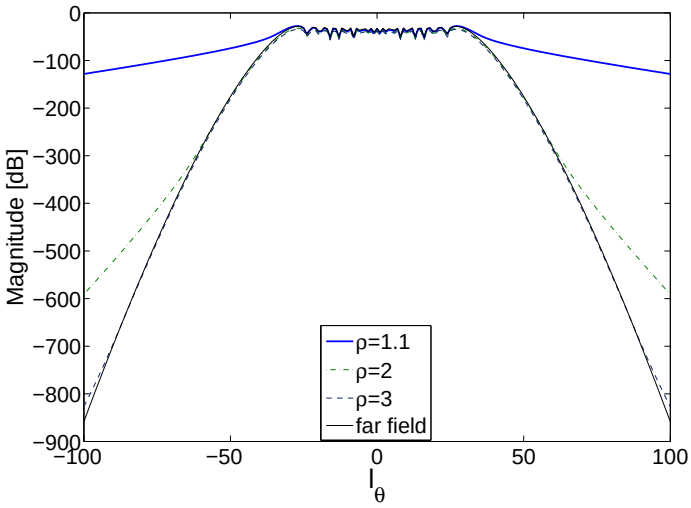
$$\hat{p}(l_\theta, \omega) = \frac{-j}{4} J_{l_\theta}\left(\frac{\omega}{c}r\right) H_{l_\theta}^{(2)}\left(\frac{\omega}{c}s\right) e^{-jl_\theta\theta_s}. \tag{4.16}$$

The spectrum obtained in (4.16) is shown in Fig. 4.4(a) for microphones located on a radius of 0.6 m and a source at 2 m from the center of the circle. When comparing the magnitude spectrum between the far field and the circular source case, we can observe that the circular case is obtained by multiplying the far field case formula (4.6) with a Hankel function as in (4.16). We show in Fig. 4.4(b) the effect of this multiplication for different ratios $\rho = \frac{s}{r}$ in a slice of the magnitude spectrum at $\omega = 10000$ rad/s. It is shown that for a source located at three times the radius of the circle, the magnitude spectrum is very similar to the far field case. When the source approaches the circle, the magnitude spectrum decays slower and slower as is shown for ρ varying from

3 to 1.1.



(a)



(b)

Figure 4.4: Spectrum of the PAF. (a) Butterfly spectrum of the PAF along a circle for a circular source. (b) Comparison of the far field PAF spectrum with the PAF spectrum for circular sources located at different distances.

Similarly to the sampling theorem described in Chapter 3, we can also develop a sampling theorem for the reconstruction of the PAF along a circle. The same approach as in Section 3.4.1 is taken and the SNR is defined in a similar

way. Consider a cut of the PAF spectrum at one temporal frequency. The spectrum only exists at integer values, namely l_θ. This cut of the spectrum is shown schematically in Fig. 4.5(a). When sampling the PAF along the angular direction with an angular spacing of $\Delta\theta$, the angular sampling frequency l_{θ_S} is defined as

$$l_{\theta_S} = \frac{2\pi}{\Delta\theta}. \tag{4.17}$$

The spectrum of the sampled PAF, as discussed in Chapter 3, is obtained by considering the original continuous angular PAF with all its spectral replicas as shown in Fig. 4.5(b). It is given by

$$\hat{p}_S(l_\theta, \omega) = \frac{1}{\Delta\theta\Delta t} \sum_{k_1,k_2=-\infty}^{\infty} \hat{p}\left(l_\theta - \frac{2\pi k_1}{\Delta\theta}, \omega - \frac{2\pi k_2}{\Delta t}\right). \tag{4.18}$$

Due to the non-perfect bandlimitedness of the spectrum of the PAF, perfect reconstruction is not achieved. A small error remains on the reconstruction depending on the decay of the spectrum outside of the butterfly region. The SNR defined in (3.30) rewrites here as

$$\mathrm{SNR}(l_{\theta_S}, \omega_0) = \frac{\sum_{l_\theta=-\infty}^{\infty} |\hat{p}(l_\theta, \omega_0)|^2}{4\sum_{l_\theta \geq \frac{l_{\theta_S}}{2}} |\hat{p}(l_\theta, \omega_0)|^2}. \tag{4.19}$$

Theorem 4.2.1. In two dimensions, assume one single source emitting in free field at a frequency $\omega = \omega_0$ located at distance s from the center of a circle of radius r. When sampling the PAF on the circle at an angular sampling frequency of l_{θ_S}, for a particular $\omega = \omega_0$, and reconstructing it using an ideal interpolator, the SNR of the reconstructed signal in the band $[-\frac{l_{\theta_S}}{2}, \frac{l_{\theta_S}}{2}]$ can be expressed as

$$\mathrm{SNR}(l_{\theta_S}, \omega_0) = \frac{\sum_{l_\theta=-\infty}^{\infty} |J_{l_\theta}\left(\frac{\omega_0}{c} r_<\right) H_{l_\theta}^{(2)}\left(\frac{\omega_0}{c} r_>\right)|^2}{4\sum_{l_\theta \geq \frac{l_{\theta_S}}{2}} |J_{l_\theta}\left(\frac{\omega_0}{c} r_<\right) H_{l_\theta}^{(2)}\left(\frac{\omega_0}{c} r_>\right)|^2}, \tag{4.20}$$

where $H_{l_\theta}^{(2)}$ is the Hankel function of order l_θ of the second kind, J_{l_θ} is the Bessel function of order l_θ, $r_<$ the lesser of r and s and $r_>$ the greater.

Numerically, (4.20) has been simulated and the results are shown in Fig. 4.6 for temporal frequencies varying from 1 to 8000 rad/s. The simulations have been performed for a source located at 2 m from a circle of radius 1 m. A numerical example is given.

Example 4.2.1. consider a circular wave emitting at 8000 rad/s (1.3 kHz).

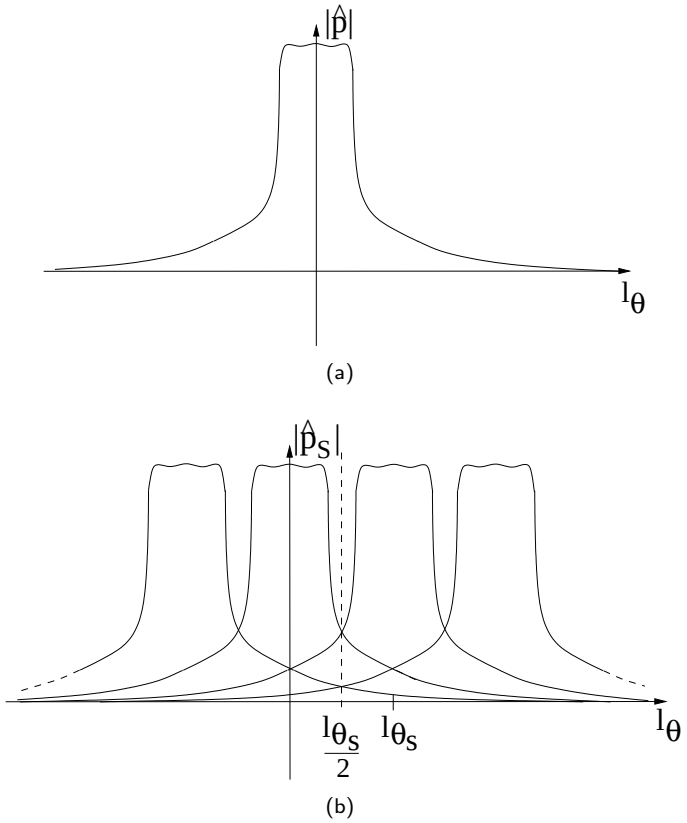


Figure 4.5: Sampling theorem of the PAF along a circle. (a) A cut at one specific temporal frequency of the spectrum of the PAF. (b) Due to the sampling of the PAF in the angular direction, the spectrum of the sampled PAF contains multiple periodic repetition of the original spectrum.

Using the approximation (4.7), one would require an angular spacing between consecutive microphones of 7.6° while from (4.20) and Fig. 4.6, it turns out that to obtain a mean SNR of 50 dB on the reconstruction, an angular spacing of 5.9° is necessary. □

4.2.2 Angular bandwidth of the 3D sound field on a circle

Up to now, the sound field along a circular array has been studied for a 2D plane wave and for a circular sound source. In a 3D setup, the sound source is a spherical source emitting in 3D. We therefore slightly modify the setup and introduce the third dimension. We consider a similar setup as in Fig. 4.3. The coordinates of the different microphones are now (r_x, r_y, r_z), with $r_x = r\cos\theta$, $r_y = r\sin\theta$ and $r_z = 0$. The sound source is located at a distance s from the

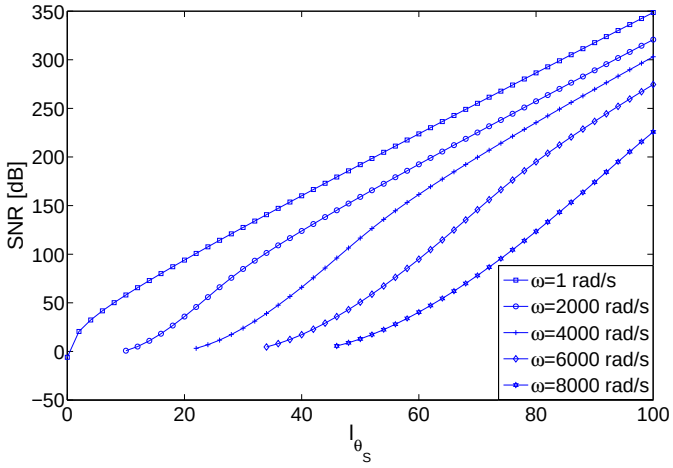


Figure 4.6: Numerical simulation for the calculation of the SNR for different ω_0.

center of the array and with coordinates (s_x, s_y, s_z) where $s_x = s\cos\theta_s$ and $s_y = s\sin\theta_s$. The distance s can be decomposed into a distance $s_p = \sqrt{s_x^2 + s_y^2}$ in the plane of the microphones and s_z with

$$|s| = \sqrt{s_p^2 + s_z^2}. \tag{4.21}$$

The different signals recorded at any angle on the circle are gathered in a function $p(\theta, t)$. When the sound source emits a Dirac signal, the microphones located on the circle measure

$$p(\theta, t) = \frac{\delta(t - h(\theta))}{4\pi c h(\theta)}, \tag{4.22}$$

where $h(\theta)$ is defined as

$$h(\theta) = \frac{\sqrt{(s_x - r\cos\theta)^2 + (s_y - r\sin\theta)^2 + s_z^2}}{c} \tag{4.23}$$

$$= \frac{\sqrt{s^2 + r^2 - 2rs_p\cos(\theta - \theta_s)}}{c}. \tag{4.24}$$

Similarly to the two-dimensional case, the function $h(\theta)$ represents, at each angular microphone position θ, the time of arrival taken by sound to propagate from the source to the different microphones. An example of this function is given in Fig. 4.7. To study the angular bandwidth of the sound field recorded on a circle, the two-dimensional Fourier transform (2D-FT) of $p(\theta, t)$ is again

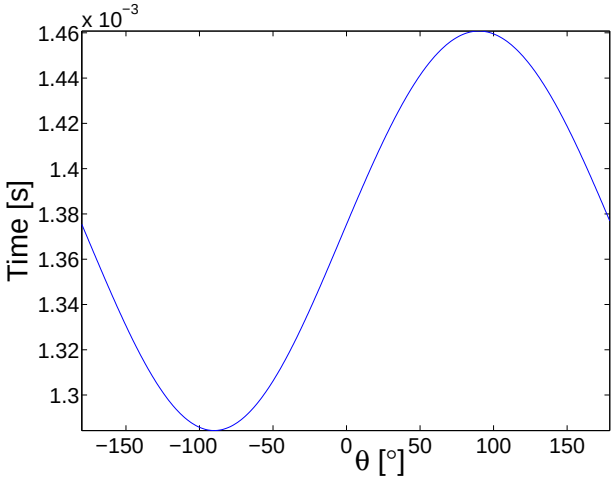


Figure 4.7: Time of arrival from source to microphone at each angular position.

studied

$$\hat{p}(l_\theta, \omega) = \frac{1}{2\pi} \int_0^{2\pi} \int_{-\infty}^{\infty} p(\theta, t) e^{-j(l_\theta \theta + \omega t)} \mathrm{d}t \mathrm{d}\theta \tag{4.25}$$

$$= \frac{1}{8\pi^2 c} \int_0^{2\pi} \frac{e^{-j\omega h(\theta)}}{h(\theta)} e^{-jl_\theta \theta} \mathrm{d}\theta. \tag{4.26}$$

A closed form solution of this equation has not been obtained yet. Therefore, to estimate the spatial support of the spectrum, we present some approximation formulae in the sequel.

In the study of the spatial bandwidth of the sound field, the attenuation depending on the distance traveled has been shown to be negligible for the study of the essential support of the PAF spectrum in [6]. It mostly affects the decay of the function. Therefore, a good estimate of the angular bandwidth corresponds to the support of the function:

$$\hat{p}(l_\theta, \omega) = \int_0^{2\pi} e^{-j\omega h(\theta)} e^{-jl_\theta \theta} \mathrm{d}\theta. \tag{4.27}$$

Note that (4.27) corresponds to the Fourier transform of a phase modulation (PM) signal where the carrier frequency would be equal to zero and the modulation function would be $h(\theta)$. The bandwidth of this signal can easily be estimated using the Carson's rule [29, 41]. For a temporal frequency ω, the bandwidth (BW) of $\hat{p}(l_\theta, \omega)$ along the angular frequency axis can be approxi-

mated by

$$\mathrm{BW}(\hat{p}(l_\theta, \omega)) \approx \max_\theta \left[\frac{dh}{d\theta}\right] \omega + W, \tag{4.28}$$

with W the bandwidth of $h(\theta)$[1].

The maxima of the first derivative of $h(\theta)$ with respect to θ can be found to be [6]

$$\frac{dh}{d\theta} = \pm \frac{\sqrt{A + 2s_p r} - \sqrt{A - 2s_p r}}{2c}, \tag{4.29}$$

with $A = s^2 + r^2$. For sources located in the same plane as the circular array, the expression of the derivative gets simpler:

- for a source located inside of the circular array,

$$\max_\theta \left[\frac{dh}{d\theta}\right] = \pm \frac{s}{c}; \tag{4.30}$$

- for a source located outside of the circular array,

$$\max_\theta \left[\frac{dh}{d\theta}\right] = \pm \frac{r}{c}. \tag{4.31}$$

Further we can also show that the maximal derivative of $h(\theta)$ associated to a source outside of the plane of the array is always smaller than the derivative for a source on the plane:

$$\left| \frac{\sqrt{A + 2s_p r} - \sqrt{A - 2s_p r}}{2c} \right| \leq \left| \frac{\min(s_p, r)}{c} \right|.$$

The maximum derivative is thus only dependent on the minimum between the distance from the center of the array to the source and the radius of the array,

$$\max_\theta \left[\frac{dh}{d\theta}\right] \leq \frac{\min(s_p, r)}{c}. \tag{4.32}$$

The signal $h(\theta)$ is a very smooth signal and therefore its bandwidth W can be shown to be approximately zero (unless in specific cases where the source is very close to the microphones as has been discussed previously in Section 4.2.1).

[1] To apply Carson's rule, either $W \ll \max_\theta \left[\frac{dh}{d\theta}\right] \omega$ or $W \gg \max_\theta \left[\frac{dh}{d\theta}\right] \omega$ needs to be satisfied.

Therefore, (4.28) becomes

$$\mathrm{BW}(\hat{p}(l_\theta, \omega)) \approx \frac{\omega}{c} \min(s_p, r). \tag{4.33}$$

In most cases the source is considered to be in the same plane as the circular array and we have then

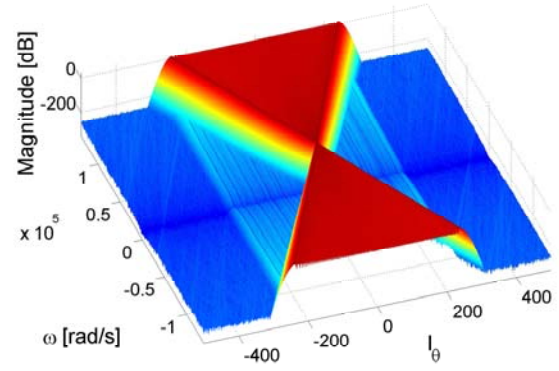
$$\mathrm{BW}(\hat{p}(l_\theta, \omega)) \approx \frac{\omega}{c} \min(s, r). \tag{4.34}$$

As indicated by (4.34), the angular frequency support increases linearly with the temporal frequency. This corresponds exactly to the butterfly support presented in Fig. 4.8(a) where most of the energy is present in the region satisfying

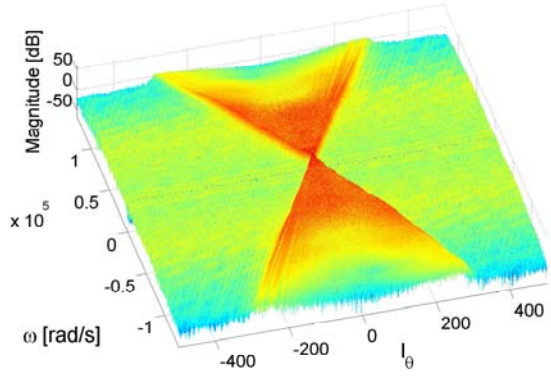
$$|l_\theta| \leq |\omega| \frac{r}{c}. \tag{4.35}$$

This spectrum was obtained by simulating the sound field measured on a circular array of radius 0.6 m with a source at 2 m from the center of the array. The same setup was followed for real measurements where 1000 room impulse responses (RIRs) were measured in a sound insulated room. The measurements were performed with a microphone every 0.36° along a circle. These measurements were performed using an automatic Pan/Tilt unit $PTU - D46 - 70$ rotating with a precision of 0.03° in the median plane. The RIRs were measured using a logarithmic sweep [86] of 2 seconds duration. The loudspeaker used was a Genelec $1029A$ and the microphone Beyerdynamic $MC - 740$. The 2D spectrum obtained from the real RIRs measurements is shown in Fig. 4.8(b). The support of the spectrum for the real measurements corresponds well to the simulations. All these measurements are available online at the following address [2].

In this chapter we only have considered the free field case where no reverberation was present. We give now a justification for the fact that measurements containing a large number of reflections also exhibit a similar magnitude spectrum. In Fig. 4.9, we present a scheme with a circular array of radius r and a source s_1 in a rectangular room. The reverberation is simulated with the image source model as described previously in Chapter 3. For simplicity we consider the source to be in the same plane as the microphones. Three virtual sources are present in the scheme. Every virtual source leads to a butterfly spectrum dictated by (4.34) where $\min(r, s) = r$. Therefore the total support is not larger with increasing number of virtual sources. Note that in the case of a source inside of the array, the free field support would satisfy (4.34) with $\min(r, s) = s$. The virtual sources would then lead to larger support, namely



(a)



(b)

Figure 4.8: 2D-FT of the sound field recorded on a circular array. (a) For a simulated case. (b) For real measurements.

(4.34) with $\min(r, s) = r$. In that case, the total support is then larger than the free field support and follows the same rule as if the original source would have been outside of the circle.

For completeness, a set of electromagnetic (EM) channel impulse responses (CIRs) is also presented in this chapter. Measurements of EM CIRs have been done in a room along a circular array at 280 positions along a circle of radius 70 cm. Similarly to the setup used for the linear array measurements in Chapter 3, the excitation signal was produced by an pseudonoise generator (Centellax $TG1P1A$) containing energy in a range between 2 GHz to 5 GHz. The impulse responses were measured by antennas using an oscilloscope Lecroy $SDA - 6000$. The CIRs are shown in Fig. 4.10(a) and the corresponding 2D

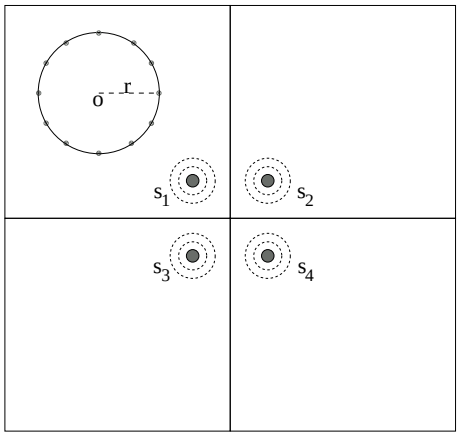


Figure 4.9: Scheme of the situation with 4 sources corresponding to reverberation.

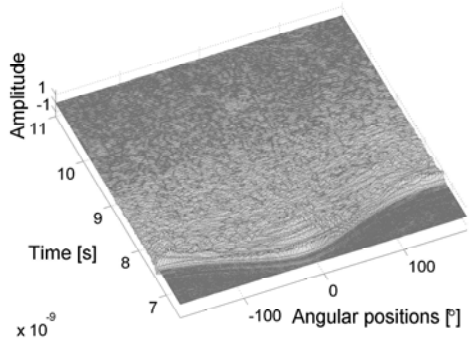
spectrum in Fig. 4.10(b). This spectrum also exhibits a butterfly shape. Nevertheless, as the excitation signal contained small energy for frequencies below 2 GHz and the used antennas only picked up signal of higher frequencies, the butterfly shape is only visible above that frequency, what explains the trapezoidal shape of the spectrum. Note that in the electromagnetic case, the support of the butterfly is now dependent on the speed of light propagation.

4.2.3 HRTF sampling and interpolation

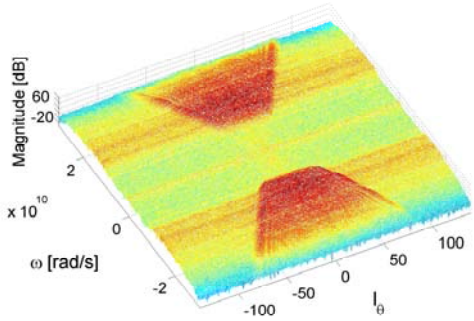
In Sections 4.2.1 and 4.2.2, the spatial bandwidth of the sound field along a circle has been studied. The same theory can be applied for the study of HRTF interpolation. HRTFs measurements are carried out in an anechoic chamber to characterize the effect of pinnae, head, and torso of a person on the perceived sound [20]. The typical setup for HRTF measurements is shown in Fig. 4.11. The loudspeakers are located along a circle around the person. The microphone is located at the entrance of the ear canal of the listener to capture the sound. By constructing $p(\theta, t)$ as the HRIRs measured at every possible angle, the support of the spectrum $\hat{p}(l_\theta, \omega)$ can be estimated using the theory presented in Section 4.2.2. With the head well centered in the middle of the loudspeaker array and the position of the microphone being $\frac{d}{2} = 9$ cm away from the center of the circle (half the spacing between the two ears) [20], (4.35) rewrites as

$$|l_\theta| \leq |\omega| \frac{d}{2c} \approx |\omega| \frac{0.09}{c}. \tag{4.36}$$

Therefore, for any temporal frequency, the necessary angular spacing between consecutive loudspeaker positions can be derived. Considering the energy outside of the butterfly spectrum as negligible, the angular sampling frequency



(a)



(b)

Figure 4.10: Experimental electromagnetic channel impulse responses measured between 2 and 5 GHz. (a) CIRs measured at different angular positions. (b) Corresponding 2D-FT of the measured CIRs.

needs to satisfy

$$l_{\theta_S} > 2|\omega_{\max}|\frac{d}{2c} \approx 2|\omega_{\max}|\frac{0.09}{c}, \tag{4.37}$$

with $\omega_{\max}$ the maximal temporal frequency present in the signal. In particular, (4.37) indicates that in order to sample HRTFs for an average adult human ($d \approx 0.18$ m) with a temporal sampling rate of 44.1 kHz, a spacing of at least $4.9°$ is necessary. Sampling the HRTFs with a too large angular spacing leads to spectral repetitions corrupting all temporal frequencies above $\omega_{\max}$ in (4.37).

HRTF interpolation

The interpolation of the dataset is done by observing that the spectrum of the HRTFs contains almost no energy outside of the butterfly region. Therefore, as

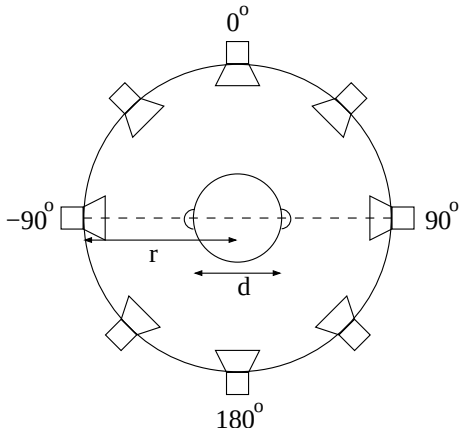


Figure 4.11: Setup for the recording of HRTFs.

the signal to be interpolated is almost bandlimited, good results are obtained using a sinc interpolator. The implementation of this interpolator happens most efficiently by zero-padding in frequency domain. This interpolation is very suitable in the case of a circular array since the Fourier transform is applied on an array that is 2π periodic. In the case of interpolation along non-periodic arrays, border effects decrease the interpolation performance as is described in Chapter 3.

Depending on the angular sampling of the database of HRTFs considered, interpolation is only applied for frequencies satisfying (4.37). Higher frequencies will not be correctly interpolated due to the spectral repetitions leading to aliasing. For these higher frequencies other techniques need to be used as will be shown in Section 4.3.

Head shadowing

The theory expressed above is valid in the case of HRTFs when the effect of the head shadowing is not considered. In practice, waves are diffracted by the head. This diffraction has to be taken into account. Diffraction has a large impact on the level of amplitude of the HRTF [20]. It also affects the shape of the function $h(\theta)$. In the model given by [42], the HRTFs are expressed as:

$$\tilde{h}(\theta, \rho, \mu) = -\frac{\rho}{\mu} e^{-i\mu\rho\Psi}, \tag{4.38}$$

with

$$\Psi(\theta, \rho, \mu) = \sum_{m=0}^{\infty} (2m+1) P_m(\cos\theta) \frac{h_m(\mu\rho)}{h'_m(\mu)},$$

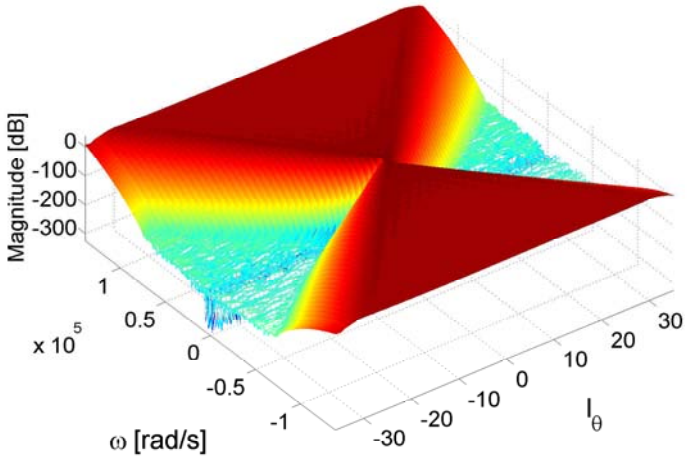
where μ is the normalized temporal frequency, P_m is a Legendre polynomial of degree m, and h_m is an m^{th} order spherical Hankel function[2]. Taking the Fourier transform of (4.38) a butterfly spectrum satisfying (4.36) is also observed, as shown in Fig. 4.12(a). The obtained spectrum considering HRTFs measured on a Kemar head [50] sampled every 5° in an anechoic chamber is shown in Fig. 4.12(b). There also, (4.37) is satisfied since a spacing of 5° results in almost no aliasing at 44.1 kHz.

4.3 HRTF Interpolation in a Temporal Envelope Domain

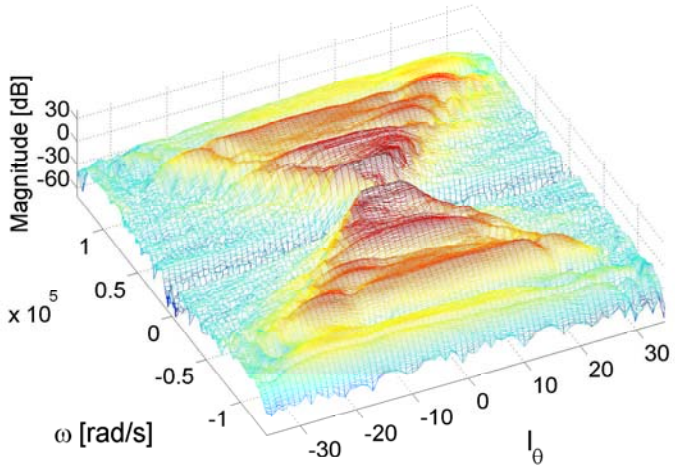
With the analysis presented in Section 4.2.2, to reconstruct HRTFs at a sampling rate of 44.1 kHz, 72 HRTFs are necessary to achieve good interpolation results with little aliasing. This number is quite large and a lot of available databases do not contain such a fine sampling. In this section, a method will be presented to increase the performance of the interpolation in case a coarser sampling of the HRTFs is at disposal for interpolation. Considering a database containing two or four times less HRTFs than the dictated number by (4.37), it will be shown that satisfactory results in a least mean squared sense can still be achieved.

The technique presented in this section is based on a subband decomposition of the HRTFs. A scheme illustrating the method is shown in Fig. 4.13. As discussed in Section 4.2, the best interpolation is achieved using the sinc interpolation when no aliasing is present. Therefore this technique is still used for the range of temporal frequencies satisfying (4.37). This frequency range is denoted as a in Fig. 4.13. For the higher frequencies a new technique is presented. In this technique each subband is considered and analyzed separately. These subbands are denoted as b in Fig. 4.13. At each angular position θ_i, the subband signal is decomposed into its complex envelope signal and carrier signal. The algorithm aligns the carriers of the different angular positions to the carrier of a reference position. Therefore, only the envelope signals are differing between the signals at various angular positions and interpolation is applied to the envelopes. Due to the smaller bandwidth of the envelopes, spatial aliasing is reduced. Fig. 4.13 illustrates the signals recorded at positions θ_1 and θ_2. In dotted lines, the envelopes of these two signals are shown. After carrier alignment of signal at position θ_2 with the carrier of the reference signal at position θ_1, the new aligned carrier at position θ_2 is shown in bold. It can

[2]Note that this notation is different than the one used in our thesis but has been kept as in the original paper [42].



(a)



(b)

Figure 4.12: 2D spectra of HRTFs: (a) using a diffraction model; (b) using measured data.

be observed that the envelope of the signal at position θ_2 is kept unmodified by this alignment. Based on the idea that the envelope signals evolve slower along the angular dimension than the HRTFs themselves, satisfactory interpolation results can be obtained even when the angular spacing is not satisfying (4.37). The technique can be summarized in four steps:

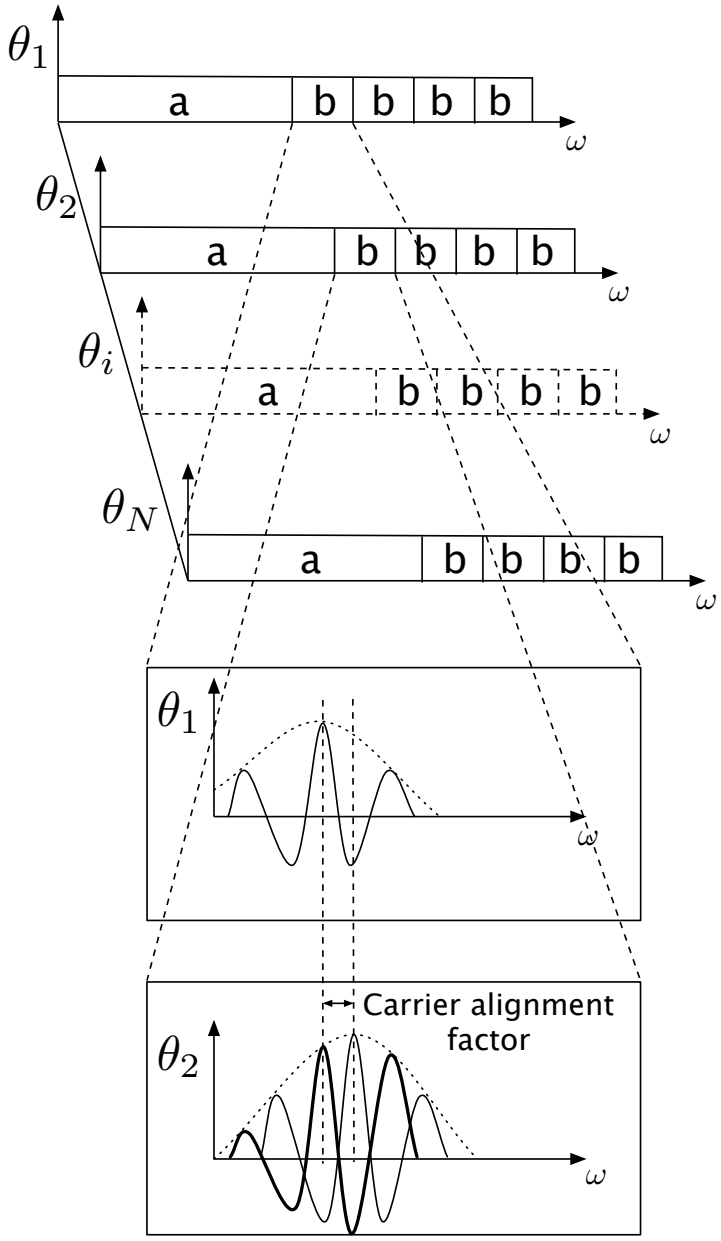


Figure 4.13: Scheme of the presented technique for HRTF interpolation. In the frequency range a, the interpolation happens with a sinc interpolator. In the frequency bands denoted as b, the technique presented in this section is used.

1. The first step consists in the estimation of the carrier alignment factors for each subband and at each angular position. The estimation of those factors is presented in Section 4.3.1. Section 4.3.2 shows further that the

angular support of the carrier aligned HRTFs is reduced when compared to the original one.

2. Once the HRTFs are aligned, the next step is to interpolate the carrier aligned HRTFs. This is discussed in Section 4.3.3.

3. The third step is to interpolate the alignment factors used to align the different HRTFs which is also discussed in Section 4.3.3.

4. In the last step, all interpolated HRTFs are obtained by multiplying each interpolated carrier aligned HRTF with its corresponding carrier alignment factor.

4.3.1 Computing the carrier alignment factors in subbands.

Each subband signal is denoted as $h_i(\theta, t)$ where the index i corresponds to the subband index and θ stands for the azimuth angle characterizing the HRTF position. The Fourier transform with respect to time of $h_i(\theta, t)$ is denoted as $\tilde{h}_i(\theta, \omega)$. The basic idea behind the algorithm is that each subband signal is considered to be made of the product of a carrier and an envelope signal. For every subband, the carriers of the different positions are aligned to the carrier of one reference position called θ_r. Considering only the positive frequencies in $\tilde{h}_i(\theta, \omega)$, this alignment is achieved by multiplying each HRTF spectrum subband of position θ by a complex number $c_i(\theta) = e^{j\alpha_{\theta,i}}$. Denote $\tilde{k}_i(\theta, \omega)$ the aligned HRTF subband, we have that

$$\tilde{k}_i(\theta, \omega) = c_i(\theta)\tilde{h}_i(\theta, \omega). \tag{4.39}$$

The complex value $c_i(\theta)$ is obtained by minimizing the error J_i defined as:

$$J_i = \int_{\omega_i}^{\omega_{i+1}} \left| \tilde{h}_i(\theta_r, \omega) - \tilde{h}_i(\theta, \omega)c_i(\theta) \right|^2 \mathrm{d}\omega, \tag{4.40}$$

with ω_i and ω_{i+1} the start and stop frequencies of band i.

The value of $c_i(\theta)$ which minimizes J_i is calculated in Appendix D:

$$c_i(\theta) = \frac{\int_{\omega_i}^{\omega_{i+1}} \tilde{h}_i(\theta_r, \omega)\tilde{h}_i^*(\theta, \omega)\mathrm{d}\omega}{\left| \int_{\omega_i}^{\omega_{i+1}} \tilde{h}_i(\theta_r, \omega)\tilde{h}_i^*(\theta, \omega)\mathrm{d}\omega \right|}. \tag{4.41}$$

By calculating $c_i(\theta)$ as in (4.41) for each azimuthal angle and multiplying the corresponding subband signal with it as in (4.39), the carrier signals of all the HRTFs are aligned to the reference signal in a least mean square sense. Fig. 4.14 shows an example of this carrier alignment procedure. The signal

measured at the reference position θ_r is shown in dashed lines together with its envelope in dashed-dotted line. The signal observed at position θ_2 is shown in full lines together with its envelope in dotted lines. Multiplying the signal with the carrier alignment factor yields to the signal in bold lines. It can be seen that both the signal at position θ_2 and its carrier aligned version have the same envelope. The carrier of the newly obtained signal is now aligned to the reference signal carrier. The angular bandwidth occupied by this new HRTF

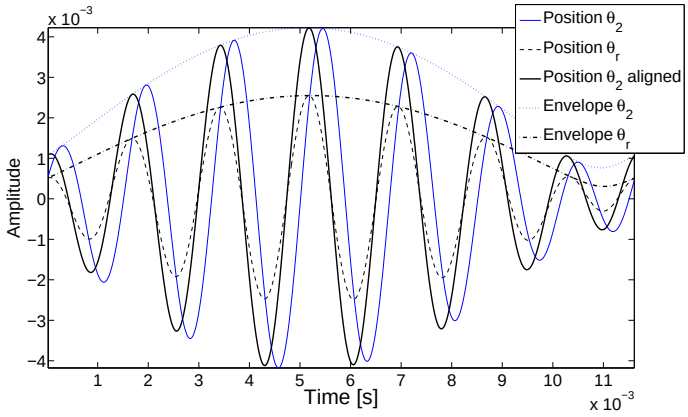


Figure 4.14: Different subband signals.

dataset is reduced as will be shown in Section 4.3.2. This reduction of the bandwidth is the key factor in the possibility of achieving satisfactory results when interpolating signals whose angular spacing is not satisfying (4.37).

4.3.2 Support of the aligned HRTFs

In this section, for each subband HRTF $h_i(\theta, t)$ the analytical signal $h_{+i}(\theta, t)$ is given by

$$h_{+i}(\theta, t) = h_i(\theta, t) + j\check{h}_i(\theta, t), \tag{4.42}$$

where $\check{h}_i(\theta, t)$ stands for the Hilbert transform of $h_i(\theta, t)$. Recall that

$$\check{h}(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{h(t)}{t - \tau} \mathrm{d}\tau. \tag{4.43}$$

The Fourier transform with respect to time of $h_{+i}(\theta, t)$ is denoted by $\tilde{h}_{+i}(\theta, \omega)$. For the positive frequencies, it can be shown that

$$\tilde{h}_{+i}(\theta, \omega) = 2\tilde{h}_i(\theta, \omega). \tag{4.44}$$

For the negative frequencies, $\tilde{h}_{+i}(\theta,\omega)=0$.

The signal $h_{+i}(\theta,t)$ can be decomposed into a carrier $g_i(\theta,t)$ and a complex envelope signal $e_i(\theta,t)$. Writing the carrier signal as $g_i(\theta,t)=e^{j(\omega_i t-\alpha_{\theta,i})}$, $h_{+i}(\theta,t)$ is obtained by

$$h_{+i}(\theta,t)=e_i(\theta,t)g_i(\theta,t)=e_i(\theta,t)e^{j(\omega_i t-\alpha_{\theta,i})}=e_i(\theta,t)c_i^*(\theta)e^{j\omega_i t}. \quad (4.45)$$

As discussed in Section 4.3.1, each carrier are aligned to the carrier of a reference HRTF. The direction of the reference HRTF is θ_r. After realignment of all HRTFs to this reference using (4.41)[3], the new HRTF set is called $k_{+i}(\theta,t)$, with

$$k_{+i}(\theta,t)=e_i(\theta,t)e^{j\omega_i t}. \quad (4.46)$$

Consider now the 2D spectrum of the HRTF for one subband. Call $\hat{h}_{+i}(l_\theta,\omega)$ the 2D-FT of $h_{+i}(\theta,t)$ and $\hat{k}_{+i}(l_\theta,\omega)$ the 2D-FT of $k_{+i}(\theta,t)$, one can write

$$\hat{k}_{+i}(l_\theta,\omega)=\frac{1}{2\pi}\int_0^{2\pi}\int_{-\infty}^{\infty}e_i(\theta,t)e^{-j(\omega-\omega_i)t}\mathrm{d}t e^{-jl_\theta\theta}\mathrm{d}\theta. \quad (4.47)$$

Calling $\hat{e}(l_\theta,\omega)$ the 2D-FT of the complex envelope signals $e_i(\theta,t)$, (4.47) can be rewritten as

$$\hat{k}_{+i}(l_\theta,\omega)=\hat{e}(l_\theta,\omega-\omega_i). \quad (4.48)$$

This shows that after carrier alignment, the 2D spectrum of the aligned HRTFs is equivalent to the support of the complex envelopes of the HRTFs shifted to the temporal frequency of the subband considered.

As mentioned in Section 4.3.1, multiplying the signal by the alignment factor does not modify the real envelope of the signal. The real envelope of the signal $h_i(\theta,t)$ is obtained by calculating

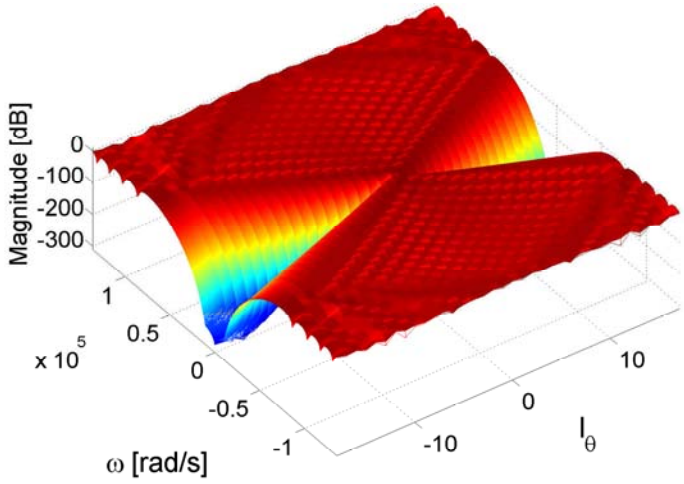
$$\text{envelope}_i(\theta,t)=|h_{+i}(\theta,t)|. \quad (4.49)$$

In the presented technique, the signals $h_{+i}(\theta,t)$ are aligned by multiplication by a phase factor $e^{j\alpha_{\theta,i}}$. This multiplication does not modify the real envelope signal since

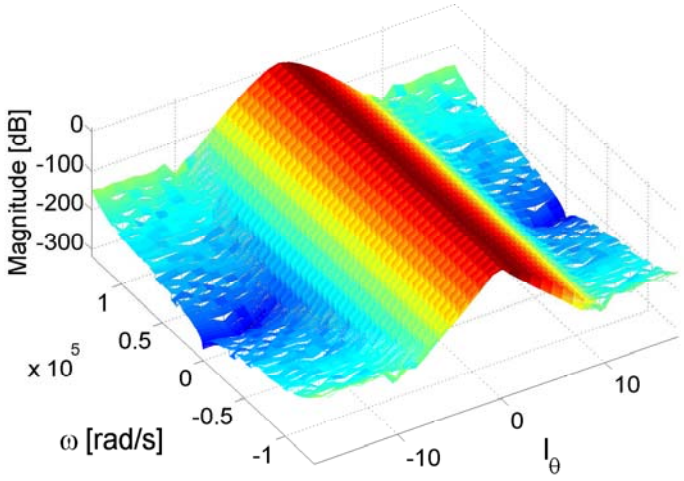
$$\text{envelope}_i(\theta,t)=|e_i(\theta,t)g_i(\theta,t)c_i(\theta)|=|e_i(\theta,t)|. \quad (4.50)$$

[3]In this analysis, we consider a perfect realignment of the different carriers to the reference one.

As an example of the efficiency of the carrier alignment technique, consider a simple model for HRTFs where only the delay and attenuation is considered [85], Fig. 4.15(a) represents the 2D-FT of this database sampled every 10°. After carrier alignment the 2D spectrum is much narrower as can be observed in Fig. 4.15(b). Comparing Fig. 4.15(a) and 4.15(b) at the particular temporal



(a)



(b)

Figure 4.15: 2D spectra of HRTF: (a) Original HRTF dataset. (b) Carrier aligned HRTFs.

frequency of 5 kHz (10000π rad/s) yields Fig. 4.16. It can be observed that the carrier aligned HRTFs have a much narrower spectral content and can therefore be sampled with a larger spacing between consecutive angular positions.

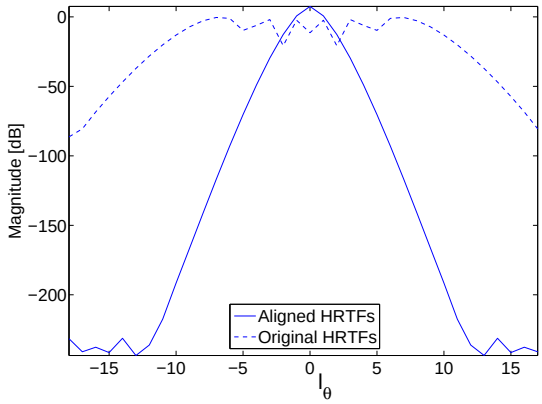


Figure 4.16: At 5 kHz, the angular frequency support of the original HRTFs in dashed lines is compared to the support of the carrier aligned HRTFs in full lines.

4.3.3 Interpolation

The interpolation of the HRTFs is carried out in three steps (corresponding to steps 2 to 4 mentioned previously). The first step is to interpolate the carrier aligned signals $k_{+i}(\theta, t)$. The second step is the interpolation of the alignment factors. The third step consists in multiplying the carrier aligned HRTF at a specific angular position with the corresponding alignment factor. The interpolated HRIR is obtained by inverse Fourier transform of the spectrum where the negative frequencies have been added. The negative frequencies are the complex conjugate of the positive frequencies to ensure real HRIRs.

For the interpolation of $k_{+i}(\theta, t)$, a simple low-pass interpolation filter using the 4 nearest neighbors is used. Similarly to the analysis performed in Section 4.2.3, it would be interesting to know how many angular positions are necessary to interpolate the HRTFs at any angular position. This number can unfortunately not easily be calculated in a closed form solution. This is due to the fact that the angular frequency support of the complex envelopes is not known in general. Nevertheless, as the angular support corresponding to the carrier aligned signals corresponds to the one of the envelopes signals, it can be concluded that the number of angular samples is reduced.

The second step is the interpolation of the carrier alignment factors $c_i(\theta)$ along the angular direction. To interpolate these complex numbers, only the phases $\alpha_{\theta,i}$ are interpolated. This interpolation is also achieved by a low-pass

filter using the 4 nearest neighbors. Nevertheless, a difficulty due to unwrapping appears when getting the phase values from the complex numbers $c_i(\theta)$. The phase $\alpha_{\theta,i}$ is only known in the range from $-\pi$ to π. Denoting $\hat{\alpha}_{\theta,i}$ the phase between $-\pi$ to π, the phase to be found satisfies

$$\alpha_{\theta,i} = \hat{\alpha}_{\theta,i} + 2k\pi, \tag{4.51}$$

with $k \in \mathbb{Z}$. The unknown is here the value of k. Interpolation between wrong phases leads to wrong alignment factors and poor performance of the proposed algorithm. The unwrapping problem can be addressed by considering a recursive algorithm where the subbands at lower temporal frequencies are used to estimate the range of values of the phase in the higher temporal frequency subband. The estimation of the new subband phase is obtained by assuming a linear phase. This estimation of the new phase allows us then to find the k in (4.51). Note that when the HRTF filter is very far from a linear phase filter, some wrong evaluation of k can be found. Also, remark that as the algorithm is recursive, the higher the temporal frequency, the larger the error can become. With our algorithm, the phases for the angular positions at the different subbands are shown in Fig. 4.17(a), (b) and (c) for different HRTF databases. In the sequel the databases will be denoted as Database 1, Database 2, and Database 3. Database 1 simply considers at each angular position a delay and an attenuation satisfying the wave equation. Database 2 uses the model of head shadowing discussed in Section 4.2.3 from [42]. Database 3 has been obtained by measurement of the HRTFs of a Kemar head [50].

In the presented technique, one of the important parameters to be chosen is the number of subbands in which the HRTFs are decomposed. Note that there exists a trade-off in the choice of the bandwidth and the number of subbands. Choosing a too large band makes it difficult to correctly align one subband HRTF at one position with its reference using only one value (the corresponding $c_i(\theta)$). In other words, the error J_i in (4.40) might remain important for large bands. Choosing the band too small leads to a very large number of bands with a more and more difficult unwrapping problem at high frequencies.

4.4 Simulations Using HRTFs Models and Measured Data

In this section, simulation results using HRTF models as well as experimental measurements are presented. The technique presented in Section 4.3 is compared with other interpolation techniques. The three databases described

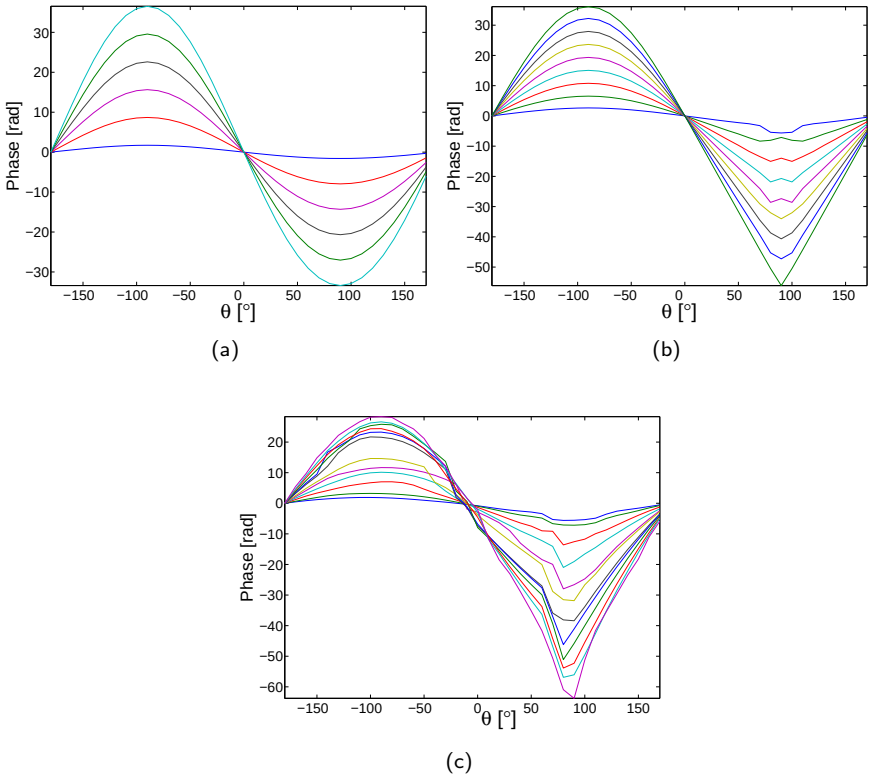


Figure 4.17: Phase corresponding to the carrier alignment factors for different frequency subbands: (a) For Database 1; (b) For Database 2; (c) For Database 3.

previously in Section 4.3.3 are used in this section. All HRTFs are sampled at a temporal sampling frequency of 44.1 kHz. The purpose of this section is not to provide a comparison of all existing techniques. Nevertheless, four methods are compared and then further discussed. The typical setup for HRTF interpolation happens as follows. From N available HRTF measurements, the purpose is to obtain after interpolation mN HRTFs where m stands for the interpolation factor. In other words, $m-1$ new measurements are obtained between two consecutive original HRTFs.

- Method 1 considers an interpolation of each time sample using a periodic sinc interpolation along the spatial direction.
- Method 2 corresponds to the new technique presented in Section 4.3.
- Method 3 corresponds to the technique presented in [81]. There, the HRTFs are first time-aligned. The alignment is obtained by finding the maximum of the cross-correlation function between the considered

HRTFs. In order to have sub-sample precision in the time delay estimation, the HRTFs are interpolated in the time domain prior to the cross-correlation estimation. Once aligned, the HRTFs are interpolated using standard interpolation. Similar results are obtained in case of linear interpolation or low-pass interpolation using the first neighbors.

- Method 4 is a new technique presented in this study based on Method 1 and Method 3. The low frequency content of the HRTFs satisfying (4.37) is interpolated using a sinc interpolator as in Method 1. The higher frequency content is then interpolated using Method 3.

To compare the performance of the different interpolation algorithms, the error obtained by interpolation of the HRTFs is studied either as a function of the different angular positions or as a function of the temporal frequency. The mean squared error (MSE) on an interpolated HRTF at one position is calculated as follows. Calling $h(\theta, n)$ the original discrete-time HRTF and $h_e(\theta, n)$ its estimated version, the MSE at an angular position θ_0 is defined as

$$\mathrm{MSE}(\theta_0) = 10\log_{10}\frac{\sum_{n=0}^{T}\left(h(\theta_0, n) - h_e(\theta_0, n)\right)^2}{\sum_{n=0}^{T} h^2(\theta_0, n)}, \tag{4.52}$$

where T stands for the number of time samples of the HRTF. When considering the frequency dependent error, the MSE is averaged over all spatial positions and studied as a function of the temporal frequency. We therefore introduce the Fourier transform with respect to time of $h(\theta, n)$ and $h_e(\theta, n)$ denoted as $\tilde{h}(\theta, f)$ and $\tilde{h}_e(\theta, f)$, respectively. The variable f stands here for the Fourier bin. The frequency dependent MSE is now defined at one temporal frequency bin f_0 as

$$\mathrm{MSE}(f_0) = 10\log_{10}\frac{\sum_{\theta=0}^{N}\left|\tilde{h}(\theta, f_0) - \tilde{h}_e(\theta, f_0)\right|^2}{\sum_{\theta=0}^{N}\left|\tilde{h}(\theta, f_0)\right|^2}, \tag{4.53}$$

In Section 4.2, considerable effort was given to study the angular bandwidth of the sound field along a circular array. It was shown that using the sinc interpolation is recommended due to the almost bandlimited character of the frequency support. Therefore, the sinc interpolation is described in detail in Section 4.4.1.

4.4.1 Sinc interpolation

In this section, we describe results obtained using the sinc interpolator. Note that due to the periodic character of the sound field, the sinc considered is not

the same as the one used in Chapter 3. Here the sinc is a periodic sinc [89]. The implementation of the sinc interpolator is done by zero-padding in the frequency domain as previously mentioned in Section 4.2.3.

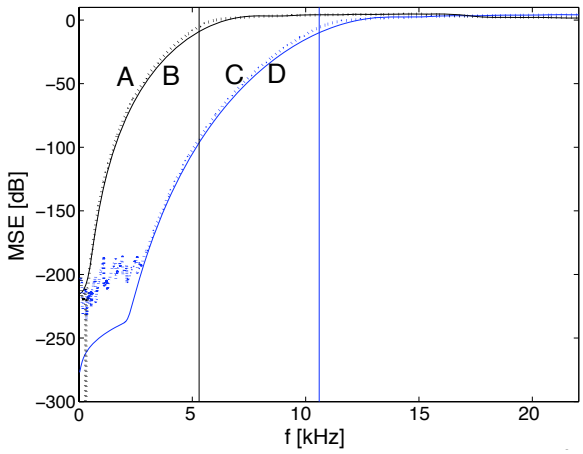
As was explained in Section 4.2, applying a sinc interpolation only makes sense when (4.37) is satisfied. This relation can be verified by observing the frequency dependent error on the interpolated HRTFs. Fig. 4.18(a) presents the frequency dependent error on HRTFs averaged over all interpolated angular positions. An interpolation of factor two was applied in the case of 20° (curves A and B) and 10° (curves C and D) angular spacing between consecutive HRTFs. Curves A and C (solid lines) were obtained using Database 1 and curves B and D (dotted lines) with Database 2. It can be observed that the curves considering the head shadowing are very close to the simpler model without considering head shadowing. Therefore it can be concluded that taking this effect into account does not modify the average error significantly. Fig. 4.18(b) presents the same results for Database 3. The interpolation error when using HRTFs every 10° is shown as a solid line and as a dotted line for an angular spacing of 20°. In both figures, the two vertical lines correspond to the maximal values of the temporal frequencies corresponding to angular samplings of 10° and 20° as given by (4.37). The two figures allow us to conclude that interpolating HRTFs for higher frequencies than the ones predicted by (4.37) leads to large errors while the interpolation error stays limited when obeying (4.37).

Consider now 36 measurements spaced every 10°. These measurements are used to interpolate HRTFs every 5°. The interpolated HRTFs are then compared with corresponding measurements which are available in the database and the normalized MSE is calculated following (4.52).

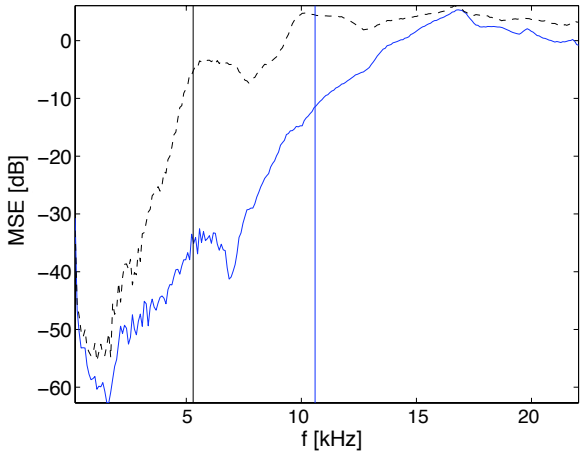
Using a spacing of 10°, interpolation of the HRTFs can only be correctly carried out up to a maximal temporal frequency of 10.8 kHz as follows from (4.37). Therefore, prior to any interpolation the HRTFs (impulse responses) are low-pass filtered using a low-pass filter with a cutoff frequency corresponding to the maximum temporal frequency associated to the specific angular spacing.

The MSE of the interpolation applied to Database 1 is shown in solid line in Fig. 4.19. The same simulation for the MSE of Database 2 is shown as the dashed line and a the full line with + for Database 3. Database 1 and 2 exhibit a MSE varying from -35 to -65 dB. The best interpolation is obtained at the positions in the proximity of -90 and 90° while slightly worse interpolation is achieved in the proximity of 0 and 180° (the angles are referenced in Fig. 4.11(b)). The MSE on the interpolation of Database 3 was slightly higher than in the simulations but still of the order of -40 dB.

These good interpolation results show that the sinc interpolation is very



(a)



(b)

Figure 4.18: Frequency dependent MSE for the interpolation of HRTF positions. (a) Simulations without considering the head (dotted lines, curves A and C) and with spherical head model (solid, curves B and D). Curves A and B are obtained with an angular spacing of $20°$ and curves C and D with $10°$. (b) Measured HRTFs for an angular spacing of $10°$ (solid) and for $20°$ (dashed). In both figures, the two vertical lines correspond to the maximum values of the temporal frequencies corresponding to angular samplings of $10°$ and $20°$ as given by (4.37).

suitable to be applied when little spatial aliasing is present or when (4.37) is satisfied.

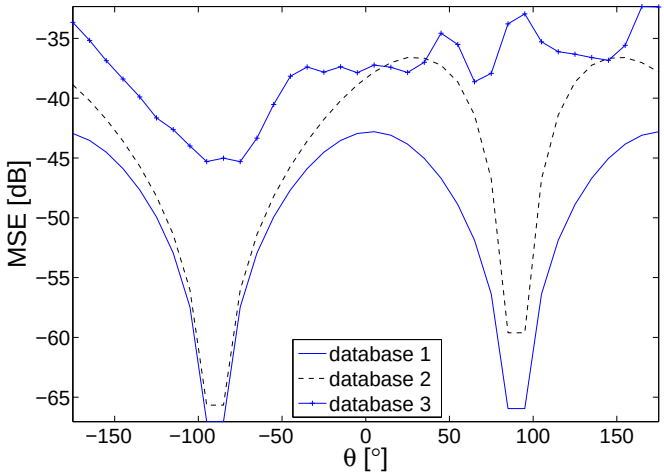


Figure 4.19: MSE on the reconstruction of HRTF Databases 1, 2 and 3 when using sinc interpolation on the low-pass signals.

4.4.2 Comparison of interpolation methods

In this section, the different Methods 1 to 4 are compared and discussed. The different methods are compared in MSE sense and the frequency dependent error is also shown for each of them. The databases used in this section contain HRTFs sampled every 5°. From these databases, we use HRTFs measured every 10° and reconstruct the dataset every 5° by means of interpolation. These interpolated HRTFs are then compared with the original HRTFs. Note that the following results are presented for the HRTFs recorded at the left ear. Therefore, for sources located in the region around −90° the ear is the ipsilateral ear, while it the contralateral ear for sources placed in the region around +90°. The performance of the different algorithms are only shown for the left ear since very similar results are obtained when considering the right ear.

Database 1

Database 1 considers the solution of the wave equation in 3D along a circular array as discussed in Section 4.2.2. The four methods have been tested and the results are presented in Fig. 4.20(a) and (b). In Fig. 4.20(a), it can be seen that Method 1 only performs well around ±90°. At the other positions, due to spatial aliasing, this technique leads to very poor results. In Fig. 4.20(b), it is shown that most of the error resulting from interpolation is due to the bad interpolation of the high frequencies.

Method 2 performs the best. The mean error is of the order of −48 dB. Since for Database 1, the HRTFs are modeled as linear phase filters the problem of

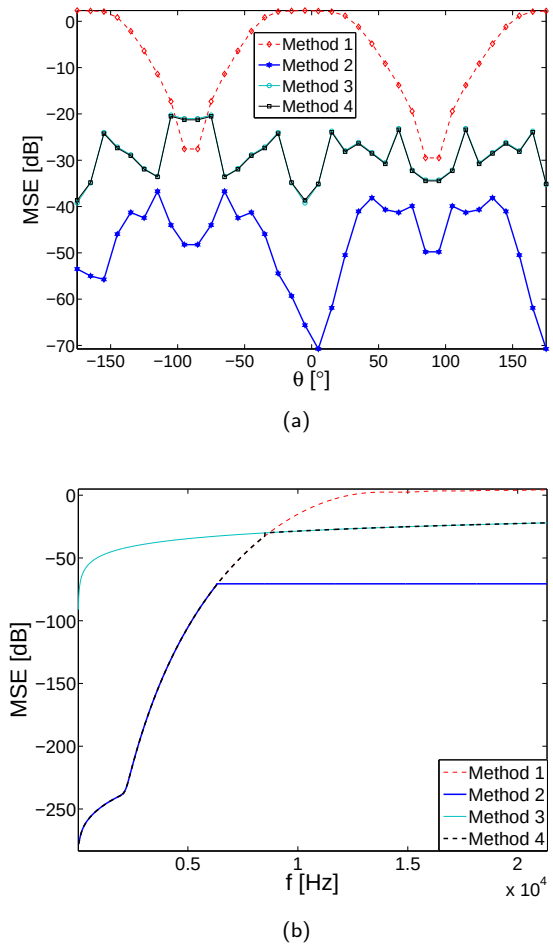


Figure 4.20: Comparison of HRTF interpolation algorithms on Database 1. (a) MSE for the different methods. (b) Frequency dependent error.

unwrapping can be well addressed. Therefore, we choose to work with narrow subbands (200 Hz). The obtained results show MSE below -40 dB at almost all angular positions. In Fig. 4.20(b), we can see that the low frequency part of the spectrum is well reconstructed but the main difference with Method 1 lies in the fact that the high frequency content of the HRTFs is well interpolated. The mean error for the higher frequencies is below -60 dB.

Method 3 aligns the different signal before interpolation using either a simple linear interpolator or a 4-tap symmetric low-pass filter. In both cases, the mean error is of about -30 dB. The frequency dependent error is shown in full lines in Fig. 4.20(b). It can be seen that the error is larger than the one presented in Method 2 but is nevertheless quite acceptable.

Method 4 considers the previous method but the low frequency part of the HRTFs is now interpolated using a sinc. It can be observed that this technique only brings little improvement. It is due to the fact that the main source of error is to be found in the higher frequencies which are similarly interpolated for Method 3 and 4. In Fig. 4.20(b) the frequency dependent error is presented in dashed bold lines.

To conclude this study for Database 1, it can be observed that the best interpolation scheme is given by Method 2.

Database 2

Database 2 considers the model by Duda et al. [42] described in Section 4.2.3. The comparison between the methods is shown in Fig. 4.21(a) and (b).

Method 1 gives similar results as in Database 1. The interpolation results are only acceptable in the region of $\pm 90°$. The frequency dependent error is very large for increasing temporal frequencies.

Method 2 gives the best overall results with a MSE over the different positions of -33 dB. In the region around $90°$ the results are worse. This is due to the fact that this region puts in evidence the head shadowing effect. The HRTFs are varying faster in this region which makes them more difficult to interpolate. The frequency dependent error shows an improvement compared to the sinc interpolation and the mean error for high temporal frequency stays limited to -30 dB.

Method 3 presents similar results as in Database 1. The MSE averaged over all angular positions is of about -20 dB. Note that the MSE gets slightly worse in the region around $-90°$.

Method 4 is very comparable to Method 3 and the overall MSE is of about -21 dB when averaged over all angular positions.

Database 2 presents a more realistic model for HRTFs interpolation. It is shown that Method 2 delivers the best results. Around the region of $-90°$ the HRTFs are interpolated with a MSE of less than -40 dB. Nevertheless, the region around $90°$ suffers poor interpolation, comparable to the one of Methods 3 and 4.

Database 3

Database 3 considers real measurements of HRTFs performed on a Kemar head. The results of the comparison are given in Fig. 4.22(a) and (b).

Method 1 presents a low MSE in the region of $\pm 90°$. The other angular positions are more poorly interpolated. The average MSE over the different

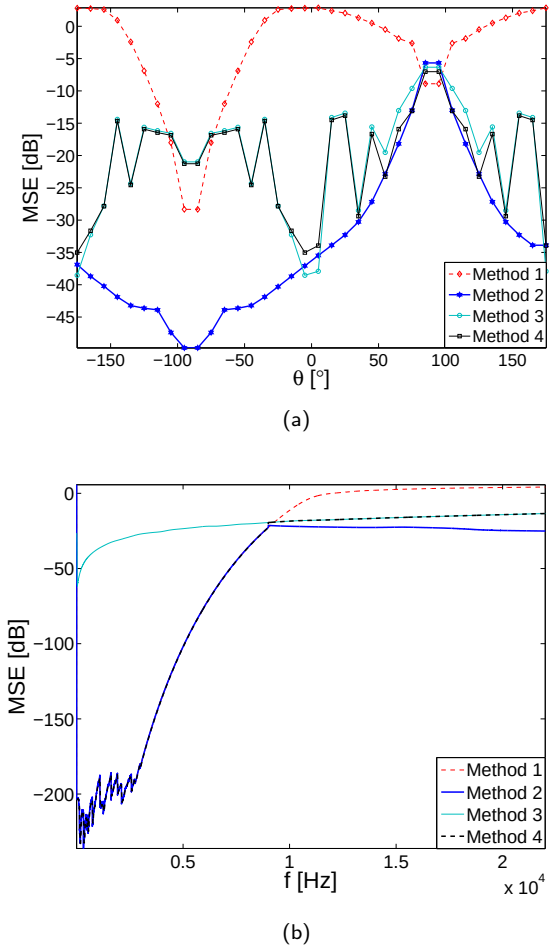


Figure 4.21: Comparison of HRTF interpolation algorithms for Database 2. (a) MSE for the different methods. (b) Frequency dependent error.

angular positions is of the order of −16dB. The frequency dependent error is shown to be large for high frequencies in Fig. 4.22(b).

Method 2 achieves the lowest average MSE of −30 dB. The best interpolation results are to be found in the region of −90°. The region around 90° is less well interpolated with a MSE of about −15 dB. Nevertheless, the frequency dependent error is the lowest of the different methods with a frequency dependent error below −20 dB for higher temporal frequencies.

Method 3 delivers interpolation results that are good except in the region of the head shadowing where the MSE increases a lot. As can be seen from Fig. 4.22(b) the frequency dependent error is quite constant over all temporal frequencies. The low frequencies are quite poorly interpolated, which explains

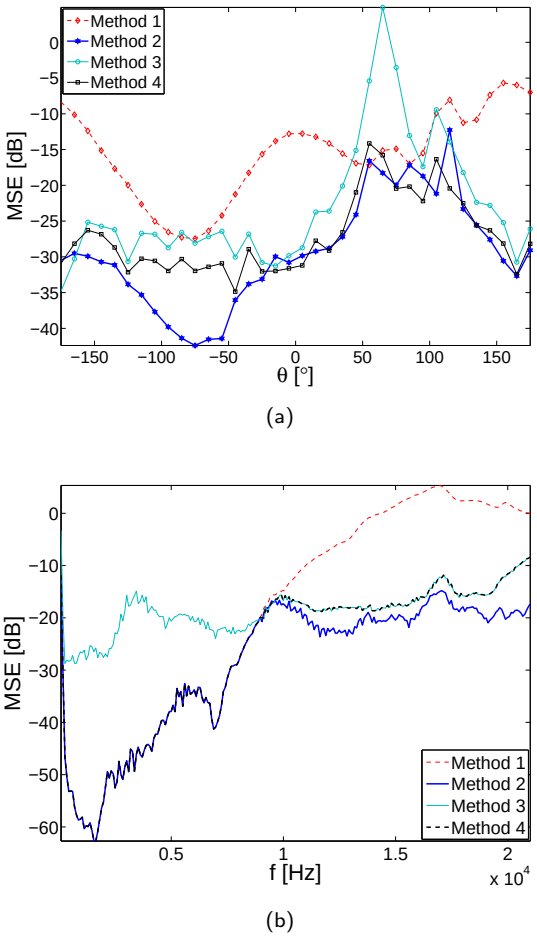


Figure 4.22: Comparison of HRTF interpolation algorithms for Database 3. (a) MSE for the different methods. (b) Frequency dependent error.

the large errors in the region of 90°.

In this scenario, using the sinc interpolation for the low frequencies allows Method 4 to achieve quite good results similar to those of Method 2. The region around 90° is a bit less well interpolated but overall the HRTFs are interpolated with an average MSE of −27 dB.

Comparing the different methods leads to the conclusion that Method 2 performs slightly better than Method 4 especially in the region around −90°. For the other regions, similar results are obtained. Nevertheless, this is an interesting finding since Method 4 is quite simple and still leads to good interpolation results.

For completeness, we have considered the case of interpolation of HRTFs

	Method 1	Method 2	Method 3	Method 4
Database 1	-5.5	-47.9	-29	-29.1
Database 2	-3.5	-33.64	-20.3	-20.7
Database 3	-15.65	-29.5	-23	-27.2
Database 3b	-6.2	-19.7	-16	-18.4

Table 4.1: Table summarizing the different MSE averaged over all interpolated positions for the different methods and databases.

every 20° to obtain HRTFs every 10°. The MSE is given in Fig. 4.23 for the 4 considered methods. The four methods behave very similarly to the case where

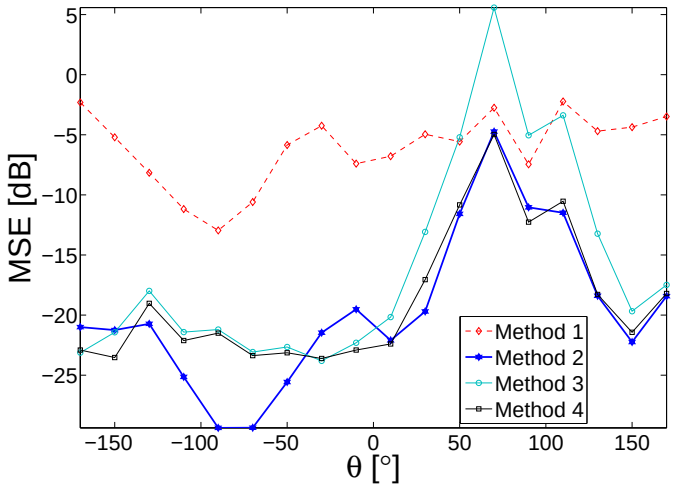


Figure 4.23: Comparison of HRTFs interpolation methods in the case of a spacing of 20° in the database.

HRTFs are interpolated from a spacing of 10°. The best total MSE is obtained for Method 2 with an averaged MSE of −20 dB. The Method 4 which is simpler in processing achieves a slightly worse MSE of −18.4 dB. Method 1 performs quite poorly with a MSE not exceeding −7 dB. Method 3 shows similar MSE than Method 4 except in the region around 90° where it performs badly.

4.4.3 Discussion

We have presented an overview of different databases tested with different interpolation methods. To summarize the results, Table 4.1 contains for each database and each method, the MSE averaged over all interpolated positions. Note that Database 3b considers the Database 3 where the HRTFs are present only every 20° as described for Fig. 4.23.

Method 2 performs the best of the different methods. In the measurements, it has been shown that the simpler Method 4 also performs quite well. The general trend present in these results shows that for sources that see the ear as the contralateral ear (source located on the other side of the ear, e.g. +90° for the left ear) interpolation is worse than in the other angular positions. On the other hand, it is known that the head shadowing effect creates an attenuation of approximatively 20 dB on the perceived sound [20]. When localizing a sound, the brain uses the signals of the two ears. For a source located at +90°, the HRTF from the source to the left ear has a low magnitude and contains the head shadowing effect while the right ear gets a very direct HRTF with higher magnitude. When interpolation of HRTFs is applied, it is likely that the localization of the source in this scenario is still achievable even when interpolation of the weak signal (left ear) is poor while being very precise for the strong signal (right ear). This topic is matter of current research and perceptual tests need to be conducted to study this scenario in more detail.

4.5 Conclusion

The sound field has been studied along a circle. The angular bandwidth of the sound has been studied in two and three dimensions and a sampling theorem has been presented to quantify the aliasing error as a function of the angular sampling frequency. Based on this study, an HRTF interpolation algorithm has been proposed. At low frequencies, where the spatial Nyquist theorem indicates that the given HRTFs can be interpolated very precisely, spatial interpolation on circles is applied. At higher frequencies, where the Nyquist theorem indicates that not enough information for precise interpolation is available, the spatial interpolation is carried in the complex temporal envelope domain in subbands to avoid aliasing. The interpolated subbands are obtained by restoring the carrier after interpolating the complex envelopes. Numerical simulations carried out with HRTF models and measured data indicate that the proposed method performs better than previous HRTF interpolation methods in a mean square sense.

Chapter 5

Dynamic Measurement of Room Impulse Responses

5.1 Introduction

One of the key elements in the success of multimedia communication systems is the capability of reproducing virtual environments to be perceived by the listener as naturally as possible. To achieve realism, either real measurements need to be introduced in the system, or the environment needs to be modeled and synthetically generated [119, 82]. In both cases, room impulse responses (RIRs) are used to give listeners the impression of being in a real environment and enveloped by the sound. In the case of headphone playback, head-related transfer functions (HRTFs) are usually added to achieve externalization and more realistic impressions. The RIRs and HRTFs to be used need to be either measured at a large number of positions in the considered space or modeled using different techniques such as the ones described in Chapter 2. The RIRs are then provided to the playback system. This is typically the case in wave field synthesis (WFS) systems where measured RIRs at a large number of positions [40] are used to increase the realism of the reproduced field. Therefore, one would like to find a way to easily and rapidly measure large sets of RIRs. The usual technique is to use a single microphone or a microphone array. Nevertheless, to capture hundreds of RIRs, the array of microphones needs to be displaced to several positions. The intrusion of a person to modify the setup (e.g. displace the array) changes greatly the characteristics of the room and the temperature field inside of the room [45] and makes the measurement very time consuming. In this chapter a technique is introduced that achieves a fast recording of a large number of RIRs. We consider two possible setups, the first

setup considers a fixed loudspeaker with a moving microphone and the second a moving loudspeaker with a fixed microphone. In both setups, the moving element follows a trajectory (e.g. circular trajectory) with constant speed. Also, the acquisition of the data is not done position after position but happens continuously along the trajectory. The movement is uniform and does not stop during the acquisition. Thanks to this setup, one avoids problems linked to abrupt stops leading to oscillations and waiting time for the microphone (or loudspeaker) to get its position. From the one-dimensional signal gathered by the microphone, the two-dimensional (2D) dataset (spatial and temporal) containing the RIRs at all the different spatial positions along the trajectory is reconstructed. The algorithm takes into account the Doppler shift inherent to the moving element and cancels its effect in the reconstructed RIRs. An analysis is performed to study the influence of the different parameters to be chosen so as to achieve the reconstruction of the RIRs (e.g. length and period of the excitation signal, frequencies contained in the excitation signal, speed of movement of the microphone, temporal and spatial frequencies of the reconstructed dataset). The trade-off existing between the speed of the moving element and the spacing between the frequencies contained in the excitation signal is discussed[1]. The presented theory is shown together with measurements. These measurements are obtained using a moving microphone holder that achieves a precision of a few hundredths of a degree when rotating in the horizontal plane. An interesting application of this setup can be found in the measurement of head-related transfer functions (HRTFs). These measurements are typically done in anechoic chambers to describe the influence of our body on the sound measured at the entrance of our ears as was previously described in Chapter 4. For these specific filters, the impulse responses to be measured are very short (on the order of a few milliseconds) since no room reflections need to be captured. The whole dataset of azimuthal angles can then be recorded in a very fast manner. In this chapter, we show that with the presented technique, measurement of all HRTFs in the horizontal plane can be achieved in less than 1 second.

5.1.1 Related work

As described previously in Chapter 2, different methods exist for the measurement of RIRs and HRTFs. In the literature, the measurements mostly happen using a fixed setup. Note that in [40, 59] room impulse responses measurements were carried out using a microphone moving along a circular trajectory.

[1] Remark that in this work, the turbulence of air due to the moving element is not taken into account in the analysis.

Nevertheless, the assumption was done that at low angular speed, the RIRs can be obtained without taking the Doppler effect into account. This is not the case in our study where the RIRs are theoretically perfectly reconstructed. Other methods using moving microphones have been presented for improving wave field extrapolation techniques. As was described in Chapter 2, the first step is to measure the field along a certain trajectory, and from the gathered information, extrapolation can be performed using the Kirchhoff-Helmholtz integral [18]. Considering the field to be extrapolated as stationary, the different measurements do not need to be performed simultaneously. In [68, 76], it is explained that using a moving microphone allows for measurement of the field in more positions and therefore better extrapolation can be achieved.

5.1.2 Contributions

The contributions of this work are the following:

- The Doppler effect is put in evidence in the two-dimensional spectrum representation of the sound pressure field when a moving source or receiver is used in the setup.
- An algorithm for the reconstruction of the static RIRs or HRTFs is presented by removing the Doppler effect present in the recording.
- A formula describing the maximal rotation speed is given in function of the different parameters of the system.
- The measurement of HRTFs at every azimuthal angle along the horizontal plane can be achieved in less than one second at a sampling frequency of 44.1 kHz.
- Experimental measurements using a moving microphone demonstrate the applicability of the algorithm and confirm the presented theory.

5.1.3 Outline

The outline of the chapter is as follows. In Section 5.2, we consider a first setup with a microphone moving along a trajectory with a uniform speed. In this scenario, the Doppler effect needs to be considered. The Doppler effect is briefly reviewed in Section 5.2.1 and it is shown in Section 5.2.2 that using the 2D spectrum representation of the sound field along a line, the Doppler effect can be put in evidence. The signal recorded by a microphone moving along a circle is then described in Section 5.2.3. The second setup considering a fixed microphone and a moving loudspeaker is explored in Section 5.3. The microphone signal is studied for a linear and a circular trajectory of the loudspeaker

in Section 5.3.1 and Section 5.3.2, respectively. Section 5.4 presents the main result of this work. It develops an algorithm to reconstruct the RIRs at every possible position, from the gathered signal by the microphone. The technique to achieve this reconstruction is presented for the two setups in Section 5.4.1 and 5.4.2. Some further remarks are given in Section 5.4.3. The theory is then compared with experimental measurements in Section 5.5. The conclusions are drawn in Section 5.6.

5.2 Moving Microphone Signal

In Chapters 3 and 4, the sound field has been studied along different spatial geometries. In this chapter, we consider two specific trajectories: a line and a circle. For both cases, the sound field has been shown to have a butterfly spectrum. This specific shape of the spectrum will be used to study the sound field gathered by a microphone in the presence of a moving element in the setup. In this section, the setup considers a fixed loudspeaker and a microphone moving along a trajectory. The dual setup will be analyzed in detail in Section 5.3. The reason to analyze these two setups separately is due to the different approach and analysis performed in the two cases. Also, as will be reminded in Section 5.2.1, the Doppler effect appearing due to the movement of an element in the setup is different in the case of a moving source or a moving microphone.

First, the Doppler effect is reviewed in Section 5.2.1. Further, the moving microphone signal is studied when the microphone is moved along a line in Section 5.2.2 and along a circle in Section 5.2.3.

5.2.1 Doppler effect

When considering a moving element, a frequency shift is observed on the recorded signal. This is known as the Doppler effect and can be expressed by the following formula[2]:

$$\omega = \omega_0 \left(\frac{c + v_r}{c + v_s} \right), \tag{5.1}$$

with c the speed of sound propagation, ω the observed frequency, ω_0 the emitted frequency, v_s the speed of movement of the source and v_r the speed of movement of the receiver. A convention for the sign of the speeds in (5.1) needs to be adopted. The positive direction is considered as the direction "listener towards source". Therefore, using this convention, the sign of the speed of the receiver is

[2]The formula is valid for the case of a movement along the same direction as the source-receiver axis.

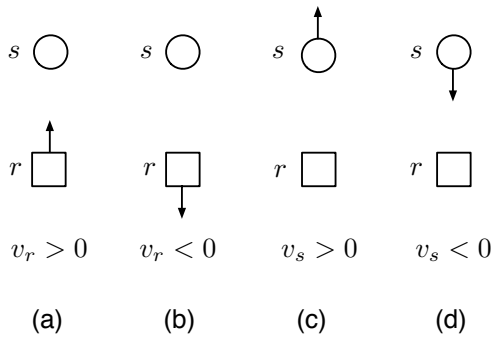


Figure 5.1: Different setups for the study of the Doppler effect.

considered as positive when the source and the receiver are coming towards each other as shown in Fig. 5.1(a). The sign of the received speed is negative when source and receiver are moving apart as in Fig. 5.1(b). The case of the moving source is analyzed in the next section. There, a source moving away from the receiver leads to a positive speed in (5.1) while when the source approaches the receiver its speed is to be considered as negative. These two situations are shown in Fig. 5.1(c) and (d), respectively. From (5.1), it is therefore observed that when source and receiver come closer to each other, the recorded frequency is increased and when the source and receiver are moving apart, the recorded frequency decreases.

5.2.2 Microphone moving along a line

The pressure recorded at all positions along a line trajectory is denoted as $p(x,t)$ as discussed in Chapter 3. The two-dimensional Fourier transform (2D-FT) of this signal is $\hat{p}(\phi,\omega)$. Consider a setup with a fixed source and a microphone moving with a constant speed of v m/s. Due to the constant speed, we have that the position x along the line is proportional to the time as

$$x = vt. \tag{5.2}$$

At each position the sound recorded by the microphone is the convolution of the source signal denoted as $s(t)$ with the corresponding RIR to be found in the RIRs database denoted as $h(x,t)$. Therefore,

$$p(x,t) = \int_{-\infty}^{\infty} s(i)h(x,t-i)\mathrm{d}i. \tag{5.3}$$

The sound recorded by the moving microphone is denoted as $r(t)$ with

$$r(t) = p(vt, t). \tag{5.4}$$

Note that (5.4) corresponds to a slice of the function $p(x, t)$ along a line of equation (5.2). The spectrum of this recorded sound can be calculated as follows:

$$\hat{r}(\gamma) \quad = \quad \int_{-\infty}^{\infty} p(vt, t) e^{-j\gamma t} \mathrm{d}t. \tag{5.5}$$

Also remark that

$$p(vt, t) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{p}(\phi, \omega) e^{j(\omega t + \phi v t)} \mathrm{d}\phi \mathrm{d}\omega. \tag{5.6}$$

Using (5.6), Expression (5.5) can be rewritten as:

$$\begin{aligned} \hat{r}(\gamma) &= \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{p}(\phi, \omega) e^{-jt(\gamma - v\phi - \omega)} \mathrm{d}t \mathrm{d}\phi \mathrm{d}\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{p}(\phi, \omega) \delta(\gamma - v\phi - \omega) \mathrm{d}\phi \mathrm{d}\omega \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{p}(\phi, -v\phi + \gamma) \mathrm{d}\phi. \end{aligned} \tag{5.7}$$

For each frequency γ, the value of the spectrum of the recorded signal is obtained by projection of the 2D spectrum $\hat{p}(\phi, \omega)$ following the direction $\vec{\mathbf{v}} = (1, -v)$ on the ω axis. This result is known as the *projection-slice theorem*: a slice in the time domain corresponds to a projection in the frequency domain [25]. This construction is presented in Fig. 5.2 and 5.3 where the Doppler effect is also put in evidence. Consider a plane source emitting a plane wave arriving on one microphone line with angle $\alpha = 0°$ as shown in Fig. 5.2(a). Consider further a moving microphone along the infinite line. The microphone is moving with a constant speed $|v|$ away from the plane source in the positive x direction. As seen in Section 5.2.1, the movement leads to a frequency shift that lowers the perceived frequency at the microphone. The receiver signal can be obtained by projection of $\hat{p}(\phi, \omega)$ along the direction $\vec{\mathbf{v}}$. As the sound pressure field is a plane wave with an arriving angle of $\alpha = 0°$, it has been shown in Section 3.3.1 that its 2D spectrum is the line of equation

$$\omega = -c\phi. \tag{5.8}$$

This spectrum is shown in bold in Fig. 5.2(b). The component of $\hat{p}(\phi, \omega)$ at frequency ω_0 is simply the point $\left(-\frac{\omega_0}{c}, \omega_0\right)$ in Fig. 5.2(b). The projection of

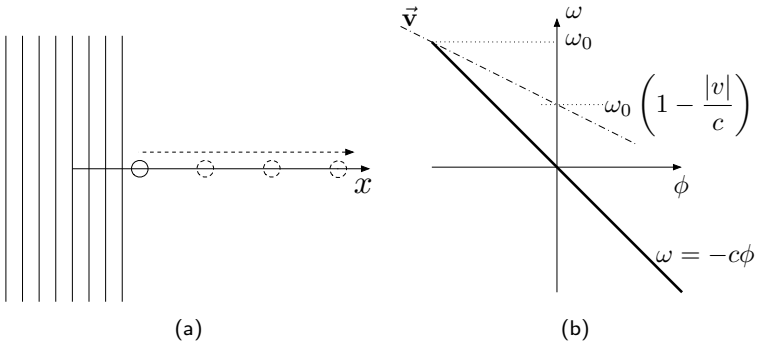


Figure 5.2: Doppler effect with a receiver moving away from the source. (a) Schematic view of the situation. (b) Analysis of the situation in the 2D-FT domain.

this point on the ω axis happens at

$$\omega = \omega_0 \left(1 - \frac{|v|}{c}\right). \tag{5.9}$$

This is exactly the result obtained when considering the Doppler effect (5.1) for a receiver moving away from the source. The speed of the receiver is $v_r = -|v|$ since we are in the case of Fig. 5.1(b).

Similarly, Fig. 5.3(a) presents the situation where the receiver is moving towards the source along the negative x direction. As can be seen in Fig. 5.3(b), the projection of the point $\left(-\frac{\omega_0}{c}, \omega_0\right)$ on the ω axis is now

$$\omega = \omega_0 \left(1 + \frac{|v|}{c}\right). \tag{5.10}$$

This corresponds to the Doppler effect for a receiver moving towards the source.

5.2.3 Microphone moving along a circle

In this section, a source emits sound and a receiver is moving along a circle. The pressure measured at the different positions along the circle is given by $p(\theta, t)$. The sound recorded by the receiver moving with an angular speed of v rad/s [3] is $r(t) = p(vt, t)$. Similarly to (5.7), it can be derived that

$$\hat{r}(\gamma) = \sum_{l_\theta=-\infty}^{\infty} \hat{p}(l_\theta, -vl_\theta + \gamma). \tag{5.11}$$

[3]Note that in the case of a circular trajectory the speed is an angular speed and has units rad/s while along a line the speed v has units m/s. We chose to keep the same letter for simplicity of notation.

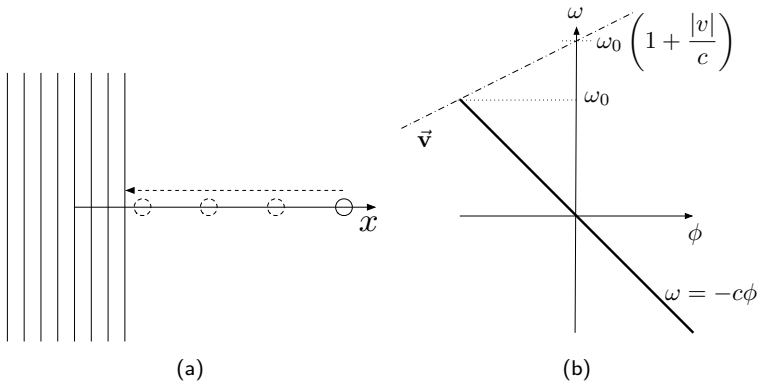


Figure 5.3: Doppler effect with a receiver moving towards the source. (a) Schematic view of the situation. (b) Analysis of the situation in the 2D-FT domain.

For completeness, note that when the source emits a periodic signal $s(t)$ with period $\frac{2\pi}{v}$, $r(t)$ becomes also periodic with the same period. In that case, we denote $r(l_\gamma)$ the Fourier series of $r(t)$ with

$$\hat{r}(l_\gamma) = \sum_{l_\theta=-\infty}^{\infty} \hat{p}(l_\theta, -vl_\theta + l_\gamma). \tag{5.12}$$

5.3 Recorded Signal for a Moving Loudspeaker

The second setup is now considered. A fixed microphone is recording the sound emitted by a source moving along a trajectory. The cases considering a linear or circular trajectory are described in Section 5.3.1 and 5.3.2, respectively.

5.3.1 Loudspeaker moving along a line

Consider a source moving along a line with speed v m/s and a receiver recording the signal at a fixed position. The purpose of the analysis is to obtain from the microphone signal $r(t)$ all the RIRs from the different positions of the source to the fixed microphone. The microphone signal can be expressed as follows:

$$r(t) = \int_{-\infty}^{\infty} s(i)h(vi, t-i)\mathrm{d}i, \tag{5.13}$$

where $s(t)$ corresponds to the emitted sound and $h(x,t)$ corresponds to the different static impulse responses between the source and the receiver. These impulse responses are functions of the time t and are different at each abscissa x of the source. Now, similarly to the analysis performed in Section 5.2.2, the spectrum of the gathered signal can be calculated. The analysis makes again

use of the projection-slice theorem [25]. This time, a projection in the time domain corresponds to a slice in the frequency domain. For the analysis, we rewrite (5.13) as follows

$$r(t) = \int_{-\infty}^{\infty} p(vi, t-i)\mathrm{d}i, \tag{5.14}$$

with

$$p(vi, t-i) = s(i)h(vi, t-i). \tag{5.15}$$

The Fourier transform of $r(t)$ is denoted as $\hat{r}(\gamma)$ with

$$\hat{r}(\gamma) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(vi, t-i)e^{-j\gamma t}\mathrm{d}i\mathrm{d}t, \tag{5.16}$$

Also,

$$p(vi, t-i) = \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \hat{p}(\phi, \omega)e^{j(\omega(t-i)+\phi vi)}\mathrm{d}\phi\mathrm{d}\omega. \tag{5.17}$$

Replacing (5.17) in (5.16) leads to

$$\begin{aligned}
\hat{r}(\gamma) &= \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \hat{p}(\phi, \omega)e^{j(\omega(t-i)+\phi vi-\gamma t)}\mathrm{d}i\mathrm{d}t\mathrm{d}\phi\mathrm{d}\omega \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \hat{p}(\phi, \omega)\delta(\phi v-\omega)\delta(\omega-\gamma)\mathrm{d}\phi\mathrm{d}\omega \\
&= \frac{1}{v}\int_{-\infty}^{\infty} \hat{p}\left(\frac{\omega}{v}, \omega\right)\delta(\omega-\gamma)\mathrm{d}\omega \\
&= \frac{1}{v}\hat{p}\left(\frac{\gamma}{v}, \gamma\right).
\end{aligned} \tag{5.18}$$

We therefore see that a projection in the spatio-temporal domain corresponds to a slice of the spectrum. The slice in the frequency domain is applied on the 2D-FT of $p(x,t)$. From (5.15) it is observed that

$$p(x,t) = s\left(\frac{x}{v}\right)h(x,t). \tag{5.19}$$

Therefore $\hat{p}(\phi,\omega)$ can be written as

$$\hat{p}(\phi,\omega) = v\hat{s}(v\phi) * \hat{h}(\phi,\omega), \tag{5.20}$$

where the convolution happens along the spatial frequency direction. Using

(5.20), (5.18) can be rewritten as

$$\hat{r}(\gamma) = \hat{s}(\gamma) * \hat{h}\left(\frac{\gamma}{v}, \gamma\right). \tag{5.21}$$

As an example, consider that the excitation signal is $s(t) = e^{j\omega_0 t}$, then

$$\hat{p}(\phi, \omega) = v\delta\left(v\phi - \omega_0\right) * \hat{h}(\phi, \omega). \tag{5.22}$$

Consider in this example a receiver r recording the signal emitted by a source s moving away with a speed $|v|$ as shown in Fig. 5.4(a). In that case, the original 2D-FT of the pressure recorded by the receiver has a support given by $\omega = -c\phi$ as shown by the dotted line in Fig. 5.4(b). In the current example, the spatial convolution in (5.22) leads to a spectrum whose support is located on the full line $\omega = -c\phi + c\frac{\omega_0}{|v|}$. The slice of the new spectrum with the dashed line, $\omega = |v|\phi$, leads to a value of the frequency, recorded by the microphone, of

$$\omega_0 \frac{c}{c + |v|}. \tag{5.23}$$

This exactly corresponds to the frequency predicted by the Doppler effect formula in the case of a source moving away from the receiver with a speed $v_s = |v|$ as shown in Fig. 5.1(c).

When the source is moving with a speed v towards the receiver, the speed $|v|$ is replaced by $-|v|$ in the equations and the observed frequency by the receiver is now

$$\omega_0 \frac{c}{c - |v|}. \tag{5.24}$$

The setup is shown in this case in Figs. 5.5(a) and (b).

5.3.2 Loudspeaker moving along a circle

Similarly to the approach presented in Section 5.3.1, the case of the loudspeaker moving along a circular trajectory is studied now. $p(\theta, t)$ represents the sound field measured at the microphone for a source located at an angle θ. Similarly to (5.19),

$$p(\theta, t) = s\left(\frac{\theta}{v}\right) h(\theta, t), \tag{5.25}$$

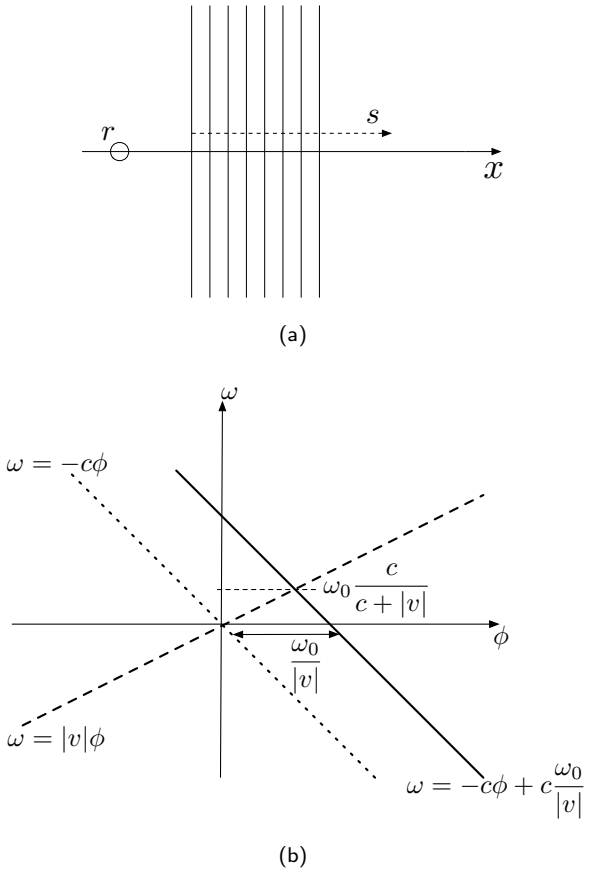


Figure 5.4: Doppler effect with a source moving away from the receiver. (a) Schematic view of the situation. (b) Analysis of the situation in the 2D-FT domain.

with $h(\theta, t)$ the different RIRs. The same theory is still applied and the Fourier transform of the signal gathered at the microphone is given by

$$\hat{r}(\gamma) = \hat{p}(\frac{\gamma}{v}, \gamma). \tag{5.26}$$

When $s(t)$ is periodic with period $\frac{2\pi}{v}$, the signal $r(t)$ becomes also periodic with the same period. In that case, it can be shown that

$$\hat{r}(l_\gamma) = \sum_{l_\theta=-\infty}^{\infty} \hat{p}(l_\theta, v l_\theta)\delta_k(v l_\theta - l_\gamma), \tag{5.27}$$

with $\delta_k(\cdot)$ describing a Kronecker symbol.

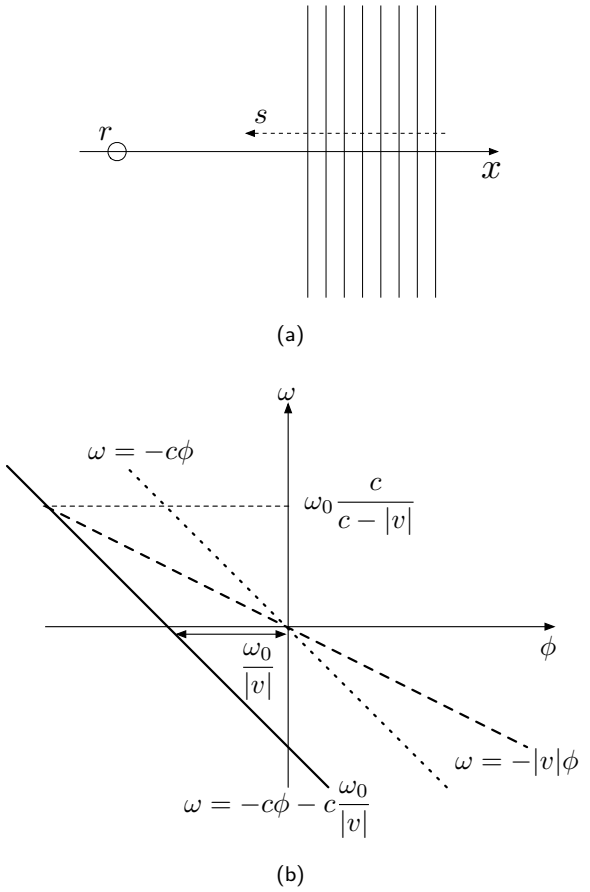


Figure 5.5: Doppler effect with a source moving towards the receiver. (a) Schematic view of the situation. (b) Analysis of the situation in the 2D-FT domain.

5.4 Spatio-Temporal Reconstruction of Room Impulse Responses

This section presents the main result of this work. Considering a moving element (microphone or loudspeaker) in the setup, the purpose is to recover the different RIRs at any position along the trajectory from the recording of the microphone. The different aspects of the technique are explored in Section 5.4.1 for a moving microphone and in Section 5.4.2 for a moving loudspeaker. The relation between the speed of movement and the spacing between the frequency components of the excitation signal is specified. Section 5.4.3 discusses different remarks related to the presented techniques and compares both setups. Application of the technique to HRTF measurements is described. Note that the technique is presented for the case of a movement along a circle but can

easily be applied for the linear trajectory or other trajectories.

5.4.1 Reconstruction algorithm for a moving microphone

With the knowledge of the emitted sound, the purpose of this technique is to reconstruct the different RIRs at any angle from the recording of the moving microphone. The speed of the moving microphone will be shown to be the key factor in the possible reconstruction of the RIRs at any possible angle. For this purpose, the 2D Fourier representation is used. Fig. 5.6 shows the magnitude of $\hat{p}(l_\theta, \omega)$ representing the spectrum of the different static signals gathered at the different angular positions along the circle. To apply our algorithm of reconstruction, we need to impose that not all temporal frequencies are present in the emitted signal. Energy is present only for the temporal frequencies shown as dashed lines in the spectrum $\hat{p}(l_\theta, \omega)$. The signal that is recorded by the microphone is given by (5.11). As described in Section 5.2, the 2D spectrum is projected following the direction $\vec{\mathbf{v}}$ on the ω axis. To be able to reconstruct the different RIRs, the emitted signal by the source has to be such that all the lines containing energy in the spectrum $\hat{p}(l_\theta, \omega)$ do not overlap once projected on the ω axis. Consider the maximal frequency emitted by the source to be

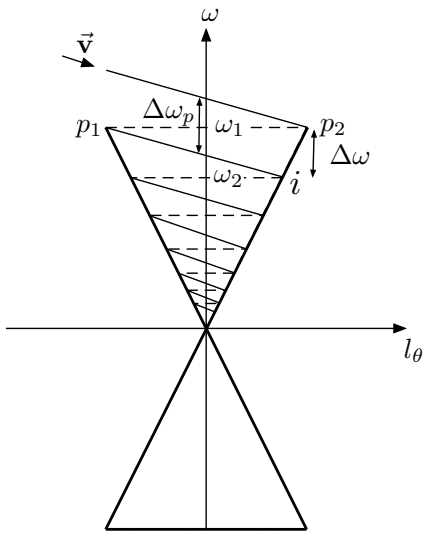


Figure 5.6: Projection of the sound pressure field for an emitting sound containing frequencies spaced following (5.31).

ω_1. In the 2D spectrum, this frequency component corresponds to the segment $|p_1 p_2|$. When projected on the ω axis, new frequency components appear in the range $\omega \in [\omega_1 - 1/2\Delta\omega_p, \omega_1 + 1/2\Delta\omega_p]$ due to the Doppler shift. To avoid any overlapping in the projections, the next frequency component emitted by

the source has to be chosen carefully. Denote by i the intersection point of the projection of the point p_1 with the butterfly spectrum. To obtain the ordinate of i, denoted as ω_2, recall that the minimal slope of the triangular spectrum is given by

$$\omega = \frac{c}{r} l_\theta. \tag{5.28}$$

The frequency ω_2 is obtained by solving a system representing the intersection between the segment $|p_1 i|$ and one side of the butterfly spectrum:

$$\begin{cases} \omega = -v l_\theta + \omega_1 (1 - \frac{rv}{c}) \\ \omega = \frac{c}{r} l_\theta. \end{cases} \tag{5.29}$$

The solution of this system leads to the value $\omega_2 = \omega_1 \left(\frac{c-rv}{c+rv}\right)$. Defining $\Delta\omega$ as the frequency spacing between the two consecutive temporal frequency components ω_1 and ω_2 emitted by the source, it can be sown

$$\Delta\omega = \omega_1 \frac{2vr}{c + rv}. \tag{5.30}$$

As can be seen in Fig. 5.6, the spacing allowed between successive temporal frequencies is diminishing for smaller temporal frequencies. Therefore, the excitation signal could be very dense at low frequencies. It can be shown that the frequencies emitted by the sources need to satisfy the following relation

$$\omega_{i+1} = \omega_1 \left(\frac{c - rv}{c + rv}\right)^i. \tag{5.31}$$

Nevertheless, to ease the calculation and make the processing simpler, a constant spacing $\Delta\omega$ between the temporal frequencies of the excitation signal is chosen, as presented in Fig. 5.7. The excitation signal is therefore a periodic signal with period $T_S = \frac{2\pi}{\Delta\omega}$. Remark that although the excitation signal only contains discrete frequencies, it is possible to determine the RIRs at every frequencies thanks to the sampling theorem in the frequency domain [89]. Reconstruction at all frequencies is possible when considering the RIR to be of finite length. The duration of the RIR to be recorded, denoted as T, has to be smaller than the sampling period T_S to avoid any temporal aliasing, i.e.

$$T < \frac{\pi(c + rv)}{rv\omega_1}. \tag{5.32}$$

To record RIRs of length T, the maximal speed at which the rotation can be

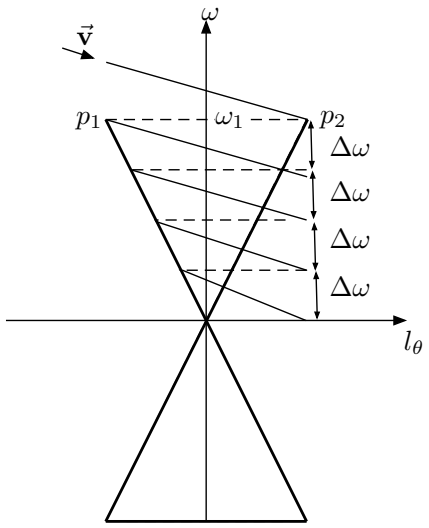


Figure 5.7: Projection of the sound field with an emitting signal containing frequencies spaced every $\Delta\omega$.

applied is obtained using (5.32):

$$v_{\max} = \frac{\pi c}{r(\omega_1 T - \pi)}. \tag{5.33}$$

5.4.2 Reconstruction algorithm for a moving loudspeaker

In case of a moving source, it has been shown that the signal gathered by the microphone is expressed by (5.13). In the frequency domain, this corresponds to the slicing of a function $\hat{p}(\phi, \omega)$ along the line $\omega = v l_\theta$. The function $\hat{p}(\phi, \omega)$ corresponds to the original 2D spectrum of the RIRs convolved with the spectrum of the emitting signal along the angular direction. In order to ensure that the slice is not cutting any part of the spectrum containing aliasing due to overlap, there needs to be some minimal spacing between the emitted frequencies. The emitted signal needs to contain frequency values ω_i satisfying

$$\omega_{i+1} = \omega_1 \left(\frac{c - rv}{c + rv} \right)^i, \tag{5.34}$$

where ω_1 stands for the maximal frequency present in the excitation signal as shown in Fig. 5.8(a). Note that this relation is exactly the same as for the moving microphone, see (5.31). For simplicity, the emitting signal $s(t)$ is again considered to be periodic and therefore its Fourier transform contains temporal frequency lines evenly spaced by $\Delta\omega$. In this case, the new setup is presented

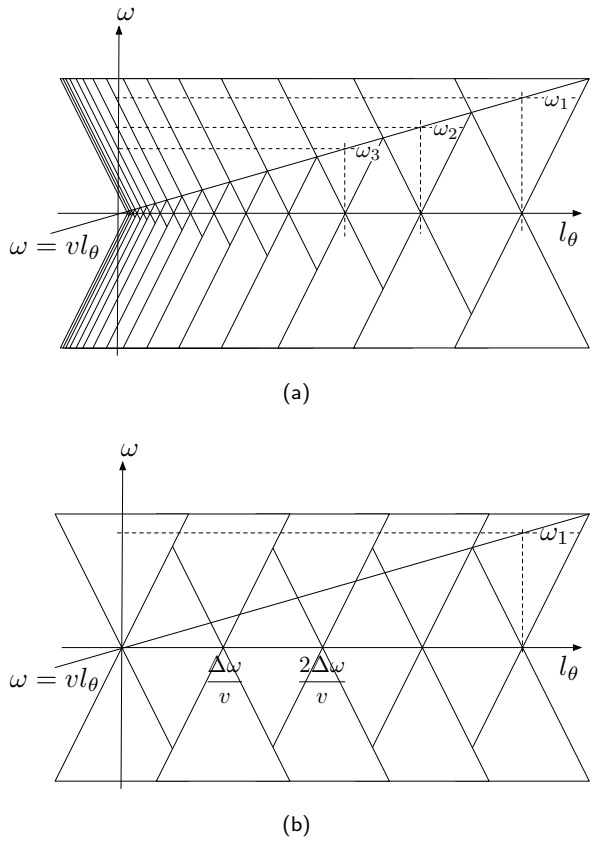


Figure 5.8: Fourier analysis of the sound gathered by a moving source. (a) The frequencies of the emitted signal are ω_i. The emitted signal is not periodic. (b) Simplification of the problem where the emitted signal is periodic.

in Fig. 5.8(b). The condition to avoid any slice containing aliasing is

$$\frac{\Delta\omega}{v} \geq \frac{2\omega_1 r}{c + rv}. \tag{5.35}$$

The signal gathered by the microphone has been shown to have support along the slice $\omega = vl_\theta$ as seen in Fig. 5.8(b). It is possible to modify this figure and to transform the slice crossing several spectral repetitions of the butterfly spectrum into multiple slices crossing the same butterfly spectrum as shown in Fig. 5.9(a). All the slices are parallel and have a slope along the vector $\vec{\mathbf{v}}_1 = (1, v)$. This slicing of the 2D spectrum can be written as:

$$\sum_{n=-\infty}^{\infty} \hat{p}(l_\theta, \omega)\delta(\omega - vl_\theta - n\Delta\omega). \tag{5.36}$$

Calculating the inverse Fourier transform of Eq. (5.36) leads to

$$\frac{T_S}{2\pi} \sum_{n=-\infty}^{\infty} p(\theta + vnT_S, t - nT_S). \tag{5.37}$$

As follows from Eq. (5.37), the slicing of the butterfly spectrum corresponds to a periodic repetition in the angular-time domain as shown in Fig. 5.9(b). The periodic repetition happens following the direction to $\vec{\mathbf{v}}_2 = (v, -1)$.

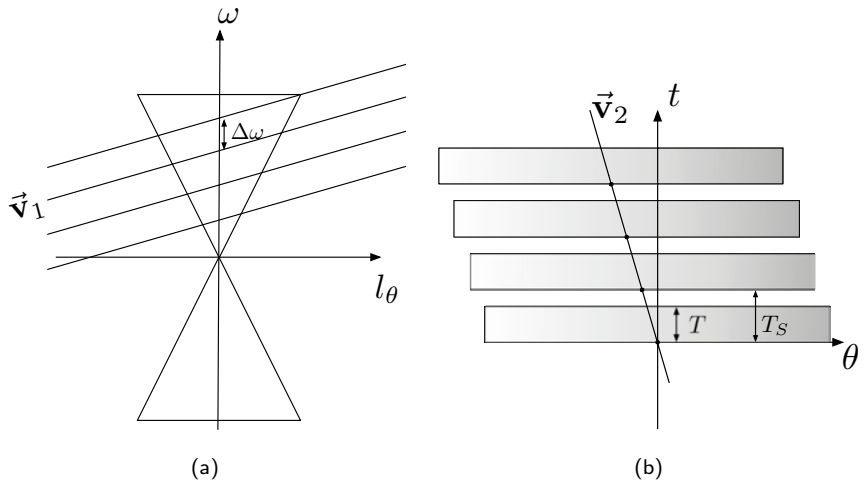


Figure 5.9: Moving source emitting periodic sound. (a) Butterfly spectrum sliced by multiple lines along the direction $\vec{\mathbf{v}}_1$. (b) Angular-time domain representation of the periodic repetitions due to the slicing of the butterfly spectrum in the frequency domain.

Considering the signal $p(\theta, t)$ to have a support in the angular-time domain ranging in $([-\pi, \pi] \times [0, T])$, it can be derived from Eq. (5.37) that the condition to avoid temporal aliasing is $T < T_S$. Similarly to the results obtained in Section 5.4.1, one finds

$$T < \frac{2\pi}{\Delta\omega} = \frac{\pi(c + rv)}{rv\omega_1}. \tag{5.38}$$

The maximal speed for the moving loudspeaker is therefore identical to the one for the moving microphone:

$$v_{\max} = \frac{\pi c}{r(\omega_1 T - \pi)}. \tag{5.39}$$

5.4.3 Discussion

To summarize, a description of the presented technique is given here. An excitation signal is transmitted to a loudspeaker. The emitted signal cannot contain every frequency but has to be designed carefully as discussed in Section 5.4.1 for the moving microphone and in Section 5.4.2 for the moving loudspeaker. It is observed that in both cases, the maximal speed of the moving element follows the same rule, namely Equation (5.33). After recording, the Fourier transform of the microphone signal is taken. In the case of the moving microphone this spectrum corresponds to $\hat{r}(l_\gamma)$ in (5.12). As discussed in Section 5.4.1, it corresponds to a projection of $\hat{p}(l_\theta, \omega)$ on the ω axis. This projection needs to be undone to recreate the 2D spectrum. This can be achieved at the condition that no overlapping was present in Fig. 5.6. This is assured when (5.33) is satisfied. In the case of the moving loudspeaker, the Fourier transform of the microphone signal $\hat{r}(l_\gamma)$ is given in (5.27). After rearranging the spectrum as described in Fig. 5.9, the 2D spectrum is obtained. To obtain the different RIRs, one simply needs in both setups to divide the obtained spectrum by the spectrum of one period of the excitation signal and take the inverse Fourier transform of the dataset to obtain the RIRs at the different spatial locations. As the obtained 2D spectrum is almost bandlimited along the angular frequency dimension as was described in Chapter 4, interpolation of the RIRs at any angle can be achieved.

A numerical example is given.

Example 5.4.1. Using (5.33), to record RIRs of 100 ms on a circular array of radius of 1m up to a frequency of 20 kHz, the rotation needs to be achieved at a maximal speed of $v = 4.9$ deg/s. The duration of one full rotation is of approximatively 74 s. □

Note that, in order to keep the periodic character of the excitation signal, one will need to adapt the speed such that the time to make a full rotation corresponds to a multiple of the period of the excitation signal.

As was shown in Section 5.4.1, due to the Doppler effect, the recorded sound can contain higher frequencies than the original emitted sound. Therefore, to avoid aliasing due to the creation of these new frequencies, one has to choose a temporal sampling frequency slightly higher than twice the largest emitted frequency component.

5.4.4 Individualized HRTF measurement

The presented technique is very suitable for the recording of very short impulse responses as is the case for HRTFs. A typical length of interest for HRTF

measurement is of the order of 10 ms. Also, as was discussed in Chapter 4, the support of the 2D spectrum has most of its energy in the region satisfying

$$|l_\theta| \leq |\omega| \frac{d}{2c}, \tag{5.40}$$

with d the diameter of the head (typically of the order of 18 cm). With this modification, (5.33) becomes

$$v_{\max} = \frac{2\pi c}{d(\omega_1 T - \pi)}. \tag{5.41}$$

The maximal speed to reconstruct all HRTFs up to 20 kHz is therefore of 542 deg/s. In only .66 s, the measurement of HRTFs for all angles in the horizontal plane can be achieved. This interesting result finds application in the fast measurement of HRTFs. This technique could be further developed to allow every person to measure his own set of HRTFs and be able to use them when listening during headphone playback. Obviously, the content of the audio playback has to be adapted to support the use of HRTFs. This is precisely the goal of the new MPEG standard that will allow the use of individualized HRTFs for optimal audio rendering using headphones [26].

5.5 Experiments

Experiments have been carried out in a sound insulated room. A loudspeaker (Genelec $1029A$) was used to generate an excitation signal, recorded by a microphone Beyerdynamic $MC-740$. In this section, only the case of the moving microphone has been considered. Rotation of the microphone was performed using a Pan/Tilt Unit $PTU-D46-70$ [4]. This device allows the rotation of the microphone with a precision of .03 deg and with a maximal rotation speed of 60 deg/s. The room, where the measurements were performed, was so that the RIRs have very small energy after 250 ms. The excitation signal was made of the sum of sinusoids spaced with 4 Hz. Each sinusoid component was given a random phase. The frequencies covered by the excitation signal ranged from 20 Hz to 22 kHz. The period of the excitation signal was of length 250 ms. To reconstruct the signal up to 22 kHz on a circular array of radius 60 cm, the maximal rotation speed is of 3 deg/s, as follows from (5.33). We chose to apply a rotation of 1.8 deg/s. Therefore, the full rotation was achieved in 200 s. The Fourier transform of the recorded signal (in blue full lines) and of the emitted signal (in black dotted lines) are shown for low frequencies in Fig. 5.10(a) and

[4]The rotation of the loudspeaker would not have been possible due to the large weight of the loudspeaker.

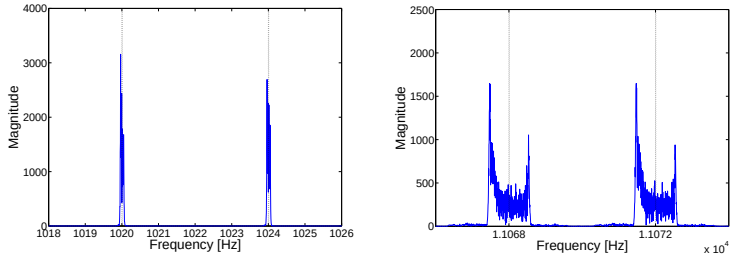


Figure 5.10: Spectrum of the moving microphone signal (a) at low frequencies (between 1018 and 1026 Hz); (b) at high frequencies (between 11066 and 11074 Hz).

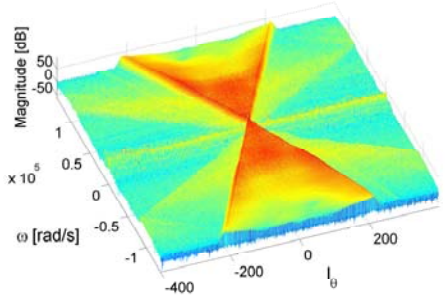


Figure 5.11: 2D-FT of the recorded data obtained after undoing the effect of the projection due to the recording by the moving microphone.

for high frequencies in Fig. 5.10(b). As discussed in Section 5.4.1, it can be observed that for the low frequencies, the projection of the triangular spectrum is very narrow while it becomes wider for larger frequencies. From the spectrum of the recorded signal, the original 2D spectrum needs to be reconstructed. This is achieved by undoing the effect of the projection appearing because of the movement of the microphone. The obtained reconstructed 2D spectrum is shown in Fig. 5.11. This spectrum is then further divided by the spectrum of one period of the excitation signal. Taking the 2D inverse Fourier transform of the result leads to the RIRs at the different angular positions. A typical RIR obtained by the algorithm is shown in Fig. 5.12. The RIR is not exactly zero at its beginning because of time aliasing. Aliasing happens since the RIR is not exactly zero after 250 ms but still contains some energy.

Estimating the correctness of the reconstructed RIRs is not a straightforward task. This is mainly due to different limitations related to the setup of the experiment. First of all, the hardware used did not allow a perfect synchronization between the moving motor and the emitted sound. Another limitation

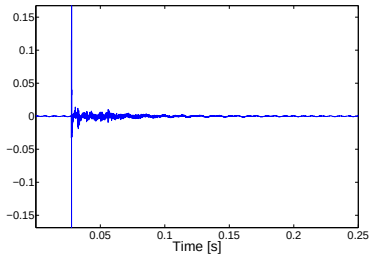


Figure 5.12: Example of a reconstructed RIR using the presented algorithm.

of the material was due to the noise emitted by the motor when the Pan/Tilt unit moves. Adaptively removing the noise or putting around the motor some sound absorbing material could be considered to attenuate this noise source. Nevertheless, the reconstructed RIRs were compared with statically measured RIRs and the mean squared error (MSE) between them was of the order of -10 to -15 dB. Note that the measurement conditions were not exactly similar for the measurement of the two sets of RIRs (one set obtained with the moving microphone and the other set using a static microphone). The temperature variation as well as the intrusion of a person in the room can lead to variations in the speed of the sound as was shown in [45, 7]. It is shown in [45] that a variation of 0.1° can create a misalignment between RIRs of more than 25 dB. These different aspects make the comparison in MSE sense difficult. Nevertheless, it can be observed that the relative time difference and attenuation between the reconstructed RIRs is very coherent and similar to the original RIRs.

A new technique to record RIRs in a very fast and easy manner has been presented and measurements have shown that the goals presented can be achieved. Some further work still needs to be addressed to solve the minor issues discussed above.

5.6 Conclusion

In this chapter, a technique was presented to record large sets of room impulse responses using a microphone or a loudspeaker moving along a trajectory. The technique reconstructs the room impulse responses at any position along a linear or a circular array. The speed of movement of the microphone (or loudspeaker) has been shown to be the key factor for the reconstruction. The theoretical aspects of the presented algorithm have been compared with experimental results.

Chapter 6

Stochastic Modeling of Spatio-Temporal Channel Impulse Responses Based on the Wave Equation

6.1 Introduction

Multipath channels between moving senders and receivers are of interest in many signal processing and communications scenarios, from acoustic echo cancellation to wireless mobile communications. Two physical constraints characterize these channels: (i) propagation is governed by the wave equation, (ii) movements of sender and/or receiver are smooth and slow as compared to propagation. In this work, we show that the resulting time-varying impulse response has a butterfly shaped spectrum, that is, its spatial bandwidth increases linearly with temporal frequency.

The method to show this uses smooth trajectories obtained by Ornstein-Uhlenbeck processes in space, in order to generate a suitable stochastic process corresponding to time-varying impulse responses. The distinguishing features of the power spectrum is its essential butterfly shape and the widening of the shape as a function of the smoothness of movement. Experimental results verify the established shape of the power spectra. Both the theoretical and experimental results are relevant for modeling of wideband channels (e.g. acoustic or ultra-wideband), since they predict the "variation bandwidth" as a function of temporal frequency. This in turn is important for channel adaptation,

equalization and power allocation.

6.1.1 Description and problem statement

Ever since listening to a passing ambulance as kids, we know that channels between a moving source and a receiver are more interesting than static scenarios. However, such time-varying channels are more challenging to characterize, analyze and equalize. The time-invariant channel leads to a convolution between its input and output and the Fourier domain is a natural setting to answer many relevant questions. The linear time-varying channel still obeys a superposition principle, but the impulse response at each instant is different. This is where physics comes into play. The impulse response corresponds to transmission of a pulse over a physical medium, governed by a differential equation, namely the wave equation. And the movement of the source is not arbitrary, but governed by Newtonian mechanics, which forces, among other things, smooth movements. This implies that two subsequent impulse responses which are close in time cannot differ that much. Consider a variation in time $\Delta\tau$ between two subsequent measured channel impulse responses (CIRs). During this variation of time, the maximal spacing Δd that has been traveled by the wave is

$$\Delta d = c\Delta\tau, \tag{6.1}$$

with c the speed of the wave propagation. At low temporal frequencies, the waves propagate with very large wavelength λ. Therefore, for $\Delta d << \lambda$, consecutive impulse response might be almost unchanged, while for high temporal frequencies, they might be completely different.

The challenge is to transform the above intuition into a mathematically tractable model. To do this, we use two ingredients as described in the block diagram of Fig. 6.1.

First, we use the Green's function which corresponds to the solution of the wave equation to an excitation pulse as described in Chapter 2. Previous work has been presented on the study of the solution of the wave equation along different geometries in the acoustic case [7,6] as well as in the electromagnetic case [37,36]. This is also the topic of the previous chapters, where, for example, the wave field has been studied along a line by considering the two-dimensional (2D) *plenacoustic function*, $p(x,t)$. This function describes the CIRs from a source to each position x along the line. Each CIRs is a function of the time t. Analysis of $p(x,t)$ was performed by considering its two-dimensional Fourier transform (2D-FT). Introducing the spatial frequency ϕ and the temporal frequency ω, it was shown that the support of the spatio-temporal CIR contained

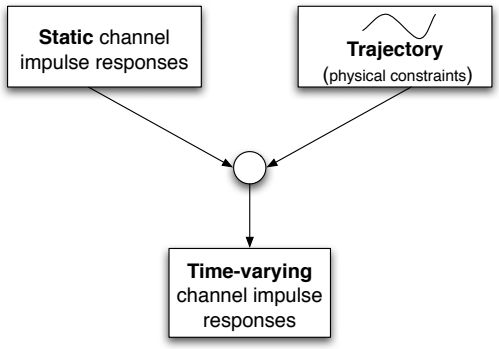


Figure 6.1: Block diagram of the general setup described in this work. On one hand, the static CIRs are studied in a region of space. On the other hand, a physically based trajectory for the movement of the receiver (or source) is presented. From these two ingredients, the time-varying CIRs corresponding to the considered system can be characterized.

most of its energy in a butterfly shaped region satisfying:

$$|\phi| \leq \frac{|\omega|}{c}. \tag{6.2}$$

For low temporal frequencies, the spatial support of the butterfly shaped spectrum is narrow since large wavelengths do not produce fast variations along the spatial dimension. For higher temporal frequencies, the wavelengths are smaller which leads to faster spatial variations of the field and larger spatial frequency support. The intuition gathered by this study is useful for the characterization and modeling of time-varying channels discussed in the present chapter[1].

Second, we introduce a smooth motion model for the moving receiver (or source or both). This is based on an Ornstein-Uhlenbeck process which can be described by a simple autoregressive relation. Simplicity and versatility are the two crucial characteristics of our autoregressive stochastic model: by tuning the order and the poles of the autoregression we can easily control the smoothness of the movement and hence describe a large family of moving receiver/source scenarios, from walking listeners in an audio room to automobile wireless stations, without forgetting the passing ambulance of our childhood. This is achieved by developing a generalization of the continuous-time Ornstein-Uhlenbeck process to any order as described in Section 6.2.

Putting these two ingredients leads to the analysis of the CIRs studied along a stochastic trajectory. Typical setups of interest include the study of

[1]More details on the deterministic spatio-temporal CIRs can be found in Chapters 3 and 4.

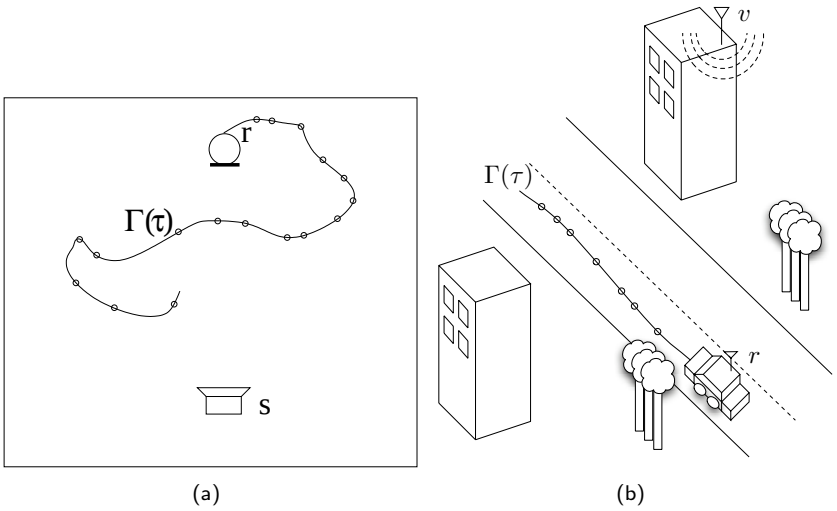


Figure 6.2: Setup of the problem. (a) For an acoustic case. (b) For an electromagnetic case.

time-varying CIRs in the acoustic and electromagnetic case with application to hands free telephony and mobile communications as shown in Fig. 6.2. In this chapter, we consider a receiver r following a smooth trajectory $\Gamma(\tau)$ given by the autoregressive model and a source s at a constant position as shown in Fig. 6.2(a) and (b). $\Gamma(\tau)$ is the random process describing the trajectory and τ is its independent variable. The variable τ represents the different instants at which we consider the curve as marked on Fig. 6.2(a) and (b). The variable τ is expressed in seconds and $\Gamma(\tau)$ in meters[2]. In this work, we are interested in modeling the channel between the source and the receiver. We therefore consider that at each position $\Gamma(\tau)$, the receiver measures the Dirac pulse emitted by the fixed source. The measured impulse response is a function of the time t representing the duration of the recorded impulse response. Considering all the impulse responses measured from each position along the random trajectory, the recorded dataset is a function of both τ and t. This dataset can be considered as a *stochastic plenacoustic function*, $P(\tau, t)$[3]. As it is not convenient to express two different times τ and t, we decide in the sequel to denote the coordinate τ as the spatial coordinate. It is a time function but is directly related to the position that is taken by the mobile at instant τ, namely $\Gamma(\tau)$. Therefore, $P(\tau, t)$ is denoted as the spatio-temporal CIR. A typical example

[2] As the velocity of the moving receiver is not constant, uniform samples of τ do not lead to uniform samples in space.

[3] An uppercase letter was chosen to put in evidence the stochastic character of the function.

of such a function is given in Fig. 6.3 where the first three peaks of the CIRs are shown. In order to characterize the channel from the source to the re-

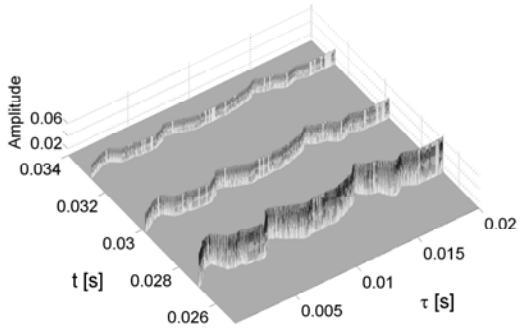


Figure 6.3: Example of a time-varying channel in function of τ describing a smoothed random trajectory and t the duration of each CIR.

ceiver, this chapter studies the autocorrelation and the power spectral density (PSD) of the CIRs. It will be shown, similarly to the deterministic case, that the PSD of the spatio-temporal CIR has a butterfly spectrum. To understand the intuition behind this specific shape, note that signals gathered along the trajectory present a high correlation at low temporal frequencies. At higher frequencies, a lower correlation is observed which explains the widening of the spectrum. The butterfly shape of the spectrum is dependent on the smoothness of the trajectory. For larger variance of the movement, the butterfly shape gets wider and "opens up". This can be shown exactly in some simple cases; for more complicated ones, approximation rules [29, 113] (well known from frequency modulation) and numerical simulations provide adequate answers. In this chapter, it is also shown that a stochastic perturbation can be added to a deterministic spatio-temporal CIR. In that case the overall PSD also has a butterfly spectrum originating from both the stochastic and the deterministic PSDs.

Knowing the spatio-temporal spectrum, the PSD of the one-dimensional signal measured at the receiver is considered and by introducing a speed of movement for the receiver, the Doppler effect is put in evidence.

With this analysis, we obtain a versatile model for time-varying channels governed by the wave equation and smooth motion. The presented theory is confirmed by experimental measurements. For this purpose, room impulse responses have been measured at hundreds of positions[4] along a trajectory and the results are comparable to the theoretical ones.

[4]All the performed measurements are available online [2].

Note that related work in the field on channel impulse responses modeling has already been previously described in Section 2.3.2.

6.1.2 Possible applications

Why is this an important problem, beyond listening to passing car equipped with sirens? First, it is a fundamental question to characterize time-varying channels in many signal processing and communications problems. This ranges from hands free telephony, where the channel variations are related to physical changes (and thermal fluctuations, among other noises) to mobile communications, where cars with varying speeds pass a base station.

Second, a good characterization of channels has algorithm implications, since many adaptive algorithms (e.g. echo cancellation in the acoustic case or multipath cancellation in the EM case) need to be tuned to channel characteristics [58]. Knowing characteristics of the spectrum can become very precious when choosing parameters of adaptive algorithms. The knowledge of the support of the spectrum can lead to a quantitative analysis of the rate of adaptation. At low temporal frequencies, the rate of change can be kept slow while it has to increase at higher temporal frequencies. This topic is matter of current research.

Third, it has legal implications. New services are submitted to spectral regulations like those of the US FCC established for ultra-wide bandwidth transmissions [3]. Characterizing the spectral variation induced by the movement of the receiver/source is of foremost importance: it allows us to take spectral regulations into account even in a dynamic scenario.

Last, physically based models can be used as a starting point for stochastic modeling of CIRs. The currently available stochastic models [94, 93, 52] could be parameterized to take physical constraints into account such as smoothness of trajectories in order to generate more realistic stochastic models for CIRs.

6.1.3 Contributions

The present work aims at developing a simple model to explain and quantify the butterfly shape of the 2D-FT of a sequence of CIRs governed by the wave equation and physically based mobility models. In particular, the contributions of this work are to

- develop a model for smooth random walks applicable for a large range of applications,
- study time-varying channel impulse responses governed by the wave equations,

- calculate and estimate the autocorrelation and power spectral density associated to these channels,
- present experimental measurements corroborating the theoretical results.

6.1.4 Outline

We develop these contributions along the following plan. As was discussed in the introduction, two main ingredients are necessary for this study. The first one describes the deterministic spatio-temporal channel impulse response and has already been discussed in details in Chapter 3. The second ingredient necessary to our study is the model of the trajectory followed by the mobile. It is described in Section 6.2. With the knowledge acquired by the two first sections, the spatio-temporal CIR is studied along a smoothed random trajectory in Section 6.3.1. The autocorrelation and the PSD associated to the CIR is described in Section 6.3.2. The influence of the poles and the order of the AR process on the smoothness of the trajectory and on the shape of the power spectral density are discussed in Sections 6.3.3. Further, the PSD of the one-dimensional signal gathered by a moving receiver is derived in Section 6.3.4[5]. Section 6.4 presents generalized results on the study of the stochastic spatio-temporal CIR in the case of multipath channels. With this study one can characterize the PSD of a channel between a fixed source and a receiver along a random trajectory. In different applications, one would like to add a deterministic trend to the movement of the receiver as discussed in Section 6.5. The presented theory is compared to simulations and experimental results in Section 6.6. The conclusions are drawn in Section 6.7.

6.2 Model for the Moving Receiver

To describe the time-varying channel that is observed by a moving receiver, one has to consider a model for the movement. One option is to choose a random walk or a random way point model [79]. Nevertheless in the present case, a totally random trajectory would not be realistic since it allows large jumps between consecutive positions. The typical movement of a receiver presents some physical constraints. It satisfies Newton's equation and therefore a certain inertia is present in the movement due to the double integration performed on the given acceleration. In this work, the control of the smoothness and the speed of variation of the movement are shown to be the key factor in the characterization of the time-varying channel. A continuous autoregressive (AR)

[5]In this chapter we only consider the case of a moving receiver and fixed source. Similar results are obtained in the dual case where the source is moving with a fixed receiver.

model is chosen to design continuous random trajectories. It allows us to easily control and modify the smoothness of the trajectory. Modifying the order of the AR model and displacing the positions of its poles allows the design of a broad range of realistic random walks by keeping the possibility to include physical constraints. In this work we present most of the results considering the case of an AR-2 model. That is, we consider the acceleration of the receiver to be a white Gaussian process. By integrating twice this process, the position of the moving receiver is smoothed as will be shown in the sequel.

In this section, we introduce the AR model to be used in the rest of the chapter. Typically, a continuous first-order AR model is denoted as an *Ornstein-Uhlenbeck* process [31]. In this work we will generalize this process to any order. For this purpose we introduce a linear model described by a state-space system in controller canonical form [78]:

$$\begin{cases} \dot{\mathbf{U}}(\tau) = F\mathbf{U}(\tau) + \mathbf{b}W(\tau) \\ Y(\tau) = \mathbf{c}^T\mathbf{U}(\tau), \end{cases} \tag{6.3}$$

with

$$F = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \\ 0 & 0 & 0 & \dots & 1 \\ -a_0 & -a_1 & -a_2 & \dots & -a_{n-1} \end{bmatrix}, \ \mathbf{b} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}, \ \mathbf{c} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}.$$

$\mathbf{U}(\tau)$ corresponds to the state variables of the system and $W(\tau)$ corresponds to the input signal. The output of the system is $Y(\tau)$. In our study, we consider the input $W(\tau)$ of the system to be a white Gaussian real stationary process. Further, the output of the system is the state variable corresponding to the position. With this choice, it can be shown that the output of the system $Y(\tau)$ follows precisely a continuous real AR model. The order of the model corresponds to the size of the matrix F in (6.3). In Section 6.2.1, the study of the autocorrelation function is presented for a receiver moving following an AR-2 law. Section 6.2.2 generalizes the results for any AR order. As will be discussed further, more smoothness can be achieved in the output of the system when increasing the order of the autoregressive model.

6.2.1 AR-2 process

In the case of an AR-2 process, the two state variables are position $X(\tau)$ and velocity $V(\tau)$. The input $W(\tau)$ is a white Gaussian stationary process. To

write the system (6.3) in two dimensions, we set

$$\mathbf{U}(\tau) = \begin{bmatrix} X(\tau) \\ V(\tau) \end{bmatrix}, \; F = \begin{bmatrix} 0 & 1 \\ -a_0 & -a_1 \end{bmatrix}, \; \mathbf{b} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \; \mathbf{c} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \tag{6.4}$$

Using (6.4), (6.3) can be rewritten as

$$\begin{cases} \ddot{X}(\tau) = -a_0 X(\tau) - a_1 \dot{X}(\tau) + W(\tau) \\ Y(\tau) = X(\tau). \end{cases} \tag{6.5}$$

Since $Y(\tau)$ is a stationary process, its correlation can be written as

$$R_Y(\Delta\tau) = \mathbb{E}[Y(\tau + \Delta\tau)Y(\tau)]. \tag{6.6}$$

If the system is stable, only the steady state term is present at the output, i.e. [78]

$$Y(\tau) = \int_{-\infty}^{\tau} \mathbf{c}^T e^{F(\tau - l)} \mathbf{b} W(l) \mathrm{d}l. \tag{6.7}$$

Therefore, $R_Y(\Delta\tau)$ for $\tau \geq 0$ can be written as follows:

$$\begin{aligned} \mathbb{E}\left[\int_{-\infty}^{\tau+\Delta\tau} \int_{-\infty}^{\tau} \mathbf{c}^T e^{F(\tau+\Delta\tau-l_1)} \mathbf{b} W(l_1) W(l_2) \mathbf{b}^T e^{F^T(\tau-l_2)} \mathbf{c} \mathrm{d}l_1 \mathrm{d}l_2 \right] \\ = \sigma^2 \int_{-\infty}^{s} \mathbf{c}^T e^{F(\tau+\Delta\tau-l)} \mathbf{b}\mathbf{b}^T e^{F^T(\tau-l)} \mathbf{c} \mathrm{d}l, \end{aligned} \tag{6.8}$$

using that $\mathbb{E}[W(l_1)W(l_2)] = \sigma^2 \delta(l_1 - l_2)$. We assume that the matrix F can be diagonalized[6]. Consequently, F can be decomposed using its eigenvalues λ_1 and λ_2 and eigenvector matrix V. Therefore,

$$e^{F\tau} = V \begin{bmatrix} e^{\lambda_1 \tau} & 0 \\ 0 & e^{\lambda_2 \tau} \end{bmatrix} V^{-1} = \begin{bmatrix} L(\tau) & M(\tau) \\ N(\tau) & O(\tau) \end{bmatrix}. \tag{6.9}$$

Note that in the sequel the eigenvalues of the matrix F are also denoted as poles of the AR process. The poles need to be negative to ensure the stability of the system. Using (6.9), it can be shown that (6.8) can be rewritten as

$$\begin{aligned} R_Y(\Delta\tau) &= \sigma^2 \int_{-\infty}^{\tau} M(\tau + \Delta\tau - l) M(\tau - l) \mathrm{d}l \\ &= \sigma^2 \int_{-\infty}^{0} M(\Delta\tau - l) M(-l) \mathrm{d}l, \end{aligned} \tag{6.10}$$

[6] If this is not the case, similar results can be derived by using the Jordan decomposition [66].

with $M(\Delta\tau - l) = C_1 e^{\lambda_1(\Delta\tau - l)} + C_2 e^{\lambda_2(\Delta\tau - l)}$ where C_1 and C_2 are obtained as $C_1 = V[1,1]V^{-1}[1,2]$ and $C_2 = V[1,2]V^{-1}[2,2]$.

By integrating (6.10), and taking into account that $R_Y(\Delta\tau)$ is an even function, we obtain

$$R_Y(\Delta\tau) = -\sigma^2 \left(e^{\lambda_1|\Delta\tau|} \left[\frac{C_1^2}{2\lambda_1} + \frac{C_1 C_2}{\lambda_1 + \lambda_2} \right] + e^{\lambda_2|\Delta\tau|} \left[\frac{C_2^2}{2\lambda_2} + \frac{C_1 C_2}{\lambda_1 + \lambda_2} \right] \right). \tag{6.11}$$

Introducing now the random vector $\mathbf{Y} = [Y(0)Y(\Delta\tau)]^T$, the correlation matrix $R_{\mathbf{Y}}$ is

$$R_{\mathbf{Y}} = \begin{bmatrix} \mathbb{E}[Y(\Delta\tau)Y(\Delta\tau)] & \mathbb{E}[Y(\Delta\tau)Y(0)] \\ \mathbb{E}[Y(0)Y(\Delta\tau)] & \mathbb{E}[Y(\Delta\tau)Y(\Delta\tau)] \end{bmatrix}, \tag{6.12}$$

while the probability density function of $\mathbf{Y}$ is given by:

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{2\pi\sqrt{|R_{\mathbf{Y}}|}} e^{\frac{-R_Y(0)}{2|R_{\mathbf{Y}}|}(y_0^2 + y_{\Delta\tau}^2 - \frac{2R_Y(\Delta\tau)}{R_Y(0)} y_{\Delta\tau} y_0)}. \tag{6.13}$$

Note that in this expression, $y_{\Delta\tau}$ corresponds to a realization of the random variable $Y(\Delta\tau)$ and y_0 to a realization of the random variable $Y(0)$.

6.2.2 Generalization to AR-N process

In the case of an AR-N process the previous formulae can be generalized. The matrix F is still considered to be diagonalizable and can therefore be decomposed in its N eigenvalues $\lambda_1 \ldots \lambda_N$ and N eigenvectors represented by the eigenvector matrix V of dimension N. The autocorrelation of the process $Y(\tau)$ is then

$$R_Y(\Delta\tau) = -\sigma^2 \left[\sum_{i=1}^{N} \left(\sum_{j=1}^{N} \frac{C_i C_j}{\lambda_i + \lambda_j} \right) e^{\lambda_i|\Delta\tau|} \right], \tag{6.14}$$

where $C_i = V[1,i]V^{-1}[i,N]$.

In order to compare different AR-N processes, one has to normalize the noise input signals. The normalization is chosen to obtain a unitary variance for the output of the system, $Y(\tau)$. For this, one simply needs to impose that

$$\sigma^2 = -\frac{1}{\sum_{i=1}^{N} \left(\sum_{j=1}^{N} \frac{C_i C_j}{\lambda_i + \lambda_j} \right)}. \tag{6.15}$$

6.3 Spatio-Temporal Channel Impulse Response

In this section, the spatio-temporal CIR is studied. The filtering process of an input signal is considered and the output process is studied both along the temporal and the spatial dimension.

For simplicity of the calculations, a few assumptions are added to the model. First, the movement of the receiver follows an AR model along one line as shown in Fig. 6.4. We consider a mean distance between source and receiver of d.

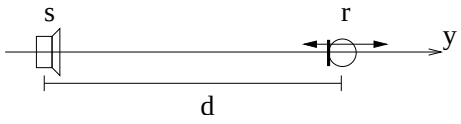


Figure 6.4: Setup of the problem: a source is emitting and a receiver moving along a line following an AR process is measuring the wave field.

For each possible position of the receiver we consider the CIR between that position and the source. The spatio-temporal CIR obeying the wave equation is given by (3.5). In our case, it rewrites as

$$P(\tau,t) = \frac{\delta(t - \frac{|Y(\tau)+d|}{c})}{4\pi|Y(\tau)+d|}. \tag{6.16}$$

Another assumption done to simplify the calculations, is to consider the receiver far enough from the source to neglect the attenuation depending on the distance, i.e. the denominator in (6.16) is considered to be constant. This can be considered as the plane wave or far-field assumption as described in Section 3.3.1. Also, we consider d to be at least a few times larger than the standard deviation of the process Y. With these assumptions, (6.16) rewrites as

$$P(\tau,t) = \delta\left(t - \frac{d+Y(\tau)}{c}\right). \tag{6.17}$$

The next section will present the filtering of an input process with the filter $P(\tau,t)$.

6.3.1 Filtering of an input process with the spatio-temporal CIR

Consider an input process $I(t)$ corresponding to the sound emitted by a source. This input is a stationary process of mean m_I and of variance σ_I^2. It is then filtered by $P(\tau,t)$ to obtain an output $O(\tau,t)$. Each realization of the process $P(\tau,t)$ is a linear and time invariant filter with respect to time t. Therefore,

we can consider the filtering of an input process $I(t)$ with each of these filters. Thanks the the stability of each realization of the process $P(\tau, t)$ (simple delay filters), the output process $O(\tau, t)$ is well defined. We therefore have that

$$O(\tau,t) = I(t) * P(\tau,t) = I(t) * \delta\left(t - \frac{d+Y(\tau)}{c}\right) = I\left(t - \frac{d+Y(\tau)}{c}\right). \quad (6.18)$$

A scheme of the situation is shown in Fig. 6.5: The mean of the output can be

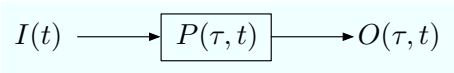


Figure 6.5: Filtering of a random process by the spatio-temporal CIR.

calculated as follows:

$$\begin{aligned} \mathbb{E}\left[I\left(t - \frac{d+Y(\tau)}{c}\right)\right] &= \mathbb{E}\left[\mathbb{E}\left[I\left(t - \frac{d+Y(\tau)}{c}\right)\middle| Y(\tau)\right]\right] \\ &= \mathbb{E}[m_I] = m_I, \end{aligned} \quad (6.19)$$

where the last equality in (6.19) follows from the stationarity of $I(t)$. The covariance of the process $O(\tau, t)$ is calculated as follows:

$$\begin{aligned} R_O(\Delta\tau, \Delta t) &= \mathbb{E}\left[I\left(t + \Delta t - \frac{d+Y(\tau+\Delta\tau)}{c}\right) I\left(t - \frac{d+Y(\tau)}{c}\right)\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[I\left(t + \Delta t - \frac{d+Y(\tau+\Delta\tau)}{c}\right)\right.\right. \\ &\qquad \left.\left. I\left(t - \frac{d+Y(\tau)}{c}\right)\middle| Y(\tau+\Delta\tau), Y(\tau)\right]\right] \\ &= \mathbb{E}\left[R_I\left(\Delta t - \frac{Y(\tau+\Delta\tau) - Y(\tau)}{c}\right)\right] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R_I\left(\Delta t - \frac{y_{\Delta\tau} - y_0}{c}\right) f_{\mathbf{Y}}(\mathbf{y}) \mathrm{d}y_{\Delta\tau}\mathrm{d}y_0. \end{aligned} \quad (6.20)$$

In Appendix E, it is shown that (6.20) can be rewritten as

$$R_O(\Delta\tau, \Delta t) = R_I(\Delta t) * \frac{c e^{-\frac{c^2(\Delta t)^2}{4(R_Y(0) - R_Y(\Delta\tau))}}}{2\sqrt{\pi\left(R_Y(0) - R_Y(\Delta\tau)\right)}}, \quad (6.21)$$

with the convolution along time t axis.

6.3.2 Power spectral density of the spatio-temporal CIR

From the analysis of the mean and the covariance of $O(\tau, t)$, it is observed that the process is wide sense stationary. Therefore, a power spectral density can be

defined. For this purpose, the Fourier transform of $R_O(\Delta\tau, \Delta t)$ with respect to time, denoted as $\tilde{R}_O(\Delta\tau, \omega)$, is first considered

$$\begin{aligned}\tilde{R}_O(\Delta\tau, \omega) &= \int_{-\infty}^{\infty} R_O(\Delta\tau, \Delta t) e^{-j\omega\Delta t} \mathrm{d}\Delta t \\ &= S_I(\omega) e^{-\left(\frac{\omega}{c}\right)^2 (R_Y(0) - R_Y(\Delta\tau))}. \end{aligned} \tag{6.22}$$

For an AR-2 process, we have

$$\tilde{R}_O(\Delta\tau, \omega) = S_I(\omega) e^{-\left(\frac{\omega}{c}\right)^2 \left[G_1\left(1 - e^{\lambda_1 |\Delta\tau|}\right) + G_2\left(1 - e^{\lambda_2 |\Delta\tau|}\right)\right]}, \tag{6.23}$$

with $G_1 = -\sigma^2 \left(\frac{C_1^2}{2\lambda_1} + \frac{C_1 C_2}{\lambda_1 + \lambda_2}\right)$ and $G_2 = -\sigma^2 \left(\frac{C_2^2}{2\lambda_2} + \frac{C_1 C_2}{\lambda_1 + \lambda_2}\right)$.

In the generalized case of an AR-N process, it can be shown that

$$\tilde{R}_O(\Delta\tau, \omega) = S_I(\omega) e^{-\left(\frac{\omega}{c}\right)^2 \left[\sum_{i=1}^{N} G_i\left(1 - e^{\lambda_i |\Delta\tau|}\right)\right]}, \tag{6.24}$$

with $G_i = -\sigma^2 \sum_{j=1}^{N} \frac{C_i C_j}{\lambda_i + \lambda_j}$. For simplicity we consider in the sequel that the emitted signal is a white Gaussian noise of unitary variance. Therefore, $S_I(\omega) = 1$ in (6.23) and (6.24).

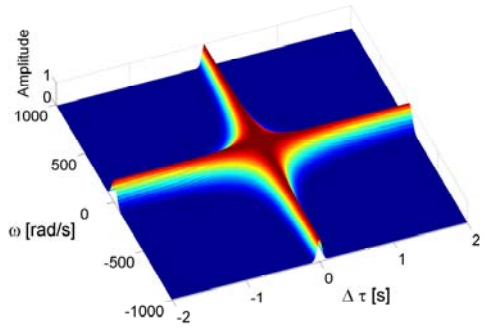
In Fig. 6.6(a), $\tilde{R}_O(\Delta\tau, \omega)$ is shown for an AR-2 process with poles (or eigenvalues) located at positions -9 and -10. Note that this AR system is stable since the poles of the system are negative as discussed previously in Section 6.2.1. It can be observed that for $\omega = 0$, the autocorrelation is unitary for any $\Delta\tau$ since there is a total correlation between any two receivers measuring a DC plane wave. For increasing temporal frequencies, it can be observed that the correlation becomes narrower, since only very closely located positions remain correlated due to the very small wavelength of the propagating wave.

The PSD $S_O(\gamma, \omega)$ is given by the Fourier transform of the correlation function, i.e.

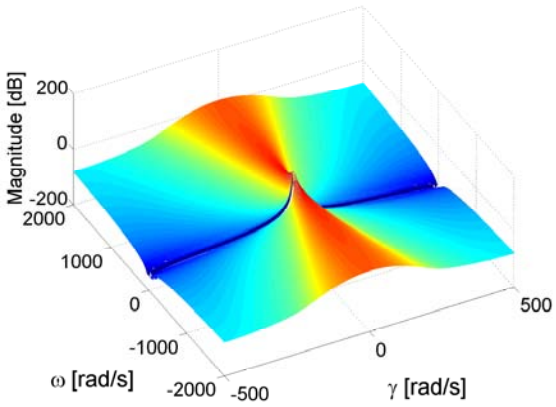
$$S_O(\gamma, \omega) = \int_{-\infty}^{\infty} e^{-j\gamma\Delta\tau} \tilde{R}_O(\Delta\tau, \omega) \mathrm{d}\Delta\tau. \tag{6.25}$$

While a closed form solution of this integral seems difficult to obtain, numerical evaluation of (6.25) is shown in Fig. 6.6(b) for a receiver moving following an AR-2 process with poles at -9 and -10. As can be seen from Fig. 6.6(b), the PSD of the spatio-temporal CIR has most of its energy contained in a butterfly region. The intuition behind this result is the following. Considering a source emitting a waveform of very low temporal frequency. If the receiver position varies slightly, the signals received at the receiver will be very correlated since the wavelength is very large. Thus, the Fourier transform of their correlation function will have a small support. When the temporal frequency of the emitted

signals increases, the received signals are less and less correlated even for small variation of the receiver position. Therefore the PSD of the spatio-temporal CIR gets a larger support for higher temporal frequencies.



(a)



(b)

Figure 6.6: Study of the spatio-temporal CIR in free field. (a) $\tilde{R}_O(\Delta\tau, \omega)$ as obtained from (6.24) with eigenvalues at -9 and-10. (b) PSD of this spatio-temporal CIR.

6.3.3 Trajectory smoothness influence on the butterfly spectrum characteristics

When observing trajectories of mobiles governed by the wave equation, a certain smoothness is present in the movement. It is the smoothness of the movement that we intend to reproduce by using AR models. To match the smoothness of the original movement, a few parameters can be modified in our model, namely the positions of the poles and the order of the AR process. In this

section, the different parameters to be chosen when designing the matching AR process are studied and their influence on the shape and widening of the 2D spectrum of the spatio-temporal CIRs is studied.

Poles locations of the AR system

By changing the poles (or eigenvalues) of the system, one can control the smoothness of the trajectory. This modifies the opening of the butterfly shaped spectrum of Fig. 6.6(b). The smaller the poles are (in absolute value), the more narrow is the spectrum, while for very large values of the poles (in absolute value), the frequency support of the PSD of the spatio-temporal CIR gets very wide. For an AR-1 process, the variation of the spatial bandwidth of the PSD has been studied in function of the position of the pole. The pole of the system has been displaced from $-.5$ to -32. For each of these positions for the pole, a cut of the PSD of the spatio-temporal CIR is shown at a temporal frequency of 1000π rad/s (500 Hz) in Fig. 6.7(a).

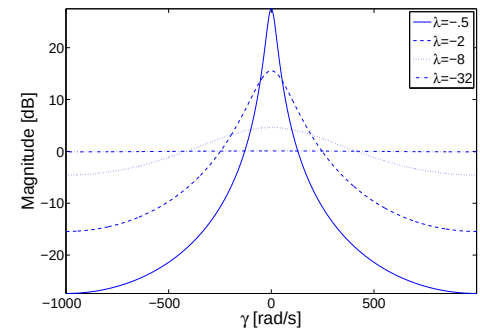
Order of the AR system

As known from system theory [66], for a fixed position of the poles, increasing the order of an AR model leads to smoother trajectories. Also, for smoother trajectories it is observed that the PSD of the spatio-temporal CIR decays faster. In Fig. 6.7(b), a cut of the PSD of a spatio-temporal CIR was considered at a temporal frequency of 500 Hz. The AR models of the three first orders are compared. The poles are all around -5, and as expected, the AR-3 model decays fastest.

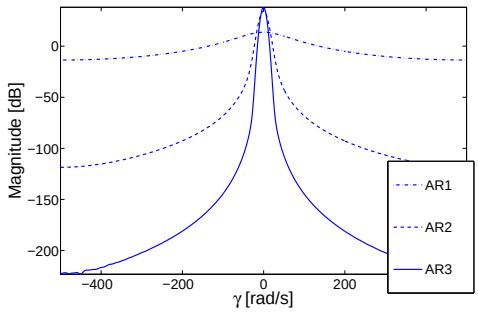
6.3.4 Power spectral density of the signal measured by a moving receiver

Until now, we have considered a continuous trajectory defining the movement of a mobile receiver using an autoregressive model as in Section 6.2. We have then studied the 2D spatio-temporal autocorrelation and PSD of the time-varying channel in Section 6.3.1 and 6.3.2. So far, we have not considered the one-dimensional signal gathered by the moving receiver, which we do now. Its PSD will be shown to be obtained from the 2D spatio-temporal PSD by means of a projection. Note that this kind of construction is very similar to the work presented in Chapter 5 for the signal gathered by a moving receiver along a deterministic circular trajectory. In Chapter 5, a similar construction is also presented for the case of a moving source.

From the model discussed in Section 6.2, the different positions along the



(a)



(b)

Figure 6.7: Influence of the different parameters of the AR processes on the PSD $S(\gamma, \omega)$. (a) For $\omega = 1000\pi$ rad/s, comparison of $S(\gamma, \omega)$ for different values of the poles of an AR-1 system ranging from $-.5$ to -32. (b) For $\omega = 1000\pi$ rad/s, comparison of $S(\gamma, \omega)$ for AR systems of different orders. The poles of the process are all located around -5.

trajectory are determined at each instant. From this trajectory, it is possible to compute the average speed of the mobile denoted as v_a. It will be dependent on the different parameters of the AR process. In this section, we introduce a scaling factor ν that allows the modification of the average speed of the mobile that now becomes νv_a.

As was previously discussed in Section 6.3.1, the filtering of an input process $I(t)$ with the spatio-temporal CIR $P(\tau, t)$ is obtained from (6.18) and can be written as follows:

$$O(\tau, t) = \int_{-\infty}^{\infty} I(u) P(\tau, t-u) \mathrm{d}u. \tag{6.26}$$

As was described in Chapter 5, the one-dimensional signal $M(t)$ gathered by

the moving receiver is given by

$$M(t) = \int_{-\infty}^{\infty} I(u)P(\nu t, t-u)\mathrm{d}u, \tag{6.27}$$

Note that

$$M(t) = O(\nu t, t). \tag{6.28}$$

The signal gathered by the receiver corresponds to a slice of the 2D function $O(\tau, t)$ along the line of equation $\tau = \nu t$. The parameter ν can therefore be seen as a time scaling.

The autocorrelation of the receiver signal is

$$R_M(\Delta t) = R_O(\nu\Delta t, \Delta t). \tag{6.29}$$

Again, it can be observed in (6.29), that the autocorrelation of the measured signal by a moving receiver corresponds to a slice of the output signal correlation $R_O(\tau, t)$ along a line of equation $\Delta\tau = \nu\Delta t$. In the frequency domain this corresponds to a projection as stated by the projection-slice theorem [25].

The PSD of the signal recorded by the moving receiver is

$$S_M(\xi) = \int_{-\infty}^{\infty} R_O(\nu\Delta t, \Delta t)e^{-j\xi\Delta t}\mathrm{d}\Delta t. \tag{6.30}$$

Also, remark that

$$R_O(\nu\Delta t, \Delta t) = \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} S_O(\gamma, \omega)e^{j(\omega\Delta t+\gamma\nu\Delta t)}\mathrm{d}\gamma\mathrm{d}\omega. \tag{6.31}$$

Using (6.31), Expression (6.30) can be rewritten as:

$$\begin{aligned} S_M(\xi) &= \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} S_O(\gamma, \omega)e^{-j\Delta t(\xi-\nu\gamma-\omega)}\mathrm{d}\Delta t\mathrm{d}\phi\mathrm{d}\omega \\ &= \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} S_O(\gamma, \omega)\delta(\xi-\nu\gamma-\omega)\mathrm{d}\gamma\mathrm{d}\omega \\ &= \frac{1}{4\pi^2}\int_{-\infty}^{\infty} S_O(\gamma, -\nu\gamma+\xi)\mathrm{d}\gamma. \end{aligned} \tag{6.32}$$

The PSD of the signal gathered by the moving receiver is given by (6.32). As mentioned earlier, the PSD can be obtained by projection of the original 2D PSD. For each frequency ξ, the value of the PSD of the recorded signal is obtained by a projection of the 2D spectrum $S_O(\gamma, \omega)$ following the direction $\vec{\boldsymbol{\nu}} = \frac{(1,-\nu)}{\sqrt{\nu^2+1}}$ on the ω-axis. This construction is presented in Fig. 6.8. The

support of the PSD of the moving receiver lies on the ω-axis as can be found out from (6.32). It is shown in bold dashed line. It can also be observed that the PSD of the moving receiver contains higher temporal frequencies than the original spectrum. This is due to the Doppler shift and is dependent on the speed of the receiver.

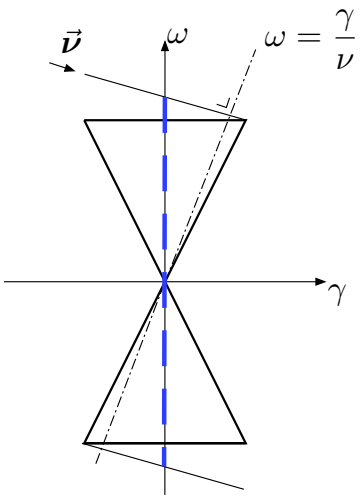


Figure 6.8: PSD of the moving receiver signal. The signal measured by the moving receiver is shown in dashed lines.

6.4 Multipath Channels

The autocorrelation function of the spatio-temporal CIR is now considered for multipath channels. The first setup discussed in Section 6.4.1 considers a single source not located along the line of the receiver movement. In Section 6.4.2, multipath reflection is studied in a rectangular room. As discussed in Section 2.3.2, a large literature exists on the modeling of multipath propagation channels. In this chapter we do not develop a new model for multipath propagation but we show that the CIRs have a specific PSD shape when a physically based realistic model is chosen for the spatial trajectory. For simplicity, a model for reverberation is considered based on the image source model as in [101]. From this model, it is shown that the butterfly shape of the studied PSD generalizes from the case of a single source in free field to a multipath model. Each multichannel path is considered in our analysis as a virtual source. Each of the considered virtual source is shown to produce a butterfly shaped PSD, therefore the total multipath channel also exhibits a similar butterfly shape.

6.4.1 Reflection from source

Consider a source s located at position $(0, 0)$ emitting the process $I(t)$ and a receiver along a trajectory described by an AR process around the position (d, l) as shown in Fig. 6.9(a). The output process $O(\tau, t)$ is given by the following

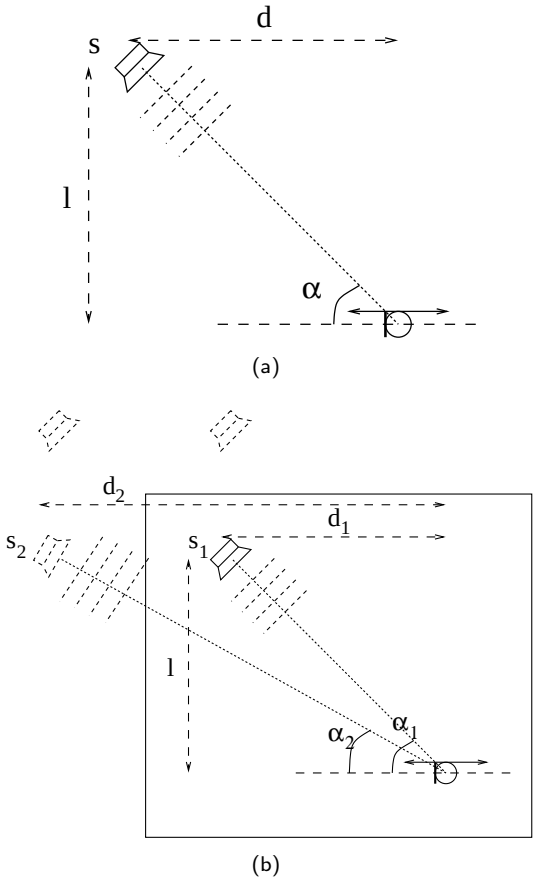


Figure 6.9: Different setups for the study of the spatio-temporal CIR. (a) Setup considering a source not aligned with the movement of the receiver. (b) Setup considering virtual sources for modeling the spatio-temporal CIRs.

expression:

$$O(\tau, t) = I\left(t - \frac{\sqrt{l^2 + (d + Y(\tau))^2}}{c}\right) \tag{6.33}$$

To simplify further calculations, linearization of (6.33) with respect to $Y(\tau)$ is performed, leading to

$$O(\tau, t) = I\left(t - \frac{\sqrt{l^2 + d^2} + \cos\alpha Y(\tau)}{c}\right), \tag{6.34}$$

with

$$\cos\alpha = \frac{d}{\sqrt{d^2 + l^2}}. \tag{6.35}$$

Remark that this linearization corresponds to the assumption that the source emits plane waves. Using results from Section 6.3.1, the autocorrelation of the output process can be written as:

$$\tilde{R}_0(\Delta\tau, \omega) = e^{j\frac{\omega}{c}\sqrt{d^2+l^2}} S_I(\omega) e^{-\left(\frac{\omega\cos\alpha}{c}\right)^2 (R_Y(0) - R_Y(\Delta\tau))}. \tag{6.36}$$

When calculating the Fourier transform of (6.36) with respect to $\Delta\tau$, we obtain

$$e^{j\frac{\omega}{c}\sqrt{d^2+l^2}} S_O(\gamma, \omega\cos\alpha), \tag{6.37}$$

with $S_O(\gamma, \omega)$ defined in (6.25). This shows that the largest support of the PSD is obtained when $\alpha = 0$. The reflections arriving with the smallest angles lead to the largest support of the PSD.

6.4.2 Multipath reflections in a rectangular room

Consider now the case of two sources. The sources s_1 and s_2 emit plane wave arriving on the receiver with angle α_1 and α_2, respectively, as shown in Fig. 6.9(b). The source s_2 corresponds to one virtual source mimicking the effect of the first reflection path.

The autocorrelation function associated to the output is first calculated for a case with two sources and then further generalized to N sources. We start with the two sources case and define $D_1 = \sqrt{d_1 + l^2}$ and $D_2 = \sqrt{d_2 + l^2}$. The new output process is defined as follows:

$$O(\tau, t) = I\left(t - \frac{D_1 + \cos\alpha_1 Y(\tau)}{c}\right) + I\left(t - \frac{D_2 + \cos\alpha_2 Y(\tau)}{c}\right). \tag{6.38}$$

The autocorrelation of the process, denoted as $R_O(\Delta\tau, \Delta t)$ is therefore given

by the following expression.

$$\mathbb{E}\left[\left[I\left(t+\Delta t-\frac{D_1+\cos\alpha_1 Y(\tau+\Delta\tau)}{c}\right)+I\left(t+\Delta t-\frac{D_2+\cos\alpha_2 Y(\tau+\Delta\tau)}{c}\right)\right]\right.$$
$$\left.\left[I\left(t-\frac{D_1+\cos\alpha_1 Y(\tau)}{c}\right)+I\left(t-\frac{D_2+\cos\alpha_2 Y(\tau)}{c}\right)\right]\right].$$

Rewriting this expression leads to

$$R_O(\Delta\tau,\Delta t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R_I\left(\Delta t-\cos\alpha_1\frac{y_{\Delta\tau}-y_0}{c}\right) f_{\mathbf{Y}}(\mathbf{y})\mathrm{d}y_{\Delta\tau}\mathrm{d}y_0 \quad (6.39)$$
$$+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R_I\left(\Delta t-\cos\alpha_2\frac{y_{\Delta\tau}-y_0}{c}\right) f_{\mathbf{Y}}(\mathbf{y})\mathrm{d}y_{\Delta\tau}\mathrm{d}y_0 \quad (6.40)$$
$$+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R_I\left(\Delta t-\frac{D_1-D_2+\cos\alpha_1 y_{\Delta\tau}-\cos\alpha_2 y_0}{c}\right) f_{\mathbf{Y}}(\mathbf{y})\mathrm{d}y_{\Delta\tau}\mathrm{d}y_0 \quad (6.41)$$
$$+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R_I\left(\Delta t-\frac{D_2-D_1+\cos\alpha_2 y_{\Delta\tau}-\cos\alpha_1 y_0}{c}\right) f_{\mathbf{Y}}(\mathbf{y})\mathrm{d}y_{\Delta\tau}\mathrm{d}y_0. \quad (6.42)$$

To calculate the Fourier transform with respect to Δt of this expression, we consider each term separately. The Fourier transform of (6.39) is

$$S_I(\omega)e^{-\left(\frac{\omega\cos\alpha_1}{c}\right)^2(R_Y(0)-R_Y(\Delta\tau))}. \quad (6.43)$$

Similarly to the technique presented in Section 6.3.1, the Fourier transform of (6.42) is

$$e^{j\omega(D_1-D_2)}S_I(\omega)e^{-\left(\frac{\omega}{c}\right)^2\left(R_Y(0)\frac{\cos^2\alpha_1+\cos^2\alpha_2}{2}-R_Y(\Delta\tau)\cos\alpha_1\cos\alpha_2\right)}. \quad (6.44)$$

We therefore have

$$\tilde{R}_O(\Delta\tau,\omega) = S_I(\omega)\left[e^{-\left(\frac{\omega\cos\alpha_1}{c}\right)^2(R_Y(0)-R_Y(\Delta\tau))}+e^{-\left(\frac{\omega\cos\alpha_2}{c}\right)^2(R_Y(0)-R_Y(\Delta\tau))}\right.$$
$$\left.+2\cos(\omega(D_1-D_2))e^{-\left(\frac{\omega}{c}\right)^2\left(R_Y(0)\frac{\cos^2\alpha_1+\cos^2\alpha_2}{2}-R_Y(\Delta\tau)\cos\alpha_1\cos\alpha_2\right)}\right]. \quad (6.45)$$

Generalizing (6.45) to N virtual sources leads to

$$\tilde{R}_O(\Delta\tau,\omega) = S_I(\omega)\left[\sum_{i=1}^{N} e^{-\left(\frac{\omega\cos\alpha_i}{c}\right)^2(R_Y(0)-R_Y(\Delta\tau))}+2\sum_{i=1}^{N}\sum_{j=i+1}^{N}\right.$$
$$\left.\cos(\omega(D_i-D_j))e^{-\left(\frac{\omega}{c}\right)^2\left(R_Y(0)\frac{\cos^2\alpha_i+\cos^2\alpha_j}{2}-R_Y(\Delta\tau)\cos\alpha_i\cos\alpha_j\right)}\right], \quad (6.46)$$

with $D_i=\sqrt{l_i^2+d_i^2}$. Note that d_i, l_i and α_i are the distances and angles from each virtual source to the receiver. The virtual sources can be obtained as

in [13]. Note that (6.46) can be rewritten for an AR-2 process as

$$\tilde{R}_O(\Delta\tau,\omega) = S_I(\omega)\left[\sum_{i=1}^{N} e^{-\left(\frac{\omega\cos\alpha_i}{c}\right)^2} e^{-\left(\frac{\omega}{c}\right)^2\left[G_1\left(1-e^{\lambda_1|\Delta\tau|}\right)+G_2\left(1-e^{\lambda_2|\Delta\tau|}\right)\right]} + 2\sum_{i=1}^{N}\sum_{j=i+1}^{N} \cos(\omega(D_i-D_j)) e^{-\left(\frac{\omega\sqrt{\cos\alpha_i\cos\alpha_j}}{c}\right)^2\left((G_1+G_2)\frac{\cos^2\alpha_i+\cos^2\alpha_j}{2\cos\alpha_i\cos\alpha_j}-G_1e^{\lambda_1|\Delta\tau|}-G_2e^{\lambda_2|\Delta\tau|}\right)}\right]. \tag{6.47}$$

To study the Fourier transform of (6.47) with respect to $\Delta\tau$, we first analyze the first sum. Each term in this sum leads to a PSD expressed by (6.37) as previously studied in Section 6.4.1. There it was shown that the largest frequency support is obtained for the smallest angle. The second sum leads to a similar analysis since the frequency ω becomes $\omega\sqrt{\cos\alpha_i\cos\alpha_j}$. Therefore, as a result of this analysis, it is observed that the smaller the angle, the larger is the spatial support. The maximal support is obtained, when for both sums, angles α_i and α_j are zero. The maximal spatial support of the PSD of (6.47) corresponds then to the PSD of the case where the source is on the line of movement of the receiver as studied in Section 6.3.2.

6.5 Mixing Deterministic and Stochastic Spatio-temporal CIR

Next to a purely stochastic spatio-temporal CIR, one may be interested in a mixture of a deterministic and a stochastic component. The most realistic scenario consists in a particular deterministic trajectory to which a stochastic perturbation is added. An example of such process corresponds to the sampling of the sound field along an array where a certain jitter exists due to the imprecision of the position of the measurement. This can be seen as the addition of a stochastic jitter to a deterministic measurement. Another example is the one presented in Fig. 6.2(b). A car is driving and receives electromagnetic signal from an antenna placed next to the freeway. Due to some stochastic variation in the traffic or in the landscape, the position of the car is jittered.

6.5.1 Deterministic channels

Consider a plane wave arriving on a line of receivers as discussed in Chapter 3. The field measured along the line is

$$p(\tau,t) = \delta\left(t-\tau\cos\alpha\right). \tag{6.48}$$

Considering the Fourier transform of (6.48) with respect to time leads to

$$\tilde{p}(\tau, \omega) = e^{-j\omega\tau\cos\alpha}. \tag{6.49}$$

Also, the 2D-FT of (6.48) is given by

$$\hat{p}(\gamma, \omega) = 2\pi\delta\left(\gamma + \omega\cos\alpha\right). \tag{6.50}$$

In the present section, the deterministic trend of the movement is considered to be a simple plane wave arriving on a receiver line. Considering more general expressions might lead to problems of the definition of the PSD of the spatio temporal CIR mixing deterministic and stochastic trends. For example, consider the following expression:

$$p(\tau, t) = \delta\left(t - f(\tau)\right). \tag{6.51}$$

As will be shown in Section 6.5.2, the function $f(\tau)$ needs to be a linear function of the variable τ to still be able to define a wide sense stationary autocorrelation function of the CIR. If the deterministic trend contains a non-linear function $f(\tau)$, linearization in little segments of the function needs to be applied. Each segment will be studied separately and will follow the presented theory. Nevertheless, the obtained autocorrelation function will have to be spatially filtered to match the size of the segments. In the frequency domain, this corresponds to a convolution along the spatial axis between the PSD and the Fourier transform of the used window. This can be considered as a short-time PSD by analogy to the short-time Fourier transform [111] used to analyze non-stationary signals.

6.5.2 Deterministic and stochastic spatio-temporal CIR

In addition to the deterministic field measured along a line, consider now a stochastic process added on the measurement to represent jitter. In this section, the output of the process, denoted as $\bar{O}(\tau, t)$, is studied when a stochastic input process is filtered by a deterministic and stochastic spatio-temporal CIR. Similarly to the technique presented in Section 6.3.1, we have that $\bar{O}(\tau, t)$ satisfies

$$\bar{O}(\tau, t) = I\left(t - \cos\alpha\left[\tau + \frac{Y(\tau)}{c}\right]\right). \tag{6.52}$$

The autocorrelation function of this process is

$$\begin{aligned} R_{\bar{O}}(\Delta\tau, \Delta t) &= \mathbb{E}\left[I\left(t + \Delta t - \cos\alpha \left[\tau + \Delta\tau + \frac{Y(\tau+\Delta\tau)}{c} \right] \right) \right. \\ &\qquad \left. I\left(t - \cos\alpha \left[\tau + \frac{Y(\tau)}{c} \right] \right) \right] \\ &= \mathbb{E}\left[R_I\left(\Delta t - \cos\alpha \left[\Delta\tau + \frac{Y(\tau+\Delta\tau) - Y(\tau)}{c} \right] \right) \right]. \end{aligned} \tag{6.53}$$

The Fourier transform of $R_{\bar{O}}(\Delta\tau, \Delta t)$ with respect to time Δt is

$$\begin{aligned} \tilde{R}_{\bar{O}}(\Delta\tau, \omega) &= e^{j\omega\cos\alpha\Delta\tau} S_I(\omega) e^{-\left(\frac{\omega\cos\alpha}{c}\right)^2 (R_Y(0) - R_Y(\Delta\tau))} \\ &= e^{j\omega\cos\alpha\Delta\tau} \tilde{R}_O(\Delta\tau, \omega), \end{aligned} \tag{6.54}$$

with $\tilde{R}_O(\Delta\tau, \omega)$ defined as in (6.36). Note that $\tilde{R}_{\bar{O}}(\Delta\tau, \omega)$ is only a function of $\Delta\tau$ and not directly of τ thanks to the choice of a linear function f in (6.51). To compute the Fourier transform of $R_{\bar{O}}(\Delta\tau, \omega)$ with respect to $\Delta\tau$, denoted as $S_{\bar{O}}(\gamma, \omega)$, it is observed that (6.54) corresponds to the product between two factors. Therefore, $S_{\bar{O}}(\gamma, \omega)$ can be written as

$$S_{\bar{O}}(\gamma, \omega) = \delta\left(\gamma + \omega\cos\alpha\right) * S_O(\gamma, \omega\cos\alpha), \tag{6.55}$$

where the convolution happens along the spatial frequency dimension γ. This convolution is shown in Fig. 6.10. Simulation have been performed and the 2D spectrum associated to the mixing of a deterministic and a stochastic process is given in Fig. 6.11. The AR-2 process with poles at -9 and -10 described in Fig. 6.6(b) has been chosen together with a deterministic plane wave with an angle of arrival of 85°.

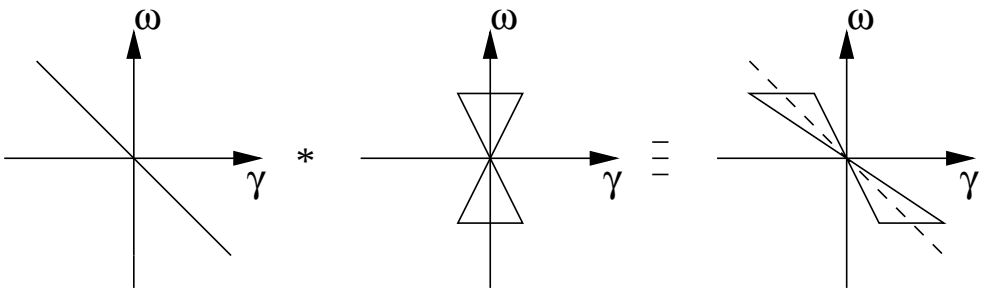


Figure 6.10: Convolution between the deterministic and stochastic 2D spectra to obtain the PSD corresponding to a process mixing a deterministic and a stochastic process.

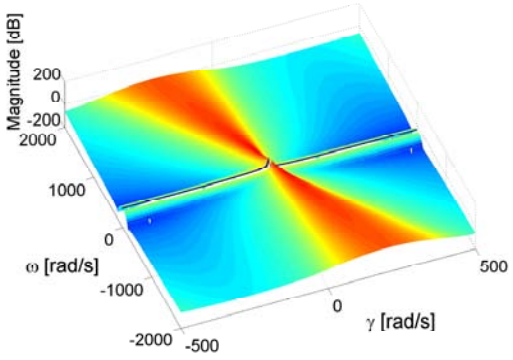


Figure 6.11: PSD associated to a process mixing a deterministic plane wave with angle of arrival of $85°$ with a stochastic AR-2 process with poles at -9 and -10.

6.6 Simulations and Experimental Measurements

In this section, we present results obtained by simulations of discrete AR processes as well as real measurements.

6.6.1 Simulation results

In a first step, simulations were carried out to show that, the higher the order of the AR process, the smoother is the trajectory, as discussed in Section 6.3.3. A smoother trajectory relates into a smoother time of arrival from the source to the receiver. The setup considered is the one presented in Fig. 6.4 and the time of arrival is defined as $d + \frac{Y(\tau)}{c}$ as in (6.17). Note that in the simulations, the continuous AR models are simulated by means of discrete AR models. The poles of the continuous system have to be translated to their corresponding discrete-time version. Considering a continuous AR system with a pole at λ, the corresponding discrete system will exhibit a pole at $e^{\lambda T}$ with T the sampling period of the discrete system. For example, a continuous AR-1 process with pole at -5, corresponds to a discrete system with a pole at position .995 for a sampling period of $T = 1/1000$ s. For this AR-1 system, it can be seen in Fig. 6.12(a) that the time of arrivals between the source and the receiver following this AR model are not very smooth. With an AR-2 process with two poles at .995, the time of arrivals get smoother as shown in Fig. 6.12(b).

Further, simulations were done to add the effect of reflections introduced by reverberation. We still consider the setup of Fig. 6.4 but the source and the receiver are inside a room. Next to the direct path, reflections on the walls need to be considered. These reflections can be seen as virtual sources. As discussed in Section 6.4, the frequency support of the PSD corresponding to

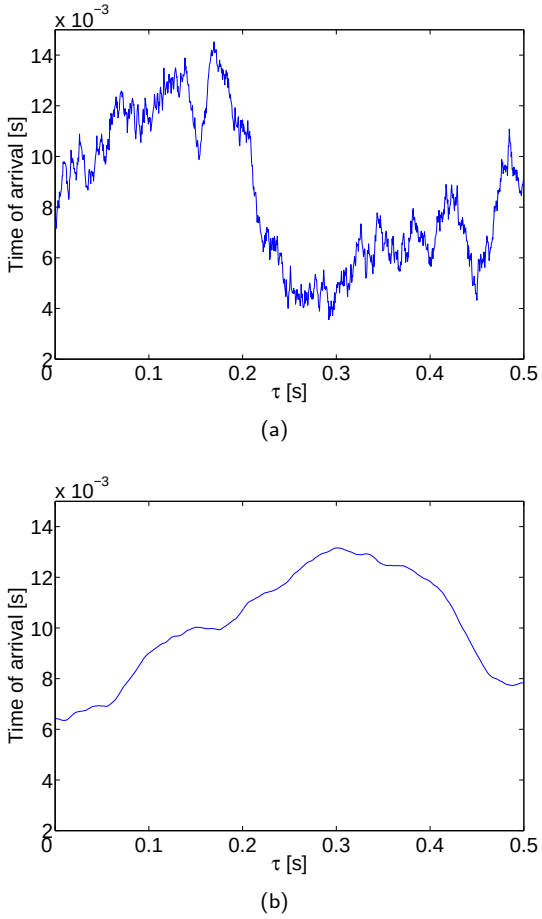


Figure 6.12: Smoothness of the AR processes. (a) Discrete AR-1 process simulated with a pole at .995. (b) Discrete AR-2 process simulated with two poles around .995.

the reverberant field is included in the support of the free field PSD. This is also observed in the simulations, where Fig. 6.13(a) represents the PSD of the spatio-temporal CIR for the free field case and Fig. 6.13(b) for a room with reverberation.

6.6.2 Experimental results

Experiments have been carried out in a sound insulated room. A loudspeaker (Genelec $1029A$) was used to generate an excitation signal, recorded by a microphone Beyerdynamic $MC-740$. The precise positioning of the microphone for the measurement of the RIRs was performed using a Pan/Tilt Unit $PTU-D46-70$. This device allows the rotation of the microphone with a

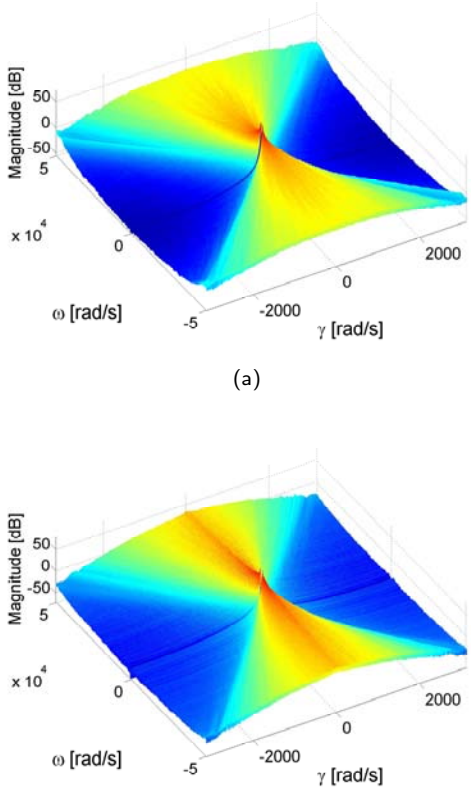
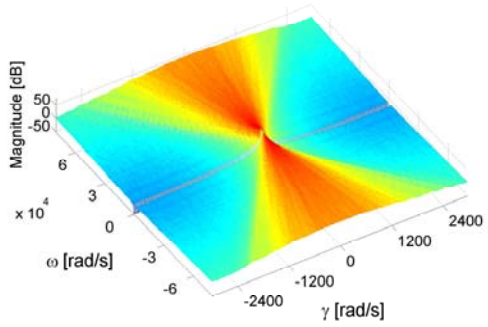
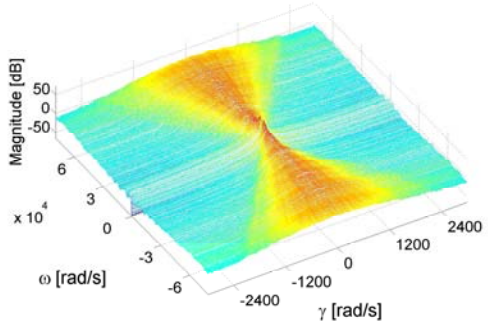


Figure 6.13: PSD of spatio-temporal CIR. (a) For a free field case. (b) For a reverberant room.

precision of .03°. Using this equipment, a set of 2000 RIRs were measured along a circle of radius of .7 m. The RIRs were measured using a logarithmic sweep [86]. Positions matching a discrete AR-2 process with poles at .995 and .99 (equivalent to a continous system with poles at -5 and -10) were chosen in the measured database and the PSD obtained for the measured database was compared to the simulated PSD. The simulated and experimental periodograms for this AR process are shown in Fig. 6.14(a) and (b). As can be seen, there is a strong agreement in the support while the measured PSD appears to be noisier.



(a)



(b)

Figure 6.14: PSD of spatio-temporal CIRs for a discrete AR-2 process with poles at .995 and .99, (a) obtained by simulation; (b) obtained using experimental room impulse responses.

6.7 Conclusion

In this chapter, we have presented a stochastic model for a spatially varying channel in the case of a receiver moving along a random trajectory with respect to a fixed source. We have modeled the trajectory of the receiver as an autoregressive model where the poles of the system control the smoothness of the path. Theoretical results are presented for the AR-2 case and generalized to any AR-N systems. The power spectral density corresponding to the CIR as a function of temporal and spatial frequency has been studied. Simulations and experimental results have been shown and compared to the presented theory. This theoretical study gives a basis for a stochastic model of time-varying channels as found in communication environments where a physical process (e.g. user movement) drives the time-varying impulse response of a channel.

6.8 Acknowledgments

The authors would like to thank Prof. A. Goldschmidt for providing us with a list of key references in the field of CIR modeling.

Chapter 7

Conclusion

7.1 Thesis Summary

The main topic of interest of this thesis lies in the sampling and interpolation of fields governed by the wave equation. While most of the results have been presented for the acoustic case, they can easily be applied to any other field satisfying the wave equation such as the electromagnetic wave equation. Only the propagation speed of the phenomenon needs to be adapted.

Note that to keep a good balance between theoretical and experimental work, each theoretical result was validated with simulations and experimental measurements. These measurements consisted mostly in acoustic and electromagnetic impulse response measurements in rooms. These measurements are available at the following address [2].

In this thesis, we have introduced the "plenacoustic function" and defined it as the sound field generated by one source or a distribution of sources. The function characterizes the sound field both along the spatial and the temporal dimensions. One of the purposes of this thesis was to answer the following question: " How many microphones need to be placed along a geometry to be able to reconstruct the sound field at any position along this geometry?" We solved this question for different geometries such as a line, a plane, a circle and the three-dimensional space filled with microphones. To solve the problem, the Fourier representation of the sound field was used.

For the linear case, we calculated the two-dimensional Fourier transform of the sound field along space and time and observed that the spectrum presented a very specific butterfly shape. While most of the energy lies in the butterfly, some small energy still remains outside of the essential support. By characterizing this amount of energy, a spatial sampling theorem was developed. This theorem presents the trade-off existing between spatial sampling frequency and

error on the reconstruction of the sound field. Also, a larger spatial sampling frequency is required to achieve a similar error on reconstruction for signals with larger maximal temporal frequencies. We compared the presented theory with measurements of room impulses responses in the acoustic and electromagnetic case.

For the case of the sound field studied along a circle, a similar angular sampling theorem was presented. Acoustic and electromagnetic room impulse responses along a circle were measured and used for comparison with the presented theory. As an application of this sampling study, we considered the problem of sampling and interpolation of head-related transfer functions (HRTFs). These filters model the effect of the shape of the head, body and pinnae on the sound measured at the entrance of the ear canal. HRTFs are widely used in audio playback over headphones to increase realism and spatial perception. Based on the sampling theorem, it was shown that HRTFs need to be sampled every 5° for reconstruction at any angle in the azimuthal plane. To reduce the number of necessary HRTFs, we developed a technique considering the decomposition of HRTFs in complex envelopes and carrier signals. With this method, good interpolation results were obtained in case of angular undersampling by a factor of two or four. We compared the technique with other existing methods for different HRTFs databases and showed that our technique achieves the best reconstruction of HRTFs in a mean squared sense.

The Fourier representation of the sound field was further used for the measurement of room impulse responses using a loudspeaker or a microphone moving along a certain trajectory (e.g. circular or linear trajectory). The presented algorithm reconstructs the different static room impulse responses from the recorded microphone signal. For this purpose, we need to carefully design the excitation signal provided to the loudspeaker. Also, the maximal speed for the moving element was given as a function of the different parameters of the system. As the main result of this work, we showed that all HRTFs in the horizontal plane can be recorded in less then one second. We verified the presented algorithm with an experimental setup using a moving microphone holder to reconstruct the different room impulse responses along the trajectory.

In the last part of the thesis, we introduced the "stochastic plenacoustic function" to describe the sound field gathered along a stochastically generated trajectory. These trajectories can model the movement of a person in a conference room in the acoustic case, or a moving car passing near an antenna in the case of the electromagnetic wave equation. We studied spatio-temporal channel impulse responses between a fixed source and a moving receiver along a stochastic trajectory. The considered trajectories are characterized by a certain smoothness due to the inertia present in any movement satisfying Newtonian

equations. We modeled the smooth trajectories using continuous-time autoregressive processes.With this model, we observed that the power spectral density of the spatio-temporal channel impulse responses presents a butterfly shape. This butterfly shape was shown to directly depend on the smoothness of the followed trajectory. We further studied this power spectral density in case of multipath reverberation. Experimental results were carried out in the acoustic case to confirm the theoretical results.

7.2 Future Research

In this section, we discuss some future research ideas and directions. Some represent minor work that still needs to be performed and other describe more general directions for future research. We present these ideas per topic covered in this thesis.

7.2.1 Channel impulse response modeling

Based on the results obtained for the modeling and the analysis of the spatio-temporal channel impulse responses, a few next steps are presented.

First, based on the different characteristics of the system (e.g. smoothness of the trajectory, signals emitted), we would like to develop approximation formulas or rules of thumb to predict the spatial bandwidth of the channel. This would correspond to find approximations for the integral given by (6.25). This information is of interest to characterize time-varying channels.

Second, a lot of adaptive algorithms, such as echo cancelation in acoustics or multipath equalization in electromagnetism, can be tuned to channel characteristics. Using information based on the butterfly support of the PSD of the channel could help in choosing the different parameters in the adaptive algorithm. Based on the intuition built in this thesis, we would like to develop a "motion-based" adaptive algorithm. The algorithm would decompose the input in temporal frequency subbands. The rate of adaptation of each subband would be following a rule dictated by the butterfly shaped PSD, which is depending on the smoothness of the motion. Typically, at low temporal frequency the rate of adaptation could be kept slow while it would increase with the temporal frequency. A quantitative analysis of the frequency dependent rate of adaptation is matter of current research.

Third, in the presented chapter, the considered trajectories represent a movement along a one-dimensional line. Generalization to 2D or 3D trajectories needs to be performed to make the analysis more realistic.

7.2.2 Impulse response measurements

A future direction of more applied research considers the developed technique in Chapter 5, where moving microphones or loudspeakers were used for impulse response measurements. We would like to develop the presented technique and create a prototype of such a measurement system. For this purpose, a few technical issues need to be resolved. First, a motorized system needs to be designed where the motor noise would be correctly attenuated. Another challenging point is the synchronization between the different parts of the system (sound card and motor). For the system to work optimally, the start (and stop) of the sound acquisition needs to be synchronized with the motor movement. Further, different aspects related to the system should be further investigated. For example, the effect of air turbulence created by the moving element on the measured impulse responses should be analyzed. Developing a working prototype could promote the use of individualized HRTFs in headphones playback. The current duration for the measurement of an HRTF database is of about one hour using standard techniques [57]. With our technique, the duration could be reduced by a factor ten to fifty. Therefore, it is really of importance to develop a system delivering fast and robust measurements.

7.2.3 HRTF interpolation

Further work on HRTFs interpolation includes another assessment method using perceptual tests. Listening tests should be carried out to validate the obtained results. In this thesis, the emphasis was given to signal processing and the purpose was to reconstruct a signal that would be as close as possible from the original one in a mean squared error sense. Nevertheless, listening tests should be carried out to test the localization performance of subjects using our interpolated HRTFs. Also, next to localization, the subjects should be asked about any timbre change or possible distortions perceived in the reconstructed sound.

Another future work in the field of HRTFs interpolation, is the development of a perceptually based assessment method. As will be described below, it is based on a conjecture that first needs to be verified. As appeared from our HRTF interpolation results, for sources located on the same side as the ear (ipsilateral ear), the interpolation results were very satisfying. For the sources located on the other side of the ear (contralateral ear), the head shadowing effect led to worse results. In [20], it is mentioned that the head shadowing effect can attenuate signals with 20 dB. For localization, the brain uses cues based on the two signals gathered at the ears [20]. In the presented technique, it is the weak signal used by the brain which might be less correctly interpolated. Based

on the conjecture that less accurate interpolation of the weak signal might not be critical for localization[1], we would like to present a new assessment criterion. Until now, the MSE has been considered for each ear separately. We would like to develop a MSE taking into account both left and right HRTFs. This adapted MSE could be expressed as:

$$\mathrm{MSE}(\theta_0) = 10\log_{10} \frac{\sum_{n=0}^{T} \left[|h_l(\theta_0, n) - h_{l,e}(\theta_0, n)|^2 + |h_r(\theta_0, n) - h_{r,e}(\theta_0, n)|^2 \right]}{\sum_{n=0}^{T} \left[|h(\theta_0, n)|^2 + |h(\theta_0, n)|^2 \right]},$$

with T the number of time sample of each HRTF. The left and right original HRTFs are denoted as $h_l(\theta_0, n)$ and $h_r(\theta_0, n)$ while the left and right estimated HRTFs are $h_{l,e}(\theta_0, n)$ and $h_{r,e}(\theta_0, n)$. This new assessment for HRTFs would have the advantage of being easily implementable and of indirectly taking some perceptual considerations into account.

[1]There is no evidence that this statement is correct, it needs to be verified.

Appendix A

Mathematical Derivations of Multidimensional PAF Spectra

In this appendix, we give the precise mathematical derivations for the calculation of the spectra of the PAF for different geometries. For the case of the infinite line, the spectrum of the PAF is given in Section A.1. For the infinite plane, the spectrum is given in Section A.2 and for the three-dimensional space, in Section A.3.

A.1 Derivation of the 2D-FT of the PAF on a line

$$p(x,t) = \frac{\delta\left(t - \frac{\sqrt{(x-x_s)^2+d^2}}{c}\right)}{4\pi\sqrt{(x-x_s)^2+d^2}}. \tag{A.1}$$

The 2D-FT of (A.1) is

$$\hat{p}(\phi,\omega) = \int_{x=-\infty}^{\infty}\int_{t=-\infty}^{\infty} p(x,t)e^{-j(\phi x+\omega t)}\mathrm{d}t\mathrm{d}x \tag{A.2}$$

Call $u = x - x_s$, therefore

$$\hat{p}(\phi,\omega) = \frac{e^{-j\phi x_s}}{4\pi}\int_{u=-\infty}^{\infty}\frac{e^{-j(\phi u+\frac{\omega}{c}\sqrt{u^2+d^2})}}{\sqrt{u^2+d^2}}\mathrm{d}u \tag{A.3}$$

$$= \frac{e^{-j\phi x_s}}{2\pi}\int_{u=0}^{\infty}\frac{e^{-j\frac{\omega}{c}\sqrt{u^2+d^2}}\cos(\phi u)}{\sqrt{u^2+d^2}}\mathrm{d}u. \tag{A.4}$$

Using existing formulae in [53], one obtains:

$$\hat{p}(\phi,\omega) = -\frac{j}{4}e^{-j\phi x_s}H_0^*\left(d\sqrt{(\frac{\omega}{c})^2-\phi^2}\right). \tag{A.5}$$

A.2 Derivation of the 3D-FT of the PAF on a plane

$$p(x_m,y_m,t) = \frac{\delta(t-\frac{a}{c})}{4\pi a},$$

with $a = \sqrt{(x_m-x_s)^2+(y_m-y_s)^2+(z_m-z_s)^2}$. Call $\mathbf{u}_m = [x_m,y_m]^{\mathrm{T}}$ and $\boldsymbol{\phi} = [\phi_x,\phi_y]^{\mathrm{T}}$. One can calculate the spectrum of this function

$$\hat{p}(\boldsymbol{\phi},\omega) = \int_{\mathbb{R}^2}\int_{\mathbb{R}}\frac{\delta(t-\frac{a}{c})}{4\pi a}e^{-j(\boldsymbol{\phi}^{\mathrm{T}}\mathbf{u}_m+\omega t)}\mathrm{d}\mathbf{u}_m\mathrm{d}t. \tag{A.6}$$

This expression can be rewritten as

$$\hat{p}(\boldsymbol{\phi},\omega) = \int_{\mathbb{R}^2}\frac{e^{-j(\boldsymbol{\phi}^{\mathrm{T}}\mathbf{u}_m+\omega\frac{a}{c})}}{4\pi a}\mathrm{d}\mathbf{u}_m. \tag{A.7}$$

Introducing $\mathbf{u} = [x_m-x_s, y_m-y_s]^{\mathrm{T}}$, the integral is rewritten as

$$\hat{p}(\boldsymbol{\phi},\omega) = e^{-j(\phi_x x_s+\phi_y y_s)}\int_{\mathbb{R}^2}\frac{e^{-j(\boldsymbol{\phi}^{\mathrm{T}}\mathbf{u}+\omega\frac{a}{c})}}{4\pi a}\mathrm{d}\mathbf{u}. \tag{A.8}$$

Call $r^2 = x^2+y^2$, $\Phi = e^{-j(\phi_x x_s+\phi_y y_s)}$, $z = z_m - z_s$ and $\phi_q^2 = \phi_x^2+\phi_y^2$. The integral can be rewritten as [25]:

$$\hat{p}(\boldsymbol{\phi},\omega) = \frac{\Phi}{2}\int_{r=0}^{+\infty}\frac{rJ_0(\phi_q r)}{\sqrt{r^2+z^2}}e^{-j\left(\omega\frac{\sqrt{r^2+z^2}}{c}\right)}\mathrm{d}r. \tag{A.9}$$

Call $m = \sqrt{r^2+z^2}$, we have that $r = \sqrt{m^2-z^2}$, and also $\mathrm{d}m = \frac{r}{\sqrt{r^2+z^2}}\mathrm{d}r$. The integral becomes

$$\hat{p}(\boldsymbol{\phi},\omega) = \frac{\Phi}{2}\int_{m=|z|}^{+\infty}J_0\left(\phi_q\sqrt{m^2-z^2}\right)e^{-j\omega\frac{m}{c}}\mathrm{d}m. \tag{A.10}$$

Using existing formulae in [53], one obtains that

$$\hat{p}(\phi_x,\phi_y,\omega) = \begin{cases} \frac{-j\Phi}{2}\frac{e^{-j|z|\sqrt{(\frac{\omega}{c})^2-\phi_q^2}}}{\sqrt{(\frac{\omega}{c})^2-\phi_q^2}} & \text{for} \quad |\phi_q| \le \frac{\omega}{c} \\ \frac{\Phi}{2}\frac{e^{-|z|\sqrt{\phi_q^2-(\frac{\omega}{c})^2}}}{\sqrt{\phi_q^2-(\frac{\omega}{c})^2}} & \text{for} \quad \frac{\omega}{c} \le |\phi_q|. \end{cases} \tag{A.11}$$

A.3 Derivation of the 4D-FT of the PAF in space

$$p(x_m, y_m, z_m, t) = \frac{\delta(t - \frac{a}{c})}{4\pi a},$$

with $a = \sqrt{(x_m - x_s)^2 + (y_m - y_s)^2 + (z_m - z_s)^2}$. Call $\mathbf{u}_m = [x_m, y_m, z_m]^{\mathrm{T}}$ and $\boldsymbol{\phi} = [\phi_x, \phi_y, \phi_z]^{\mathrm{T}}$. The spectrum of this function can be calculated

$$\hat{p}(\boldsymbol{\phi}, \omega) = \int_{\mathbb{R}^3} \int_{\mathbb{R}} \frac{\delta(t - \frac{a}{c})}{4\pi a} e^{-j(\boldsymbol{\phi}^{\mathrm{T}} \mathbf{u}_m + \omega t)} \mathrm{d}\mathbf{u}_m \mathrm{d}t. \tag{A.12}$$

By introducing $\mathbf{u} = [x_m - x_s, y_m - y_s, z_m - z_s]^{\mathrm{T}}$, the integral can be rewritten as

$$\hat{p}(\boldsymbol{\phi}, \omega) = e^{-j(\phi_x x_s + \phi_y y_s + \phi_z z_s)} \int_{\mathbb{R}^3} \frac{e^{-j(\boldsymbol{\phi}^{\mathrm{T}} \mathbf{u} + \omega \frac{a}{c})}}{4\pi a} \mathrm{d}\mathbf{u}. \tag{A.13}$$

Call $\phi_s^2 = \phi_x^2 + \phi_y^2 + \phi_z^2$, the integral can be rewritten as [25]:

$$\hat{p}(\boldsymbol{\phi}, \omega) = e^{-j(\phi_x x_s + \phi_y y_s + \phi_z z_s)} \int_{m=0}^{+\infty} \sin(\phi_s m) e^{-jm\frac{\omega}{c}} \mathrm{d}m. \tag{A.14}$$

The solution of this integral can be found in [53]:

$$\hat{p}(\phi_x, \phi_y, \phi_z, \omega) = \frac{e^{-j(\phi_x x_s + \phi_y y_s + \phi_z z_s)}}{\phi_s^2 - (\frac{\omega}{c})^2}. \tag{A.15}$$

Appendix B

Spatial Frequency Decay of the PAF for Microphone Lines Non-Parallel to a Wall

In this appendix, we consider the case of a line of microphones not parallel to a wall. We also assume that the line of microphones is included in a plane parallel to the floor and the ceiling of the room. Two possible configurations are studied: the case where the coefficient of direction of the line is rational or non-rational. For simplicity we consider in the rest of this section that $\frac{L_y}{L_x} \in \mathbb{Q}$.

1. The case of a line with rational coefficient of direction is first considered. In Fig. B.1 a room is shown with a source and all associated virtual sources. The line where the field is to be studied follows a direction $\mathbf{g} = [\mathbf{g_x}, \mathbf{g_y}]$, with $\frac{g_y}{g_x}$ a rational number. It can be observed that the sound pressure measured on the distance denoted as L along $\mathbf{g}$ is enough to know the pressure on the whole line. This follows from the periodicity of the sources as shown in Fig. B.1. To study the spatial decay of the spectrum of the PAF corresponding to this line, the same formalism as in Section 3.3.2 is followed. Due to the periodicity L of the scheme, the spatial frequency of the spectrum is only defined for discrete values $n\phi_0$, with $\phi_0 = \frac{2\pi}{L}$.

   To obtain the spatial decay of the spectrum, one still needs to study the effect of the infinite number of sources in the region ranging from abscissa 0 to L along $\mathbf{g}$. One can remark that due to the rational character of the coefficient of direction, there exists a periodicity in the sources along the direction $\mathbf{g}_\perp$ being orthogonal to $\mathbf{g}$. This periodicity depends on

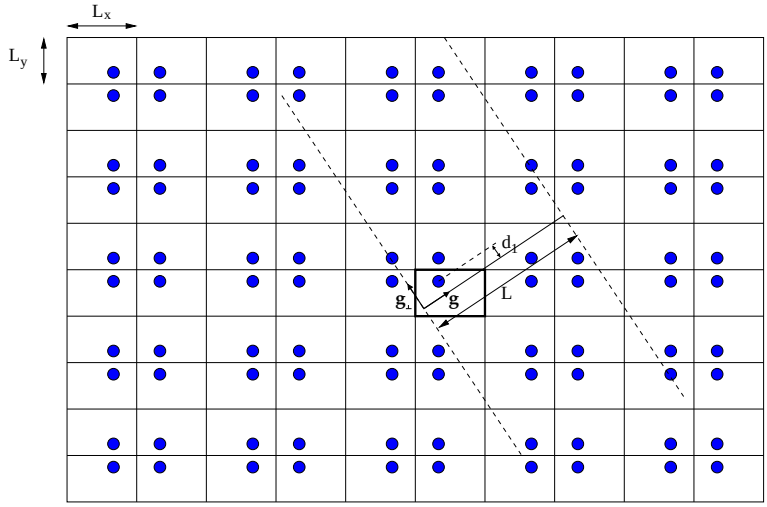


Figure B.1: A room is shown with its virtual sources. In the middle of the figure, the original room is shown in bold. A line with direction **g** is shown.

the direction **g** but also on the size of the room. This periodicity is $L_p = \sqrt{(2\alpha L_x)^2 + (2\beta L_y)^2}$, with α and β the smallest possible integer numbers satisfying the relation

$$\frac{\alpha}{\beta} = \frac{-g_y L_y}{g_x L_x}.$$

Knowing this periodicity, the spectrum can easily be obtained similarly to the derivation of Section 3.3.2. The only difference is that here more than four virtual sources need to be considered and that the periodicity L_y needs to be replaced by L_p. The final result is here also that

$$\hat{p}(n\phi_0, \omega) = O(e^{-d_1 n\phi_0}),$$

with d_1 the distance between the closest source and the line of microphones.

2. The case of a line having a non-rational coefficient of direction is now considered. In order to study the decay of the spectrum of the PAF on that line, one studies the decay of the spectrum of the plane that contains this line and that is also parallel to the ceiling and the floor of that room. Further, by using the projection-slice theorem [25], one can study the spectrum corresponding to a specific line of the PAF (our line of interest). By observing the repetitions of the sources, one can observe that the PAF on the plane is periodic both in the x and in the y directions with periodicity $2L_x$ and $2L_y$ respectively. Therefore, the spectrum of

the PAF is only defined for discrete values m and n with $\phi_{0x} = \frac{\pi}{L_x}$ and $\phi_{0y} = \frac{\pi}{L_y}$, $\hat{p}(m\phi_{0x}, n\phi_{0y}, \omega)$. Without considering the periodicity, one can remark that the PAF spectrum can also be rewritten as:

$$\hat{p}(\phi_x, \phi_y, \omega) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \delta(\phi_x - m\phi_{0x})\delta(\phi_y - n\phi_{0y})\hat{p}(m\phi_{0x}, n\phi_{0y}, \omega). \quad \text{(B.1)}$$

In order to study the spectrum of the PAF along the line of interest, one needs to apply the projection-slice theorem. This theorem says that to study a slice of the PAF in time domain (in this case along the line of interest), one needs to project the spectrum along the corresponding line in frequency. Consider that the line of interest has a direction $\mathbf{g} = [g_x, g_y]$, the projection in the frequency domain happens along the direction $\mathbf{g}_\perp = [-g_y, g_x]$ as shown in Fig. B.2. Further, call ϕ_g the abscissa along the line G. Also, consider the lines $G_\perp(\phi_g)$ orthogonal to $\mathbf{g}$ and at positions ϕ_g on G. The spectrum of the PAF along the line of interest is given by:

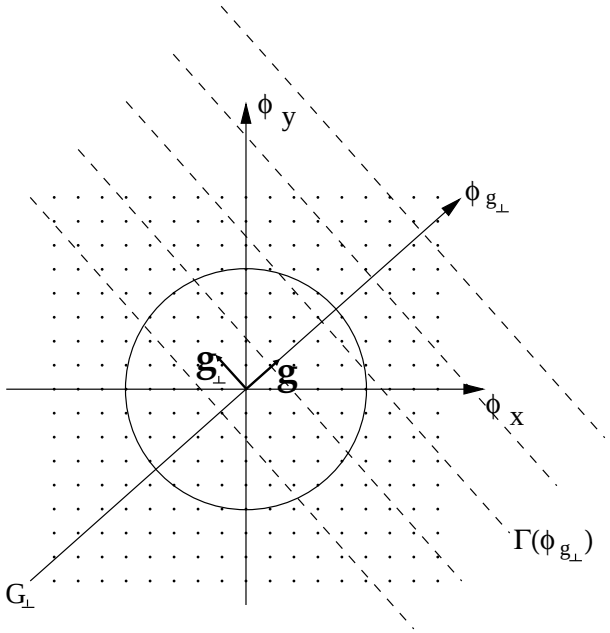


Figure B.2: Top view of the spectrum of the PAF for the plane of microphones. The spectrum is only defined for discrete values. The spectrum of the line of interest corresponds to the projection the spectrum of the PAF along directions $G_\perp$.

$$\hat{p}_{\mathbf{g}}(\phi_g, \omega) = \int_{G_\perp(\phi_g)} \hat{p}(\phi_x, \phi_y, \omega) \mathrm{d}s, \quad \text{(B.2)}$$

with s the abscissa along $G_\perp(\phi_g)$.

As $\frac{g_y}{g_x}$ is a non-rational value, it can be observed that each sample of the spectrum of the plane is projected on distinct positions along G. When the source is in the same plane as the line of microphones, it was shown in (3.51) that the decay is not anymore exponential but becomes only linear. Therefore, in this case, the decay of the PAF spectrum decays at least linearly. This decay is much slower than the one obtained in the case of a rational direction.

When no source is located inside the plane of interest, we consider the more general case where sources are also repeated along the z axis. The decay of the spectrum will be shown to become exponential at the condition that no sources are located inside the plane of the microphones. We consider a plane of microphones parallel to the floor and the ceiling containing the line G and the original source with all its virtual sources outside of that plane. As shown in the image method [13], the original source is first repeated to create the seven first virtual sources. Then, these eight sources are further repeated with periodicity $2L_x$, $2L_y$ and $2L_z$ in the x, y and z directions respectively. Similarly to the 2D case, the PAF is periodic along the x and y directions and therefore the PAF spectrum is discrete in ϕ_x and ϕ_y. What differs from the 2D case, is that each of the eight original sources to be considered is now repeated along the z axis with periodicity $2L_z$. For the simplicity of the calculations, we only consider one of the eight original sources and calculate the spectrum of the PAF for that source repeated in the three directions. Further the other seven original sources are considered and it will be shown that the other sources are negligible when studying the decay of the PAF spectrum for large spatial frequencies.

Due to the construction of the virtual sources, we have that among the eight first sources, four are located at distance z_1 from the plane of interest, with $z_1 \leq L_z$. The other four sources are located at a distance $2L_z - z_1$ from the plane. Call s_1 one of the sources that is separated from the plane by a distance z_1. The spectrum of the PAF along the plane in the presence of s_1 without repetitions is (for large spatial frequencies):

$$\hat{p}(m\phi_{0x}, n\phi_{0y}, \omega) = e^{-j(m\phi_{0x}x_s + n\phi_{0y}y_s)} \frac{e^{-z_1\Gamma(m,n)}}{2\Gamma(m,n)}, \tag{B.3}$$

with $\Gamma(m,n) = \sqrt{(m\phi_{0x})^2 + (n\phi_{0y})^2}$.

When considering s_1 with all its repetitions along the z axis, we obtain

$$\hat{p}(m\phi_{0x}, n\phi_{0y}, \omega) = e^{-j(m\phi_{0x}x_s + n\phi_{0y}y_s)} \sum_{i=-\infty}^{\infty} \frac{e^{-|z_1+2L_z i|\Gamma(m,n)}}{2\Gamma(m,n)}$$
$$= \frac{e^{-j(m\phi_{0x}x_s + n\phi_{0y}y_s)}}{2\Gamma(m,n)} \sum_{i=0}^{\infty} \left(e^{-(z_1+2L_z i)\Gamma(m,n)} + e^{-(z_1'+2L_z i)\Gamma(m,n)} \right),$$

with $z_1' = 2L_z - z_1$. We therefore can write that

$$\hat{p}(m\phi_{0x}, n\phi_{0y}, \omega) = \frac{e^{-j(m\phi_{0x}x_s + n\phi_{0y}y_s)}}{2\Gamma(m,n)} \left(\frac{e^{-z_1\Gamma(m,n)}}{1 - e^{-2L_z\Gamma(m,n)}} + \frac{e^{(z_1-2L_z)\Gamma(m,n)}}{1 - e^{-2L_z\Gamma(m,n)}} \right).$$

Considering now, the other sources, we can observe that asymptotically for large m and n, the decay of the spectrum is of the following order:

$$\hat{p}(m\phi_{0x}, n\phi_{0y}, \omega) = O\left(\frac{e^{-z_1\Gamma(m,n)}}{\Gamma(m,n)} \right). \tag{B.4}$$

It is thus shown that the PAF spectrum decays exponentially when studied in a 3D environment.

Here again, as $\frac{g_y}{g_x}$ is a non-rational value, it can be observed that each sample of the spectrum of the plane is projected on distinct positions along G. Call $\phi_g(m_0, n_0)$ the abscissa of the projection of a specific point of the spectrum with coordinates (m_0, n_0). We have that $\phi_g(m_0, n_0) < \Gamma(m_0, n_0)$. Therefore we have that the projection of the PAF spectrum on G decays at least as fast as (B.4).

Finally, once the spectrum of the infinite line is studied, one can still consider the fact that the sound field is only studied inside the room. This part happens similarily to the spatial windowing in Section 3.5.

Appendix C

Temporal Frequency Decay of the PAF in a Rectangular Room

We have observed in (3.28) that the PAF spectrum could be expressed as

$$\hat{p}(n_0\phi_0,\omega) \sim c_4(n\phi_0)\sum_{i=-\infty}^{\infty}\left(\frac{e^{-jD_{1,i}\frac{\omega}{c}}}{\sqrt{D_{1,i}\omega}}+\frac{e^{-jD_{2,i}\frac{\omega}{c}}}{\sqrt{D_{2,i}\omega}}\right).$$

In this appendix, we investigate the convergence of $\hat{p}(n_0\phi_0,\omega)$.

The first and second terms are denoted as $\hat{m}_1(\omega) = c_4(n\phi_0)\sum_{i=-\infty}^{\infty}\frac{e^{-jD_{1,i}\frac{\omega}{c}}}{\sqrt{D_{1,i}\omega}}$ and $\hat{m}_2(\omega) = c_4(n\phi_0)\sum_{i=-\infty}^{\infty}\frac{e^{-jD_{2,i}\frac{\omega}{c}}}{\sqrt{D_{2,i}\omega}}$, respectively. The term $\hat{m}_1$ can be rewritten as:

$$\begin{aligned}\hat{m}_1(\omega) &= \frac{c_4(n\phi_0)}{\sqrt{\omega}}\left(e^{-jd_1\frac{\omega}{c}}+\sum_{i=1}^{\infty}\frac{e^{-j\frac{\omega}{c}(d_1+2L_yi)}}{\sqrt{d_1+2L_yi}}+\sum_{i=-\infty}^{-1}\frac{e^{-j\frac{\omega}{c}|d_1+2L_yi|}}{\sqrt{|d_1+2L_yi|}}\right)\\ &= \frac{c_4(n\phi_0)}{\sqrt{\omega}}\left(e^{-jd_1\frac{\omega}{c}}+\sum_{i=1}^{\infty}\frac{e^{-j\frac{\omega}{c}(d_1+2L_yi)}}{\sqrt{d_1+2L_yi}}+e^{-jd_1^{'}\frac{\omega}{c}}+\sum_{i=1}^{\infty}\frac{e^{-j\frac{\omega}{c}(d_1^{'}+2L_yi)}}{\sqrt{d_1^{'}+2L_yi}}\right)\end{aligned}$$

with $d_1^{'} = 2L_y - d_1$. Call the second and fourth terms of the previous expression $\hat{u}_1(\omega)$ and $\hat{u}_2(\omega)$, respectively. These two expression are shown to converge using the Dirichlet convergence test[1]. Considering $\hat{u}_1(\omega)$, the Dirichlet convergence test can be applied[2] with $\{a_n\} = e^{-j\frac{\omega}{c}(d_1+2L_yi)}$ and $\{b_n\} = \frac{1}{\sqrt{d_1+2L_yi}}$.

[1] Let $\{a_n\}$ and $\{b_n\}$ be sequences of real numbers such that $\{\sum_{i=0}^{n} a_i\}$ is bounded and $\{b_n\}$ decreases with 0 as limit. Then $\sum_{n=0}^{\infty} a_n b_n$ converges.

[2] In this case we have to split the exponential in a_n in its sine and cosine expression to be

This proves the convergence of $\hat{u}_1(\omega)$. Similarly, it can also be shown that $\hat{u}_2(\omega)$ converges. Furthermore, $\hat{u}_1(\omega)$ and $\hat{u}_2(\omega)$ are periodic function of ω. Therefore, they can be upperbounded and are converging for all possible values of ω. This leads us to the conclusion that $\hat{s}_1(\omega)$ behaves asymptotically as

$$\hat{s}_1(\omega) = \frac{c_5(\omega)}{\sqrt{\omega}},$$

with $c_5(\omega)$ a bounded function of ω. We obtain the same result for $\hat{s}_2(\omega)$ and therefore we have that

$$\hat{p}(n_0\phi_0, \omega) \sim \frac{c_6(\omega)}{\sqrt{\omega}},$$

with $c_6(\omega)$ a bounded function of ω.

able to apply the convergence test since it was defined for real expressions.

Appendix D

Optimal Alignment of HRTF Carriers

In this section we present the derivation of the complex exponential $c_i(\theta) = e^{j\alpha_{\theta,i}}$ minimizing the following cost function.

$$J_i = \int_{\omega_i}^{\omega_{i+1}} \left| \tilde{h}_i(\theta_r, \omega) - \tilde{h}_i(\theta, \omega) c_i(\theta) \right|^2 \mathrm{d}\omega, \tag{D.1}$$

with ω_i and ω_{i+1} the start and stop frequencies of band i. First, let us rewrite J_i as follows:

$$\begin{aligned} J_i = \int_{\omega_i}^{\omega_{i+1}} \Big[|\tilde{h}_i(\theta_r, \omega)|^2 + |\tilde{h}_i(\theta, \omega)|^2 - \tilde{h}_i^*(\theta, \omega)\tilde{h}_i(\theta, \omega)c_i(\theta) \\ - \tilde{h}_i(\theta, \omega)\tilde{h}_i^*(\theta, \omega)c_i^*(\theta) \Big] \mathrm{d}\omega. \end{aligned} \tag{D.2}$$

The value of $c_i(\theta)$ which minimizes J_i is obtained by calculating the root of the first derivative of J_i with respect to $\alpha_{\theta,i}$. By deriving J_i a second time with respect to $\alpha_{\theta,i}$, it can then be verified wether the root corresponds to a minimum or a maximum of the function.

$$\frac{\partial J_i}{\partial \alpha_{\theta,i}} = j \int_{\omega_i}^{\omega_{i+1}} \left(-\tilde{h}_i^*(\theta, \omega)\tilde{h}_i(\theta, \omega)e^{j\alpha_{\theta,i}} + \tilde{h}_i(\theta, \omega)\tilde{h}_i^*(\theta, \omega)e^{-j\alpha_{\theta,i}} \right) \mathrm{d}\omega = 0. \tag{D.3}$$

From (D.3), we have that

$$e^{j2\alpha_{\theta,i}} = \frac{\int_{\omega_i}^{\omega_{i+1}} \tilde{h}_i(\theta, \omega)\tilde{h}_i^*(\theta, \omega)\mathrm{d}\omega}{\int_{\omega_i}^{\omega_{i+1}} \tilde{h}_i^*(\theta, \omega)\tilde{h}_i(\theta, \omega)\mathrm{d}\omega}. \tag{D.4}$$

Therefore, the roots of $\frac{\partial J_i}{\partial \alpha_{\theta,i}}$ are

$$e^{j\alpha_{\theta,i}} = \pm\sqrt{\frac{\int_{\omega_i}^{\omega_{i+1}} \tilde{h}_i(\theta,\omega)\tilde{h}_i^*(\theta,\omega)\mathrm{d}\omega}{\int_{\omega_i}^{\omega_{i+1}} \tilde{h}_i^*(\theta,\omega)\tilde{h}_i(\theta,\omega)\mathrm{d}\omega}}. \tag{D.5}$$

The solution in (D.5) is equivalent to these two solutions since the numerator is the complex conjugate of the denominator

$$e^{j\alpha_{\theta,i}} = \pm\frac{\int_{\omega_i}^{\omega_{i+1}} \tilde{h}_i(\theta,\omega)\tilde{h}_i^*(\theta,\omega)\mathrm{d}\omega}{\left|\int_{\omega_i}^{\omega_{i+1}} \tilde{h}_i(\theta,\omega)\tilde{h}_i^*(\theta,\omega)\mathrm{d}\omega\right|}. \tag{D.6}$$

One of the solution corresponds to the minimum of J_i and the other to the maximum. Calculating the second derivative $\frac{\partial^2 J_i}{\partial \alpha_{\theta,i}^2}$ will allow us to derive which sign in (D.5) corresponds to the minimum of the function.

$$\frac{\partial^2 J_i}{\partial \alpha_{\theta,i}^2} = e^{j\alpha_{\theta,i}} \int_{\omega_i}^{\omega_{i+1}} \tilde{h}_i^*(\theta,\omega)\tilde{h}_i(\theta,\omega)\mathrm{d}\omega + e^{-j\alpha_{\theta,i}} \int_{\omega_i}^{\omega_{i+1}} \tilde{h}_i(\theta,\omega)\tilde{h}_i^*(\theta,\omega)\mathrm{d}\omega. \tag{D.7}$$

Estimating (D.7) for the two roots of the first derivative in (D.6) leads to

$$\begin{aligned}\left.\frac{\partial^2 J_i}{\partial d^2}\right|_{e^{j\alpha_{\theta,i}}} = &\left(\pm\frac{\int_{\omega_i}^{\omega_{i+1}} \tilde{h}_i(\theta,\omega)\tilde{h}_i^*(\theta,\omega)\mathrm{d}\omega}{\left|\int_{\omega_i}^{\omega_{i+1}} \tilde{h}_i(\theta,\omega)\tilde{h}_i^*(\theta,\omega)\mathrm{d}\omega\right|}\right) \int_{\omega_i}^{\omega_{i+1}} \tilde{h}_i^*(\theta,\omega)\tilde{h}_i(\theta,\omega)\mathrm{d}\omega \\ &+ \left(\pm\frac{\int_{\omega_i}^{\omega_{i+1}} \tilde{h}_i^*(\theta,\omega)\tilde{h}_i(\theta,\omega)\mathrm{d}\omega}{\left|\int_{\omega_i}^{\omega_{i+1}} \tilde{h}_i^*(\theta,\omega)\tilde{h}_i(\theta,\omega)\mathrm{d}\omega\right|}\right) \int_{\omega_i}^{\omega_{i+1}} \tilde{h}_i(\theta,\omega)\tilde{h}_i^*(\theta,\omega)\mathrm{d}\omega.\end{aligned} \tag{D.8}$$

The product of a complex value with its complex conjugate leads to a positive value. Therefore, the solution with the $+$ in (D.8) leads to a positive second derivative and the solution with the $-$ leads to a negative second derivative. Therefore, to minimize the cost function J_i, the solution is

$$c_i(\theta) = \frac{\int_{\omega_i}^{\omega_{i+1}} \tilde{h}_i(\theta_r,\omega)\tilde{h}_i^*(\theta,\omega)\mathrm{d}\omega}{\left|\int_{\omega_i}^{\omega_{i+1}} \tilde{h}_i(\theta_r,\omega)\tilde{h}_i^*(\theta,\omega)\mathrm{d}\omega\right|}. \tag{D.9}$$

Appendix E

Filtering of an Input Process by the Stochastic PAF

In this appendix, we present a mathematical derivation showing that (6.20) can be seen as a convolution and that it can be rewritten as in (6.21). Equation (6.20) is

$$R_O(\Delta\tau, \Delta t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R_I\left(\Delta t - \frac{y_{\Delta\tau} - y_0}{c}\right) f_{\mathbf{Y}}(\mathbf{y})\mathrm{d}y_{\Delta\tau}\mathrm{d}y_0. \tag{E.1}$$

We are showing that (E.1) can be rewritten as a general convolution in two dimensions

$$D(u, v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A(u - u', v - v')B(u', v')\mathrm{d}u'\mathrm{d}v', \tag{E.2}$$

where $A(u, v) = R_I(\frac{u-v}{c})$, $B(u', v') = f_{\mathbf{Y}}(\mathbf{y})$, $u = c\Delta t$, $v = 0$, $u' = y_{\Delta\tau}$ and $v' = y_0$. Note that (E.2) can easily be calculated in the Fourier domain as

$$S_D(\omega_u, \omega_v) = S_A(\omega_u, \omega_v)S_B(\omega_u, \omega_v), \tag{E.3}$$

with $S_A(\omega_u, \omega_v)$, $S_B(\omega_u, \omega_v)$ and $S_D(\omega_u, \omega_v)$ the 2D-FT of $A(u, v)$, $B(u, v)$ and $D(u, v)$, respectively.

The PSD $S_A(\omega_u, \omega_v)$ can be expressed as follows:

$$\begin{aligned} S_A(\omega_u, \omega_v) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R_I\left(\frac{u - v}{c}\right) e^{-j(\omega_u u + \omega_v v)}\mathrm{d}u\mathrm{d}v & \text{(E.4)} \\ &= cS_I(c\omega_u)\delta(\omega_u + \omega_v), & \text{(E.5)} \end{aligned}$$

with $S_I(\omega)$ the power spectral density of the input process $I(t)$. The 2D-FT of

$B(u,v)$ is

$$\begin{aligned} S_B(\omega_u,\omega_v) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} B(u,v)e^{-j(\omega_u u+\omega_v v)}\mathrm{d}u\mathrm{d}v \\ &= C\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-a(u^2+v^2-2buv)}e^{-j(\omega_u u+\omega_v v)}\mathrm{d}u\mathrm{d}v, \end{aligned} \quad \text{(E.6)}$$

with $a=\frac{R_Y(0)}{2|R_{\mathbf{Y}}|}$ and $b=\frac{R_Y(\tau)}{R_Y(0)}$ and $C=\frac{1}{2\pi\sqrt{|R_{\mathbf{Y}}|}}$. Relation (E.6) is obtained by considering (6.13). Further simplification of (E.6) leads to

$$\begin{aligned} S_B(\omega_u,\omega_v) &= C\int_{-\infty}^{\infty} e^{-a(1-b^2)v^2}e^{-j\omega_v v}\int_{-\infty}^{\infty} e^{-a(u-bv)^2}e^{-j\omega_u u}\mathrm{d}u\mathrm{d}v \\ &= C\sqrt{\frac{\pi}{a}}e^{-\frac{\omega_u^2}{4a}}\sqrt{\frac{\pi}{a(1-b^2)}}e^{-\frac{(\omega_u b+\omega_v)^2}{4a(1-b^2)}} \\ &= e^{-\frac{1}{4a}\left(\omega_u^2+\frac{(\omega_u b+\omega_v)^2}{(1-b^2)}\right)} \end{aligned} \quad \text{(E.7)}$$

From (E.2), we have that $R_O(\Delta\tau,\Delta t)=D(c\Delta t,0)$ with $D(c\Delta t,0)$ given by

$$\begin{aligned} D(c\Delta t,0) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} cS_I(c\omega_u)\delta(\omega_u+\omega_v)S_B(\omega_u,\omega_v)e^{j\omega_u c\Delta t}\mathrm{d}\omega_v\mathrm{d}\omega_u \\ &= \int_{-\infty}^{\infty} cS_I(c\omega_u)S_B(\omega_u,-\omega_u)e^{j\omega_u c\Delta t}\mathrm{d}\omega_u \\ &= \int_{-\infty}^{\infty} S_I(\omega)S_B(\frac{\omega}{c},-\frac{\omega}{c})e^{j\omega\Delta t}\mathrm{d}\omega, \end{aligned} \quad \text{(E.8)}$$

with $\frac{\omega}{c}=\omega_u$. To calculate $S_B(\frac{\omega}{c},-\frac{\omega}{c})$, (E.7) is used:

$$\begin{aligned} S_B(\frac{\omega}{c},-\frac{\omega}{c}) &= e^{-\frac{1}{4a}\left(\left(\frac{\omega}{c}\right)^2+\frac{\left(\frac{\omega}{c}\right)^2(1-b)^2}{(1-b^2)}\right)} \\ &= e^{-\left(\frac{\omega}{c}\right)^2(R_Y(0)-R_Y(\Delta\tau))}. \end{aligned} \quad \text{(E.9)}$$

Relation (E.8) can be seen as the inverse Fourier transform of the product of the two spectra $S_I(\omega)$ and $S_B(\frac{\omega}{c},-\frac{\omega}{c})$. Therefore, $R_O(\Delta\tau,\Delta t)$ is the convolution of the inverse Fourier transforms of these two spectra and is given by

$$R_O(\Delta\tau,\Delta t)=R_I(\Delta t)*\frac{ce^{-\frac{c^2(\Delta t)^2}{4\left(R_Y(0)-R_Y(\Delta\tau)\right)}}}{2\sqrt{\pi\left(R_Y(0)-R_Y(\Delta\tau)\right)}}. \quad \text{(E.10)}$$

Bibliography

[1] http://www.itakura.nuee.nagoya-u.ac.jp/HRTF/.

[2] http://lcavwww.epfl.ch/~thibaut/Measurements/.

[3] U. S. Federal Communications Commission. FCC 0248: First Report and Order.

[4] E. Adelson and J. Bergen, "The plenoptic function and the elements of early vision," in *Computational Models of Visual Processing*. MIT Press, 1991, pp. 3–20.

[5] T. Ajdler, L. Sbaiz, and M. Vetterli, "The plenacoustic function and its sampling," Ecole Polytechnique Federale de Lausanne, Tech. Rep., 2005.

[6] ——, "The plenacoustic function on the circle with application to HRTF interpolation," in *IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, 2005, pp. 273–276.

[7] ——, "The plenacoustic function and its sampling," *IEEE Transactions on Signal Processing*, vol. 54, no. 10, pp. 3790–3804, 2006.

[8] ——, "Room impulse responses measurement using a moving microphone," in *120th AES Convention*, Paris, 2006.

[9] T. Ajdler and M. Vetterli, "The plenacoustic function and its sampling," in *IEEE Benelux Workshop on Model Based Processing and audio Coding, MPCA*, 2002.

[10] ——, "Acoustic based rendering by interpolation of the plenacoustic function," in *SPIE/IS&T Visual Communication and Image Processing Conference*, vol. 5150, 2003, pp. 1337–1346.

[11] ——, "The plenacoustic function, sampling and reconstruction," in *IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, vol. 5, 2003, pp. 616–19.

[12] V. Algazi, R. Duda, D. Thompson, and C. Avendano, "The CIPIC HRTF database," in *IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics*, 2001.

[13] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, pp. 943–950, 1979.

[14] C. A. Balanis, *Antenna Theory: Analysis and Design.* John Wiley and Sons, 2005.

[15] D. R. Begault, *3-D Sound for Virtual Reality and Multimedia.* Academic Press, Cambridge, MA, 1994.

[16] P. Bello, "Characterization of randomly time-varying linear channels," *IEEE Transactions on Communication Systems*, pp. 360–393, 1963.

[17] A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," *J. Acoust. Soc. Am.*, vol. 93, no. 5, pp. 2764–2778, May 1993.

[18] A. Berkhout, *Applied Seismic Wave Theory.* Elsevier Science, 1987.

[19] A. Berkhout, D. de Vries, and J. Sonke, "Array technology and wave field analysis in enclosures," *J. Acoust. Soc. Am.*, vol. 102, pp. 2757–2770, 1997.

[20] J. Blauert, *Spatial Hearing.* MIT press, 2001.

[21] M. Blommer and G. Wakefield, "Pole-zero approximations for HRTF using a logarithmic error criterium," *IEEE Trans. Speech Audio Processing*, vol. 5, pp. 278–287, 1997.

[22] J. Borish, "Extension of the image model to arbitrary polyhedra," *J. Acoust. Soc. Am.*, vol. 75, pp. 1827–1835, 1984.

[23] J. Borish and J. Angell, "An efficient algorithm for measuring the impulse response using pseudorandom noise," *J. Audio Eng. Soc*, vol. 31, no. 7, pp. 478–488, 1983.

[24] D. Botteldooren, "Finite-difference time-domain simulation of low-frequency room acoustic problems," *J. Acous. Soc. Am.*, vol. 98, pp. 3302–3308, 1995.

[25] R. N. Bracewell, *The Fourier Transform and its Applications.* McGraw-Hill, 2000.

[26] J. Breebaart, J. Herre, L. Villemoes, C. Jin, K. Kjörling, and J. Plogsties, "Multichannel goes mobile: Mpeg surround binaural rendering," in *29th AES Conference*, 2006.

[27] C. Brown and D. R.O., "An efficient HRTF model for 3-d sound," in *Workshop on Applications of signal processing to audio and acoustics*, 1997.

[28] S. Carlile, C. Jin, and V. Van Raad, "Continuous virtual auditory space using HRTF interpolation: acoustic and psychophysical errors," in *Int. Symposium on Multimedia Information Processing*, 2000.

[29] A. Carlson, *Communication Systems*. McGraw-Hill, 1986.

[30] J. Chai, X. Tong, S. Chan, and H. Shum, "Plenoptic sampling," in *Conference on Computer graphics*, 2000, pp. 307–318.

[31] C. Charalambous and N. Menemenlis, "Stochastic models for short-term multipath fading channels: Chi-square and ornstein-uhlenbeck processes," *Proceedings of the 38th IEEE Conference on Decision and Control*, pp. 4959–4964, 1999.

[32] J. Chen, B. Van Veen, and K. Hecox, "A spatial feature extraction and regularization model for the head-related transfer function," *J. Acoust. Soc. Am.*, vol. 97, pp. 439–452, 1995.

[33] C. Cheng and G. H. Wakefield, "Introduction to Head-Related Transfer Func- tions (HRTFs): Representations of HRTFs in time, frequency and space," *J. Audio. Eng. Soc.*, pp. 231–249, 2001.

[34] F. Christensen, C. Jensen, and H. Moller, "The design of valdemar - an artificial head for binaural recording purposes," *109th AES Convention*, 2000.

[35] R. H. Clarke, "A statistical theory of mobile-radio reception," *Bell Sys. Tech. J.*, vol. 47, 1968.

[36] J. Coleman, "Ping-pong sample times on a linear array halve the nyquist rate," in *IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, 2004, pp. 925–928.

[37] ——, "Three-phase sample timing on a wideband triangular array of 4/3 the usual density reduces the nyquist rate for far-field signals by two thirds," in *The 38th Annual Asilomar Conference on Signals, Systems, and Computers, Pacific Grove CA, USA*, 2004.

[38] D. Colton and R. Kress, *Inverse Acoustic and Electromagnetic Scattering Theory*. Springer-Verlag, 1992.

[39] D. de Vries and M. Boone, "Wave field synthesis and analysis using array technology," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, New York*, 1999.

[40] D. de Vries, S. Brix, and E. Hulsebos, "Measurements of church impulse responses using a circular microphone array for natural spatial reproduction of a choir concert recording," in *21th AES Conference*, 2002.

[41] M. Do and M. Vetterli, "On the bandlimitedness of the plenoptic function," in *IEEE IEEE International Conference on Image Processing*, Genoa, Italy, 2005.

[42] R. O. Duda and W. Martens, "Range dependence of the response of a spherical head model," *J. Acoust. Soc. Am.*, vol. 104, pp. 3048–3058, 1998.

[43] R. Duraiswami, D. N. Zotkin, and N. Gumerov, "Interpolation and range extrapolation of HRTFs," in *IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, vol. 4, 2004, pp. 45–48.

[44] N. Durlach, A. Rigopulos, X. Pang, W. Woods, K. A., H. Colburn, and E. Wenzel, "Ion the externalization of auditory images," *Presence: Teleoperators and Virtual Environments*, vol. 1, pp. 251–257, 1992.

[45] G. Elko, E. Diethorn, and T. Gänsler, "Room impulse response variation due to thermal fluctuation and its impact on acoustic echo cancellation," in *IWAENC*, Kyoto, Japan, 2003, pp. 67–70.

[46] R. Ertel, P. Cardieri, K. Sowerby, R. T.S, and R. J.H, "Overview of spatial channel models for antenna array communication systems," *IEEE Pers. Commun. Mag.*, pp. 10–22, 1998.

[47] C. Faller, "Parametric coding of spatial audio," Ph.D. dissertation, Ecole Polytechnique Fédérale de Lausanne, 2004.

[48] T. Funkhouser, N. Tsingos, I. Carlbom, G. Elko, M. Sondhi, and J. West, "Modeling sound reflection and diffraction in architectural environments with beam tracing," in *Forum Acusticum*, 2002.

[49] M. Gans, "A power spectral theory of propagation in the mobile radio environment," *IEEE Trans. Vehic. Tech.*, vol. 21, pp. 27 – 38, 1972.

[50] B. Gardner and K. Martin, "HRTF measurements of a Kemar dummy-head microphone," MIT Media Lab, Tech. Rep., 1994.

[51] U. Gerlach, "Linear mathematics in infinite dimensions: Signals boundary value problems and special functions," Ohio State University, Course notes, 2006, http://www.math.ohio-state.edu/~gerlach/math/BVtypset/.

[52] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[53] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 4th ed. Academic Press, 1965.

[54] N. Gumerov and R. Duraiswami, *Fast Multipole Methods for the Helmholtz Equation in Three Dimensions*. Elsevier Science, 2005.

[55] W. Hartmann, "Localization of sound in rooms," *J. Acoust. Soc. Am.*, vol. 74, pp. 1380–1391, 1983.

[56] ——, "Localization of sound in rooms ii:the effects of a single reflecting surface," *J. Acoust. Soc. Am.*, vol. 78, pp. 524–533, 1985.

[57] K. Hartung, J. Braasch, and S. J. Sterbing, "Comparison of different methods for the interpolation of head-related transfer functions," in *AES 16th Int. Conf.*, 1999, pp. 319–329.

[58] S. Haykin, *Adaptive Filter Theory*. Prentice Hall, September 2001.

[59] E. Hulsebos and D. de Vries, "Parameterization and reproduction of concert hall acoustics measured with a circular microphone array," in *112th AES Convention*, 2002.

[60] E. Hulsebos, D. de Vries, and E. Bourdillat, "Improved microphone array configurations for auralization of sound fields by wave field synthesis," in *110th AES Convention*, 2001.

[61] E. Hulsebos, T. Schuurmans, D. de Vries, and R. Boone, "Circular microphone array for discrete multichannel audio recording," *114th AES Convention*, 2003.

[62] W. Jakes, *Microwave Mobile Communications*. John Wiley and Sons, 1974.

[63] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques*. Prentice-Hall, Englewood Cliffs, 1993.

[64] H. M. Jones, R. Kennedy, and T. Abhayapala, "On dimensionality of multipath fields: Spatial extent and richness," in *Proceedings of International Conference on Acoustics, Sensor and Signal Processing (ICASSP'02)*, 2002, pp. 2837–2840.

[65] H. Jones, "On multipath spatial diversity in wireless multiuser communications," Ph.D. dissertation, The Australian National University, 2001.

[66] T. Kailath, *Linear Systems*. Prentice-Hall, 1980.

[67] R. S. Kennedy, *Fading Dispersive Communication Channels*. New York: Wiley Interscience, 1969.

[68] Y.-H. Kim, "Can we hear the shape of a noise source?" in *The 18th International Congress on Acoustics*, 2004.

[69] D. Kistler and F. L. Wightman, "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," *J. Acoust. Soc. Am.*, vol. 91, pp. 1637–1647, 1992.

[70] M. Kleiner, B.-I. Dalenbäck, and P. Svensson, "Auralization - an overview," *J. Audio Eng. Soc.*, vol. 41, pp. 861–875, 1993.

[71] M. Kubovy and D. Van Valkenburg, "Auditory and visual objects," *Cognition*, vol. 80, pp. 97–126, 2001.

[72] A. Kulkarni and H. Colburn, "Infinite-impulse response models of the head-related transfer function," *J. Acoust. Soc. Am.*, vol. 115, pp. 1714–1728, 2004.

[73] A. Kulkarni, S. Isabelle, and H. Colburn, "On the minimum-phase approximation of head-related transfer functions," *IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics*, 1995.

[74] A. Kulowski, "Algorithms representation of the ray tracing technique," *Applied Acoustics*, vol. 18, pp. 449–469, 1985.

[75] H. Kutruff, *Room Acoustics*. Elsevier, 1991.

[76] H.-S. Kwon and Y.-H. Kim, "Moving frame technique for planar acoustic holography," *J. Acoust. Soc. Am.*, vol. 103, pp. 1734–1741, 1998.

[77] V. Larcher, J.-M. Jot, J. Guyard, and O. Warusfel, "Study and comparison of efficient methods for 3D audio spatialization based on linear decomposition of HRTF data," in *AES 108th Int. Convention*, 2000.

[78] B. Lathi, *Signal Processing and Linear Systems*. Oxford University Press, 1998.

[79] J.-Y. Le Boudec and M. Vojnovic, "Perfect Simulation and Stationarity of a Class of Mobility Models," *IEEE INFOCOM*, 2005.

[80] W. Lee, *Mobile Communications Engineering*. New York: McGraw Hill, 1982.

[81] M. Matsumoto, S. Yamanaka, M. Tohyama, and H. Nomura, "Effect of time arrival correction on the accuracy of binaural room impulse response interpolation," *J. Audio Eng. Soc.*, pp. 56–61, Jan./Feb. 2004.

[82] J. Miller, "SLAB: A software-based real-time virtual acoustic environment rendering system," in *International Conference on Auditory Display*, Espoo, Finland, 2001.

[83] P. Minaar, J. Plogsties, and F. Christensen, "Directional resolution of head-related transfer functions required in binaural synthesis," *J. Audio Eng. Soc.*, vol. 53, pp. 919–929, 2005.

[84] H. Moller, "Fundamentals of binaural technology," *Appl. Acoust.*, vol. 36, pp. 171–218, 1992.

[85] P. Morse and K. Ingard, *Theoretical Acoustics*. McGraw-Hill, 1968.

[86] S. Müller and P. Massarani, "Transfer function measurement with sweeps," *J. Audio Eng. Soc.*, vol. 49, pp. 443–471, 2001.

[87] P. S. Naidu, *Sensor Array Signal Processing*. CRC Press, 2001.

[88] G. Naylor, "Odeon - another hybrid room acoustical model," *Applied Acoustics*, vol. 38, pp. 131–143, 1993.

[89] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing*. Prentice–Hall, 1999.

[90] S. Petrausch, S. Spors, and R. Rabenstein, "Simulation and visualization of room compensation for wave field synthesis with the functional transformation method," in *119th AES Convention*, 2005.

[91] M. Poletti, "A unified theory of horizontal holographic sound systems," *J. Audio Eng. Soc*, pp. 1155–1182, 2000.

[92] ——, "Effect of noise and transducer variability on the performance of circular microphone arrays," *114th AES Convention*, 2003.

[93] A. Ridolfi, "Power spectra of random spikes and related complex signals," Ph.D. dissertation, Ecole Polytechnique Fédérale de Lausanne, 2004.

[94] A. Ridolfi and M. Win, "Power Spectra of Multipath Faded Pulse Trains," in *IEEE International Symposium on Information Theory*, 2005, pp. 102–106.

[95] S. M. Robeson, "Spherical methods for spatial interpolation: Review and evaluation," *Cartography and Geographic Information Systems*, vol. 24, no. 1, pp. 3–20, 1997.

[96] P. R. Runkle, M. A. Blommer, and G. H. Wakefield, "Comparison of head-related transfer function interpolation methods," in *Proc. IEEE Workshop on Appl. of Sig. Proc. to Audio and Acoust.*, 1995.

[97] A. A. M. Saleh and R. A. Valenzuela, "A physical spatio-temporal model of multipath propagation channels," *IEEE J. Sel. Areas Commun.*, vol. 5, p. 128137, 1987.

[98] L. Savioja, *Modeling Techniques for Virtual Acoustics.* Doctoral thesis, Helsinki University of Technology, Telecommunications Software and Multimedia Laboratory, Report TML-A3, 1999.

[99] L. Savioja, J. Backman, A. Järvinen, and T. Takala, "Waveguide mesh method for low-frequency simulation of room acoustics," *15th Int. Congr. Acoust. (ICA 95)*, vol. 2, pp. 637–640, 1995.

[100] T. Schetelig and R. Rabenstein, "Simulation of three-dimensional sound propagation with multidimensional wave digital filters," in *IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, 1998, pp. 3537–3540.

[101] P. Sharma, P. Sachetta, C. Thompson, and K. Chandra, "Channel models for indoor wireless transmission," in *Proceedings of International Conference on Third Generation Wireless and Beyond*, 2001, pp. 48–53.

[102] J. Smith, "Physical modeling synthesis update," *Computer Music J.*, vol. 20, pp. 44–56, 1996.

[103] S. Spors, "Active listening room compensation for spatial sound reproduction systems," Ph.D. dissertation, Erlangen-Nurnberg Universitat, 2005.

[104] S. Spors and R. Rabenstein, "Spatial aliasing artifacts produced by linear and circular loudspeaker arrays used for wave field synthesis," *120th AES Convention*, 2006.

[105] S. Spors, H. Teutsch, and R. Rabenstein, "High-quality acoustic rendering with wave field synthesis," in *Proc. of Vision, Modeling, and Visualization*, 2002, pp. 101–108.

[106] T. Theußl, T. Möller, and E. Gröller, "Optimal regular volume sampling," in *IEEE Visualization*, 2001, p. 9198.

[107] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[108] P. Vaidyanathan, *Multirate systems and filter banks*. Prentice Hall, 1992.

[109] R. A. Valenzuela, "A ray tracing approach for predicting indoor for wireless transmission," in *43rd IEEE Vehicular Technology Conf.*, 1993, pp. 214–218.

[110] E. N. G. Verheijen, "Sound reproduction by wave field synthesis," Ph.D. dissertation, Delft University of Technology, 1997.

[111] M. Vetterli and J. Kovačević, *Wavelets and Subband Coding*, ser. Signal Processing. Prentice Hall, 1995.

[112] E. G. Williams, *Fourier Acoustics*. Academic Press, 1999.

[113] J. Wozencraft and I. Jakobs, *Principles of Communication Engineering*. John Wiley and Sons, 1965.

[114] P. Zetterberg, P. Espensen, and P. Mogensen, "A downlink beam steering technique for gsm/dcs1800/pcs1900," in *IEEE PIMRC*, vol. 2, 1996, pp. 535–539.

[115] C. Zhang and T. Chen, "Generalized plenoptic function," Carneggie Mellon University, Tech. Rep. 01-06, September 2001.

[116] ——, "A survey on image-based rendering - representation, sampling and compression," Carneggie Mellon University, Tech. Rep. 03-03, 2003.

[117] L. J. Ziomek, *Fundamentals of Acoustic Field Theory and Space-Time Signal Processing*. CRC Press, 1995.

[118] U. Zölzer, *DAFX, Digital Audio Effects*. Wiley, 2002.

[119] D. Zotkin, R. Duraiswami, and L. Davis, "Rendering localized spatial audio in a virtual auditory space," *IEEE Transactions on Multimedia*, vol. 6, pp. 553–564, 2004.

[120] D. Zotkin, R. Duraiswami, E. Grassi, and N. Gumerov, "Fast head related transfer function measurement via reciprocity," *In press J. Acoust. Soc. Am.*, 2006.

Thibaut Ajdler

Audiovisual Communications Laboratory
Swiss Federal Institute of Technology Lausanne (EPFL)
1015 Lausanne, Switzerland

Personal

Date of birth:	December 10, 1979.
Nationality:	Belgian.
Civil status:	Married.

Education

2002-2006 **Ph.D.** candidate in Department of Communication Systems, Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland.

1996-2001 **M.Sc.** in Electrical Engineering, Vrije Universiteit Brussel, Belgium.

Professional Experience

2002 - present **Research and Teaching Assistant, Audiovisual Communications Laboratory (LCAV), EPFL, Switzerland**

- Research in signal processing applied to audio and acoustics
- Teaching assistant for undergraduate and graduate signal processing related classes
- Supervised students for semester and thesis projects

2003 **Internship at Intel Corporation, Santa Clara, CA, USA**

- Worked on source localization in distributed sensor networks under the supervision of Dr. I. Kozintsev
- Developed algorithms and a working system to localize multiple acoustic sources using distributed sensors

2000 **Research Assistant at Massachusetts Institute of Technology (MIT), Cambridge, MA, USA**

- Worked under the supervision of Dr. D. Dabby in the field of musical digital signal processing
- Developed and implemented various algorithms for musical variations using chaotic mapping

Awards and honors

- Awarded the Belgian American Education Foundation (BAEF) Fellowship
- Ranked 2nd out of more than 300 participants at the entry exam of the Polytechnique School of the Free University of Brussels (ULB).

Publications

Journal papers

1. T. Ajdler, L. Sbaiz and M. Vetterli, "The plenacoustic function and its sampling," IEEE Transactions on Signal Processing, Vol. 54, Nr. 10, pp. 3790-3804, 2006.

2. T. Ajdler, L. Sbaiz, C. Faller and M. Vetterli, "Interpolation of head-related transfer functions considering acoustics and temporal envelopes," *submitted to the Journal of the Audio Engineering Society*, 2006.

3. T. Ajdler, L. Sbaiz and M. Vetterli, "Room impulse responses measurement using a moving microphone or loudspeaker," *submitted to the Journal of the Acoustical Society of America*, 2006.

4. T. Ajdler, L. Sbaiz, A. Ridolfi and M. Vetterli, "Spatio-temporal channels have butterfly shaped spectra," *To be submitted to IEEE Transactions on Signal processing*, 2006.

Conference papers

1. T. Ajdler, R.L. Konsbruck, O. Roy, L. Sbaiz, E. Telatar and M. Vetterli, "Spatio-temporal sampling and distributed compression of the sound field," in *Proc. of European Signal Processing Conference (EUSIPCO) (invited paper)*, Sept. 2006.

2. T. Ajdler, L. Sbaiz and M. Vetterli, "Room impulse responses measurement using a moving microphone," in *Proc. of 120th Convention of the AES*, Paris, May, 2006.

3. T. Ajdler, L. Sbaiz, A. Ridolfi and M. Vetterli, "On a stochastic version of the plenacoustic function," in *Proc. of IEEE Conference on Acoustics, Speech and Signal Processing*, Vol. 4, pp. 1125-1128, 2006.

4. T. Ajdler, C. Faller, L. Sbaiz and M. Vetterli, " Head Related Transfer Functions Interpolation Considering Acoustics," in *Proc. of 118th Convention of the AES*, Barcelona, May, 2005.

5. T. Ajdler, L. Sbaiz and M. Vetterli, "The plenacoustic function on the circle with application to HRTF interpolation," in *Proc. of IEEE Conference on Acoustics, Speech and Signal Processing*, Vol. 3, pp. 273 - 276, 2005.

6. T. Ajdler, I. Kozintsev, R. Lienhart and M. Vetterli, "Acoustic Source Localization in Distributed Sensor Networks," in *Proc. of Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, Vol. 2, pp. 1328-1332, 2004.

7. T. Ajdler and M. Vetterli, "The Plenacoustic Function, Sampling and Reconstruction," in *Proc. of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY, pp. 147, 2003.

8. T. Ajdler, R. Cristescu, P.L. Dragotti, M. Gastpar, I. Maravic and M. Vetterli, "Distributed Signal Processing and Communications: on the Interaction of Sources and Channels," in *Proc. of IEEE Conference on Acoustics, Speech and Signal Processing*, Vol. 4, pp. 852-855, 2003.

9. T. Ajdler and M. Vetterli, "The Plenacoustic Function, Sampling and Reconstruction," in *Proc. of IEEE Conference on Acoustics, Speech and Signal Processing*, Vol. 5, pp. 616-619, 2003.

10. T. Ajdler and M. Vetterli, "Acoustic Based Rendering by Interpolation of the Plenacoustic Function," in *Proc. of SPIE/IS&T Visual Communications and Image Processing Conference (invited paper)*, Vol. 5150, pp. 1337-1346, 2003.

11. T. Ajdler and M. Vetterli, "The Plenacoustic Function and its Sampling," in *Proc. of IEEE Benelux Workshop on Model Based Processing and Coding of Audio (MPCA)*, Nov. 2002.

Technical report

1. T. Ajdler, L. Sbaiz and M. Vetterli, "The Plenacoustic Function and its Sampling," *Technical Report, EPFL*, 2005.

Thesis

1. T. Ajdler, "Comparison of pitch recognition algorithms," *M.Sc. Thesis Vrije Universiteit Brussel, Belgium*, 2001.